1995

Visualization And Analysis Of Multiple Sclerosis Lesions In Magnetic Resonance Images

J Ross Mitchell

National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Visualization and Analysis of Multiple Sclerosis Lesions in Magnetic Resonance Images

by

J. Ross Mitchell
Department of Medical Biophysics

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
September 22, 1995

ISBN 0-612-03477-1

Name J. Ross Mitchell

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

MEDICAL BIOPHYSICS — SUBJECT TERM

0160 — SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Abstract

Magnetic Resonance Imaging (MRI) provides very sensitive indication of the presence and extent of the lesions of Multiple Sclerosis (MS). As a result, quantifying changes in the number and extent of lesions in MR images has been used to indicate disease activity in clinical trials of new therapies. However, lesion quantification is long, arduous, and subject to variability between and within operators. These factors limit the application of quantitative techniques and increase the duration, and thus cost, of clinical trials which rely on their measurements to assess therapeutic effect.

A computerised system was developed for assisted 3-D quantification and analysis of MS lesions in standard spin-echo MR exams. Manual and assisted quantification were compared using repeated measurements of lesions in MR exams of a phantom and MS patient. Results indicate that assisted quantification reduced inter-operator variability by 1/3, and reduced intra-operator variability by 1/2. The minimum significant change between two successive measurements of lesion volume by the same operator, was 0.64 cm^3 for manual quantification, and 0.42 cm^3 for assisted quantification. For two different operators making successive measurements, the minimum significant change was 0.94 cm^3 for manual quantification, but only 0.47 cm^3 for assisted quantification.

Repeated measurements were also used to determine the impact on operator variability of: a) lesion quantification in high-field (1.5T) versus mid-field (0.5T) exams; and, b) an anisotropic diffusion filter algorithm which reduces image noise without blurring or moving object boundaries. Results suggest that inter- and intra-operator variability are reduced by anisotropic filtering, and by quantification in 0.5T exams. Reduced operator variabilities may result from higher detail signal-to-noise ratios (dSNR's) in 0.5T and filtered exams.

Finally, pathological change may occur within a lesion without a corresponding change in volume. Therefore, a new technique was developed to provide lesion

composition information from MR exam intensities. Analysis of serial exams of 3 MS patients revealed changes in the intensity spectra within lesions, even when their volume remained constant. Together, assisted lesion quantification and analysis may provide additional insight into disease activity, and improve the sensitivity of clinical trials of new therapies.

Acknowledgments

I would like to begin by thanking my supervisor, Dr. Aaron Fenster. He had the courage to accept me as a Computer Scientist, and the patience and perseverance to transform me into a Medical Physicist. In addition, I would like to thank Aaron for his generosity, both intellectual and financial, and for his amazing insight into my research and solutions to my problems. I also thank the other members of my supervisory committee: Dr. Chris Ellis and Dr. Steve Karlik. They have guided me during my studies. Their direction and advice have been invaluable. They put the "Medical" in Biophysics, never letting me forget the importance of physiology and the patient.

I would like to thank all the members of the Imaging Research Labs, through my many years here, for their support and open and hospitable nature. It's a great place to work. In particular, I would like to thank: Dr. Brian Rutt for always being available to read, comment and revise, even when he's *really* busy, and for teaching me so much about MRI; Dr. David Holdsworth and Dr. Maria Drangova, as role models, for showing me how its done, and done right!; Dr. Richard Frayne for so much help, for the time spent imaging, and for his thorough explanations of the intricacies of MRI; and, finally, to the "Boys Club down in The Hole" (Michel Moreau, Blaine Chronik, Chris Bowen, Louis Lauzon, Joe Gati, Ken Chu) I express my thanks for the lively conversations, great ideas, and always constructive criticism.

I would like to thank my other collaborators: Dr. Don Lee; Dr. George Rice; Craig Jones; and, Dr. Mike Eliasziw. They have all contributed to the experimentation, preparation and revision of my manuscripts. I also thank Jeffery Verhagen, Shi Sherebrin, and Lori Gardi, for their efforts repeatedly quantifying lesions in MR exams!

I acknowledge the financial support of the Multiple Sclerosis Society of Canada, the Medical Research Council of Canada, the Natural Sciences and Engi-

neering Research Council of Canada, and the Sterling-Winthrop Imaging Research Institute.

Finally, I thank the members of my family: Sheela, my wife, for her commitment, emotional support, putting up with the long nights in the lab, teaching me to be prompt, and for just saying yes!; Sheela's parents, Lalu and Charu, for support and great food; and to my own parents, George and Connie, for their faith, encouragement and financial support. It wouldn't have been possible without family.

Table of Contents

List of Tables

List of Figures

CHAPTER 1 INTRODUCTION

1.1 Multiple Sclerosis

1.1.1 General Description

Multiple Sclerosis (MS) is a neurological disorder afflicting nearly 1 in 1000 North Americans.[1] Initial symptoms occur typically between 10 and 50 years of age, with two thirds of the afflicted being female. MS is the most common of a family of diseases involving destruction of the myelin sheaths surrounding axons of the central nervous system (CNS). These sheaths act as insulators, and aid in the transmission of action potentials along the axonal membrane. Myelin loss, or demyelination, interferes with or disrupts neural transmission. The resulting symptoms vary widely, depending upon the extent of the damage, and its location within the CNS. However, common symptoms include: "stiffness"; spasticity or weakness in the limbs, particularly the legs; numbness; nystagmus (rapid oscillation of the eyeballs); slurred speech; dimmed vision; dementia and depression; fatigue; bladder and bowel dysfunction; and, sudden short seizures.[2]

MS is an autoimmune disorder; that is, myelin loss is the result of an immune mediated response against self.[3] Briefly, MS initially may involve defects in the blood-brain barrier (BBB) - the endothelial cells lining brain capillaries which mediate transport of macromolecules between the vascular system and the CNS. Defects in the BBB may allow immune components access to the CNS, where they recognize myelin basic protein as a foreign antigen, then attack it. Ultimately, this leads to demyelination, and may involve loss of oligodendrocytes (myelin producing cells) and axons themselves.[4,5] Macroscopically, the demyelinated regions appear as multiple grey, somewhat translucent plaques, or scleroses (thus the disease name).[4] The pathology of MS is described in more detail in Chapter 5, while excellent reviews are provided by Matthews et. al.,[4] Prineas,[3] and Steinman.[5]

1.1.2 Diagnosis and Clinical Course

Positive diagnosis is dependant upon the following criteria: 1) symptoms which indicate CNS white matter (WM) disease; 2) symptoms disseminated in time (separate attacks, or continual progression); 3) symptoms disseminated over space (involving different neural sites); and, 4) demonstration of clinically silent lesions through laboratory or imaging studies.[2] Through combinations of these criteria the diagnosis can vary from "Clinically Definite" to "Laboratory Supported Definite" to "Clinically Probable" to finally, "Laboratory Supported Probable", with gradations within each category.

After onset, the disease often follows a relapsing/remitting course in which attacks are followed by a latent phase.[2] Attacks are often sudden, occurring over days, hours or even minutes,[4] with a gradual easing of symptoms over the ensuing days or weeks.[2] Remissions may vary in length from a few weeks to years.[4] Often, the relapsing/remitting stage gives way to a progressive stage (chronic progressive), although some patients are progressive from the onset (primary progressive). Progression is defined as a gradual clinical deterioration without clear attacks or remissions.[2] Late age of onset appears to increase the chance of primary progression, and reduce the delay between the relapsing/remitting and chronic progressive stages. Typically half of all patients are categorized as progressive within 10 years of onset. The effect of MS on life-span varies, but some studies have shown that it reduces life expectancy by 10 years for men, and 15 years for women.[4] Disability is determined clinically, and graded in clinical trials using a scale such as Kurtzke's Expanded Disability Status Scale (EDSS).[6] This scale, summarized in Table 1.1, uses an objective assessment of the patient's capabilities to assign a grade of disability from 0 (normal) to 10 (death from MS). Patients can be examined and graded repeatedly during the course of a clinical trial to determine the effect of therapy. Yet strictly clinical evaluations, like the EDSS, may not reflect changes in clinically silent lesions.[7] These lesions can be demonstrated, however, using Magnetic Resonance Imaging (MRI). MRI allows

Grade	Description	Criteria
0	normal	normal neurological exam
1-2	minimal disorder	minimal abnormality
3-4	mild disorder	not sufficient to impede normal activities
5-6	disordered	not housebound, can walk, but not capable of a full day's work
7-9	severely disordered	limited to a wheelchair, or bed. Grade 9 patients are helpless.
10	death from MS	from brainstem involvement, cardiorespiratory failure, pneumonia via bed ridden state, or sepsis. Prior grade is 8 or 9.

Table 1.1 Kurtzke's Expanded Disability Status Scale (EDSS)

The EDSS is used to grade disorders due to Multiple Sclerosis. Patients are assigned a numerical value from 0 to 10 based on an objective assessment of their neurological, and physical capabilities.

the direct, non-invasive visualization of MS lesions, and can provide a sensitive indication of their change in time.[1,7,8] Thus, MRI, when combined with clinical evaluation, can provide more sensitive diagnosis and monitoring of this disease. Indeed, MRI is now considered an essential component of clinical trials to evaluate new therapies.[1]

1.2 Magnetic Resonance Imaging

1.2.1 General Description

Magnetic Resonance Imaging is a relatively new imaging technology, first demonstrated in 1972.[9] In the ensuing 2 decades, the development and application of this technology flourished, and continues to expand. MRI has become a valuable component of the radiologist's "tool chest" and is essential for the diagnosis of many disorders. Among the many advantages of MRI are that: a) it does not employ ionizing radiation; b) it can provide excellent resolution; c) it can provide images whose contrast is based upon many different physical characteristics,

including the density of hydrogen protons, tissue specific magnetic parameters, flowing or moving tissue, and diffusion; and, c) it can provide many of these images along arbitrary axes and in three-dimensions (3-D).

Magnetic resonance imaging is based upon the interaction between atomic nuclei in the object being imaged and strong externally applied magnetic fields. For the experiments described in this thesis, the objects are human patients or volunteers, or tissue mimicking test objects, called phantoms. The strong external magnetic fields were applied using commercial MR imaging systems (General Electric Signa, GE Medical Systems, Milwaukee) with field strengths of 0.5 Tesla or 1.5 Tesla. These same systems were used to induce and measure signals in the patient or phantom, and reconstruct images from the signals. Complete descriptions of these processes are described elsewhere,[10-12] and only a brief summary is presented here.

When an object is placed inside the MR imager, its atomic nuclei with a net magnetic moment align parallel or antiparallel to the main magnetic field, $\vec{B}_0$. The number of nuclei aligned with the main field is slightly larger than the number aligned against it, creating a net magnetic moment in the direction of $\vec{B}_0$. The nuclei, and thus the net magnetic moment, precess about the direction of $\vec{B}_0$ at a frequency given by the Larmor equation:[9]

$$f = \frac{\gamma}{2\pi}|\vec{B}_0| \tag{1.1}$$

where f is the frequency in hertz, γ is the gyromagnetic ratio. The hydrogen nucleus in water is the most important nucleus in medical imaging, since it is not only the most common nucleus with a magnetic moment, but also gives the strongest MR signal per nucleus. All of the experiments described in this thesis utilize images produced from MRI of hydrogen nuclei. To produce a signal which can be measured by the MR imager, radiofrequency (RF) energy at the Larmour frequency is applied to the object to tip the nuclei, or spins, into the plane which is

perpendicular, or transverse, to the main field. The RF energy also has the effect of ensuring that the transverse spins all have the same phase immediately following application of an RF pulse. The magnetization precessing in the transverse plane can then be detected using an RF receive coil placed near the object being imaged.

Spatial information can be encoded into the measured signal by applying a spatially varying magnetic gradient, $\vec{G}$, to the main magnetic field, $\vec{B}_0$. Equation 1.1 then becomes:[13]

$$f = \frac{\gamma}{2\pi} \cdot \left| (B_0 + \vec{G} \bullet \vec{r})\, \hat{z} \right| \tag{1.2}$$

where B_0 is the magnitude of the main magnetic field, $\vec{r}$ is the spatial location, and $\hat{z}$ is the unit vector in the direction of the main magnetic field. Equation 1.2 reveals, for example, that in the presence of a uniform gradient, frequency varies linearly with spatial location. Therefore, the frequency content of a signal measured in the presence of a known gradient provides information about the signal location. The Fourier Transformation may be used to convert frequency information to spatial information and construct an image from measured signals.

1.2.2 Relaxation Times T1 and T2

After the spins have been tipped into the transverse plane, they gradually return or relax back to the longitudinal or steady state condition with net magnetization parallel to $\vec{B}_0$. Specifically, the longitudinal magnetization M_z at time, t, is given by:[14]

$$M_z = M_0 \cdot \left(1 - e^{\frac{-t}{T1}} \right) \tag{1.3}$$

where M_0 is the initial or steady state magnetization and T1 is a tissue dependent time constant. T1 relaxation is due to interactions between the spins and the surrounding material, or lattice.

The signal measured in the receive coil is directly proportional to the net magnitude of spins in the transverse plane. These spins interact with each other and the local magnetic environment to produce small variations in the local magnetic field. In turn, this results in variations in the precessional frequency of the transverse spins which reduces the phase coherence imposed by the RF pulse. Loss of phase coherence reduces the net magnitude of the spins, and reduces the signal measured in the coil. The net transverse magnetization, M_{xy}, at time t is given by:[14]

$$M_{xy} = M_o \cdot e^{\frac{-t}{T2}} \tag{1.4}$$

where T2 is a tissue specific time constant. M_0 in equations 1.3 and 1.4 is proportional to the density of hydrogen protons in the tissue, ρ. Table 1.2 lists values for T1, T2 and ρ for white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF).

	WM	GM	CSF
T1 (ms)	624	1090	3302
T2 (ms)	62	104	2269
ρ	0.72	0.84	1.00

Table 1.2 T1, T2 and proton density for various brain tissues

The T1 (longitudinal relaxation time), T2 (transverse relaxation time) and proton density (ρ) for white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF), at 1.5T. The values in this table were obtained from Fischer *et. al*[15], and Condon *et. al.*[16] ρ is relative to that of CSF.

1.2.3 Spin Echo Imaging

Most MR imaging techniques utilize differences in T1, T2 or ρ to provide contrast between different tissues. One of the most common of these techniques is spin-echo imaging.[12] Spin echoes are formed by applying an RF pulse to tip the longitudinal magnetization 90 degrees into the transverse plane where the spins begin to lose phase coherence. Then, after a time TE/2 another RF pulse is applied to flip the transverse magnetization 180 degrees. This causes spins which had gained phase to lose phase, and spins which had lost phase to gain phase. After another time TE/2 the transverse spins again have phase coherence, and an "echo" is formed. During the echo the receive coil is turned on in the presence of a gradient, and the signal measured. By varying the spatial gradients and repeating this process, signals can be measured at a range of frequencies, and thus with information regarding a range of spatial locations. The Fourier Transform may be used to convert the measured signals into an image with intensity at coordinate (x,y) approximated by:[12]

$$S_{x,y} = \rho \cdot \left(1 - \exp\left(\frac{-TR}{T1}\right)\right) \exp\left(\frac{-TE}{T2}\right) \tag{1.5}$$

where $S_{x,y}$ is the signal intensity at location (x,y), TR is the time between spin-echo experiments, and TE is the time between the 90 degree RF pulse, and the formation of the echo. TR and TE can be varied to alter the contrast characteristics between tissues. In general, contrast based on T1 differences is provided by imaging sequences with a short TR (< 800 ms) and short TE (< 30ms). Contrast based on T2 differences is provided by sequences with long TR (> 1500 ms) and with TE near the T2 of the tissue of interest. Finally, contrast based on ρ differences is provided by sequences with long TR and as short a TE as possible.[12] Figure 1.1 shows spin-echo MR images weighted by T1, T2 and ρ, acquired at the same slice through the author's brain.

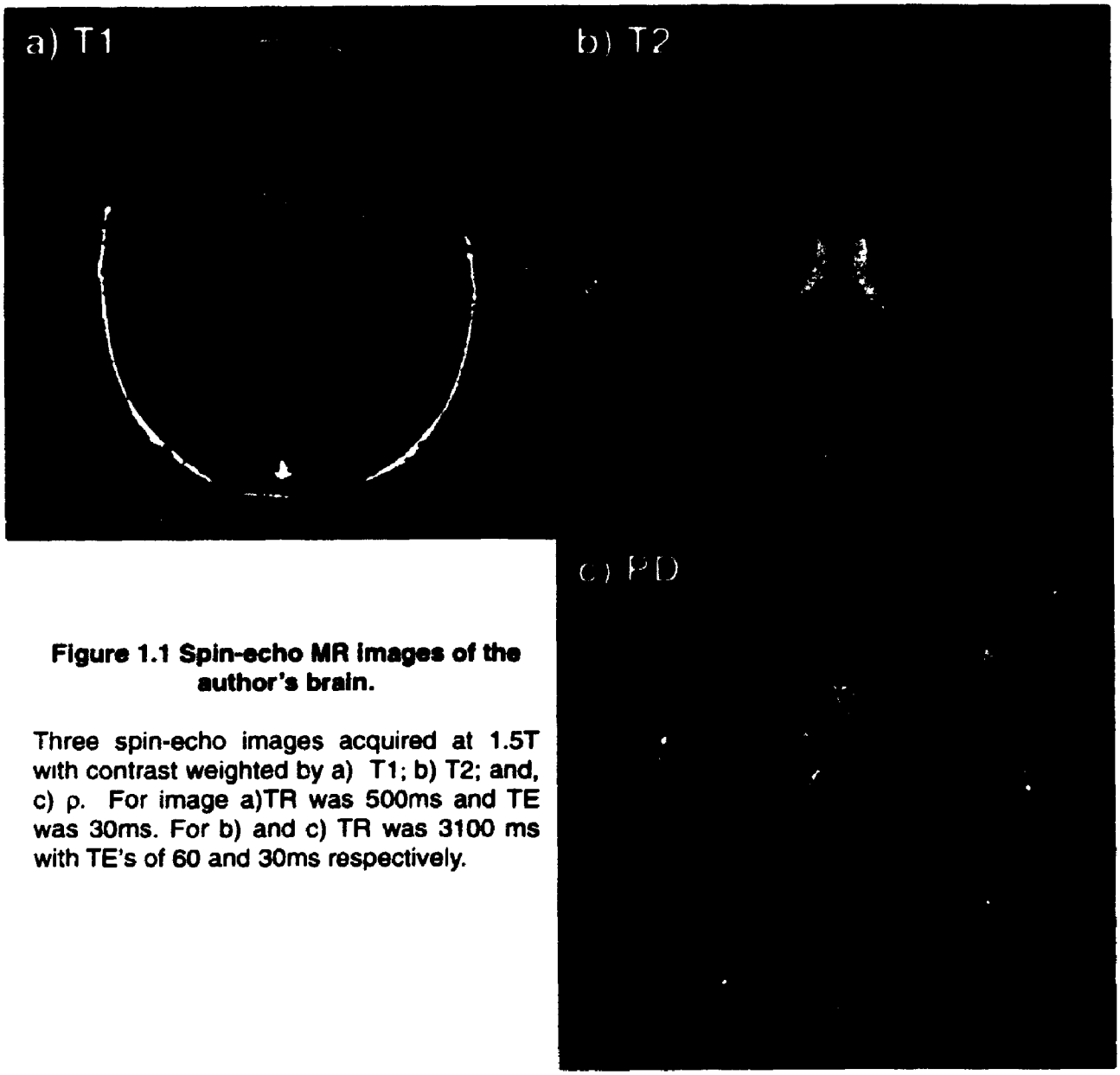


Figure 1.1 Spin-echo MR images of the author's brain.

Three spin-echo images acquired at 1.5T with contrast weighted by a) T1; b) T2; and, c) ρ. For image a)TR was 500ms and TE was 30ms. For b) and c) TR was 3100 ms with TE's of 60 and 30ms respectively.

1.3 MRI and Multiple Sclerosis

The variety of contrast mechanisms available to MRI means that it is well suited to neurological imaging applications, and imaging of Multiple Sclerosis in particular.[1,17] Although the relationship between lesion pathology and MR signal still is not completely understood, demyelination and edema at lesion sites tends to raise ρ and lengthen T1 and T2, compared to WM. Differences in T1 and T2

between WM, GM, CSF (listed in Table 1.2) and lesions may be exploited to provide large contrast between these tissues, even when little contrast is provided by x-ray based techniques such as computed tomography (CT).[12] Indeed, the first application of MRI to MS was by Young et. al. in 1981,[18] who examined 10 patients using both CT and MRI and demonstrated over 100 lesions in the MR images that were not evident in the CT images. Thereafter, a number of investigators began to acquire MR imaging systems, and apply them in clinical trials.[19-36] T2 weighted spin-echo sequences were determined to be very sensitive for MS lesion detection,[19] and have become the norm for clinical trials.[1] Exams typically consist of a series of images with in-plane resolution of 1 mm or less acquired from 5 mm thick axial slices with a slice gap of 2.5 mm or less. At each slice, two images are formed: a proton density weighted, early echo image in which CSF is less intense than GM; and, a T2 weighted, late echo image in which lesions are more intense than GM.[1] Figure 1.2 shows early and late echo images acquired from the same slice in the brain of an MS patient. In each image, a number of periventricular (PV) and deep white matter (DWM) lesions are indicated.

1.3.1 Quantification of MS lesions in MR images

In some clinical trials, assessment of changes in the MR exams has been used to determine disease presence and activity. For example, Thompson *et. al.*[27] and Feinstein *et. al.*[34] assigned a numerical score between 1 and 3 to each lesion based on measurement of its diameter. Cumulative score was then used to estimate lesion burden in the MR exam. Khoury *et. al.*[36] compared old and new MR images and counted the number of lesions and the number of gadolinium-DTPA enhancing lesions on T1 weighted MR exams. Changes in these numbers between exams were used to estimate improvement or deterioration in disease state. Similarly, Harris *et. al.*[28] counted the number of new gadolinium-DTPA enhancing lesions on T1 weighted MR exams, and Wiebe *et. al.*[31] determined the number, size, and gadolinium-DTPA enhancement on MR exams. In the latter experiment, lesion size was determined by visually estimating the largest lesion

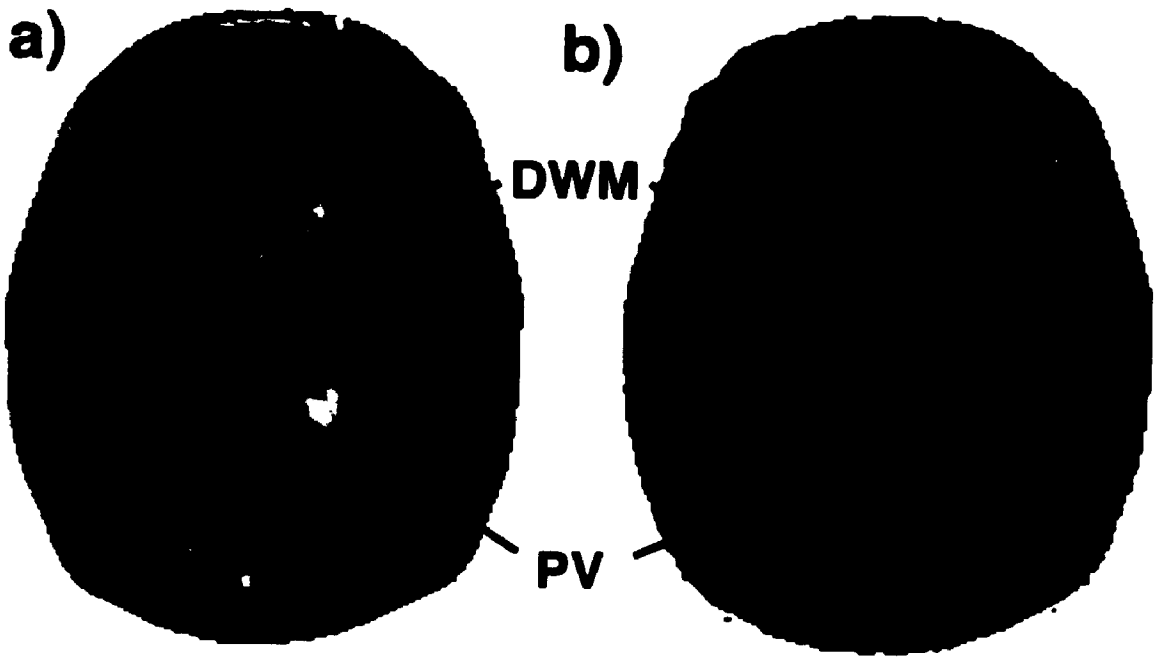


Figure 1.2 Spin-echo images from one slice in an MS patient's brain.

a) Early echo (TE=30ms) and b) late echo (TE=80ms) images from the same slice in the brain of an MS patient acquired at 1.5T. TR for both images was 2500ms. a) is proton density weighted, with GM more intense than CSF. b) is T2 weighted with lesions more intense than GM. MS lesions appear bright in both images. Deep white matter (DWM) and periventricular (PV) lesions are indicated in each image.

diameter. Morrissey *et. al.*[33] compared current and previous MR exams and counted the number of new lesions unequivocally identified as such by two neuroradiologists acting independently.

More quantitative assessments of lesion burden have also been used. Typically, these have utilized a computer workstation to assist in some aspect of lesion quantification. Much of the early research on computer-based lesion quantification was performed at the University of British Columbia, by D.W. Paty *et. al.*[21-24] Their technique is to view the MR images on a video display, then use a trackball to outline manually the lesion boundaries in each image. The computer then counts the number of voxels contained within the boundaries, and multiplies by the voxel volume to determine the lesion burden. Kappos *et. al.*[20] used a similar

procedure, except they outlined lesion boundaries on film reproductions of the MR images. More recently, Goodkin *et. al.*[30] used a computer console to outline manually lesions identified by a neuroradiologist.

However, even on a computer workstation, manual outlining of lesion boundaries is long and arduous and requires a knowledgeable operator. Therefore, many investigators have developed computerised global quantification schemes which determine the total lesion volume in an MR exam in a single operation. Armspach *et. al.*[35] empirically chose an intensity threshold in a T2 weighted image with high WM contrast. All voxels brighter than the threshold were labelled as lesion and their total volume measured. Pannizzo *et. al.*[37] reduced the operator dependence of this technique by using an algorithm to select the threshold level. Their algorithm examined the GM distribution in an intensity histogram from a T2 weighted image, then determined the maximum GM intensity, and labelled all brighter voxels as lesion.

Then in 1985 Vannier *et. al.* applied multispectral analysis techniques, originally developed for remote sensing, to the quantification of tissues in an MR exam.[38] Unlike earlier techniques, multispectral techniques utilize the additional information provided in multi-echo MR exams. Multispectral quantification is described in more detail in Chapters 2 and 5. Briefly, the images produced at each echo in an MR exam provide views of the same anatomy, but with different contrast characteristics. The additional contrast information can be used to better differentiate tissues. One common scheme is to construct a multidimensional histogram in which each axis corresponds to the range of intensities in one of the MR images. As a result, the histogram, or "feature space", shows the intensity distribution in the MR exam. Peaks or clusters in the feature space which correspond to tissues of interest can then be identified and used to label the corresponding tissue voxels in the MR exam.

Kikinis *et. al.*[39] and Jackson *et. al.*[40,41] used this scheme with an operator who identified and labelled sample voxels from CSF, "brain" (WM and GM), MS lesion, and "other" tissues (muscle and fat). The labelled data were then used to identify and color-code the feature space regions corresponding to each of these tissues. Identification of these regions was based on a statistical analysis of the labelled data, and the intensity distribution in feature space. The feature space regions were used then to color and label all voxels in the MR exam corresponding to each tissue. The total number of lesion voxels thus identified were counted and used to estimate lesion burden. Kamber *et. al.*[42] used a model of brain anatomy to further extend multispectral quantification. Their model provided the prior probabilities of GM, WM, and CSF for each voxel in the MR exam after it was transformed onto a standardized brain anatomy. The prior probabilities were then used to constrain identification of lesion voxels to regions of probable WM. Incorporation of the model reduced the number of false-positive lesions.

Artificial intelligence and artificial neural networks have also been used to quantify MS lesions. Raff *et. al.*[43] used an artificial neural network to identify MS lesions in a single echo, T2 weighted MR exam. They used 42 MR images acquired at approximately the same axial slice from healthy volunteers. These images were transformed onto an arbitrary standard, and used to train the neural network about "normal" anatomy. Images from MS patients were then fed into the network, which produced an image of "abnormal" tissues. A threshold was used to identify lesions in this image, and their volume was calculated. Zijdenbos *et. al.*[44] also used a neural network to identify WM lesions. However, their technique was multispectral, utilizing ρ, T2 and T1 weighted images. Like Kikinis *et. al.*, an operator identified and labelled regions of WM, GM, CSF and lesion. The labelled regions were used to train a neural network, which then operated on the entire MR exam to identify lesions. Finally, Raya obtained preliminary results from a rule-base system.[45] His approach was to examine voxels from various combinations of ρ and T2 weighted images (for example, their difference, or their sum

divided by their difference) then use a set of "if..then" statements to classify each voxel as "brain", parenchyma, CSF or MS lesion.

1.3.2 The Variability of Lesion Quantification

Even with the sophisticated quantification schemes described previously, there remains some disagreement about whether or not lesion burden measured from MRI correlates with clinical measures of disability. Some studies have found little or no correlation,[27] while others have found significant correlation.[36] The discrepancy may be due in part to variability both in clinical and MRI based measures.[25] Variability results both when different people make measurements (inter-operator variability) and when the same person repeats measurements (intra-operator variability). Often, variability is expressed as a coefficient of variation (COV), defined as:[46]

$$\mathrm{COV} = \frac{s}{\bar{x}} \bullet 100\% \tag{1.6}$$

where s is the standard deviation in the measurements, and $\bar{x}$ is the mean measured volume.

To date, only a few investigators have examined the inter- and intra-operator variability of MS lesion quantification. Paty *et. al.* had operators perform repeated outlining of lesions in a standard set of images over time.[21-24] From these measurements they calculated an intra-operator COV of 6% to 7%, and an inter-operator COV of 13% to 23%. Jackson *et. al.* had 4 operators make 4 repeated measurements in 5 MR exams using multispectral classification.[40] They reported a mean intra-operator COV of 6.5%, but a mean inter-operator COV of only 5.7%. Their results imply that the variability between operators was *less* than the variability within operators. However, intra-operator variability is actually one component of inter-operator variability, and therefore should be less than or equal to inter-operator variability. Unfortunately they did not specify the equations used to

calculate lesion quantification variability. Yet if they applied equations they used later to evaluate the variability of normal brain quantification[41] then this result may be due to the pooling of measurements to compare the mean results between operators when calculating inter-operator variability.

Goodkin *et. al.* eliminated inter-operator variability by having 1 operator make 3 repeated measurements of lesions in 1 exam using manual outlining.[30] They found that the intra-operator COV varied with the mean lesion volume ($\bar{x}$ in equation 1.6). In particular, for small lesions ($\bar{x} < 0.67$ cm^3) the COV was 23%, and for large lesions ($\bar{x} > 0.67$ cm^3) the COV was 12%. They conclude that the COV is difficult to compare between experiments without taking the mean lesion volume into account. Zijdenbos *et. al.* used a reliability coefficient and two measures of geometric agreement to compare manual outlining to artificial neural network based multispectral classification.[44] They had 2 operators make 2 repeated measurements using both quantification techniques in 6 exams of patients with Alzheimer's Dementia. The geometric measures indicated better operator agreement with the neural network technique. However, the reliability coefficient indicated poorer operator agreement with this technique. Reliability coefficients alone are also difficult to compare between experiments since they depend upon the variability in the subjects being measured in each experiment, and this is likely to vary.[47]

Many authors have emphasized the importance of operator variability on clinical trial design, and analysis.[1,3,8,30] Yet any measure of operator variability will have greater clinical utility if it is expressed in the same units as the measurements, and if it is independent of the particular operators and lesions selected for analysis. This will allow clinicians to predict absolute uncertainty, even when operators and lesions change, and thus allow better comparison of variability between clinical trials. In addition, it will allow evaluation of new quantification techniques and the impact on variability of new image processing algorithms, and such things as magnetic field strength, to be established.

Finally, lesion quantification in MR images alone may not reflect compositional or physiological changes inside lesions. However, these changes may provide valuable information on the efficacy of treatment and the success of clinical trials. As a result, many authors recommend development of alternative methods of lesion quantification and analysis from MR images.[1,4,7,8,48] However, the arduous and complex nature of many analysis techniques limits their application on a general basis to clinical trials. Also, variability in new and existing techniques increases the duration, and thus the cost, of clinical trials which rely on their measurements as an indicator of disease activity.[1] Thus, development of new analysis techniques which are easy to use, less variable, and which provide compositional information is critical for increased application, reduced cost, and improved sensitivity of clinical trials.

1.4 Research Goal

The goals of my research were: 1) to develop a computer assisted technique to allow the quick, reliable, and simple quantification of MS lesions in 3-D; 2) to develop a procedure to determine the variability and reliability of the technique; 3) to use this procedure to compare my new quantification technique to manual quantification; 4) to determine the impact of magnetic field strength and a noise reduction filter on quantification; 5) to use the technique to provide compositional information about MS lesions from standard spin-echo MR exams; and, 6) to apply the technique to the analysis of serial MR exams of MS patients involved in current clinical trials.

To fulfill these objectives, I developed a graphical, interactive computer system which allows visualization and quantification of multi-echo MR exams.[49] The system provides facilities for both manual delineation of lesion boundaries, and computer assisted identification of lesion regions in 2-D or 3-D. The assisted quantification is an implementation of a new technique I developed based on mul-

tispectral analysis of multi-echo MR exams.[38] The variability and reliability of lesion quantification were determined by 6 operators who made 5 repeated measurements of 5 lesions using both quantification techniques. Measurements were taken in filtered and unfiltered 1.5T and 0.5T MR exams of an MS patient. The data took over 10 months to acquire, and a total of 1200 lesion volumes were measured (i.e. 6 operators x 5 repeated measurements x 5 lesions x 2 techniques x 4 exams). Statistical techniques for the simultaneous assessment of inter- and intra-operator variability were applied to this data and used to determine the standard error of measurement and reliability of manual and assisted lesion quantification.[47]

To provide information about lesion composition I developed a technique to calibrate MR exam intensities over time, then compress the multidimensional intensitv information in the exams into 1-D. The resulting histograms provide information about the intensity composition, and thus the underlying proton density and T2, of MS lesions. Finally, I applied my system to monitor the volume and intensity composition over time of several lesions in a chronic-progressive and two relapsing/remitting MS patients involved in a clinical trial at University Hospital, London Ontario.

1.5 Thesis Outline

The body of work presented in this thesis consists of four papers published or submitted for publication in refereed journals. Each paper addresses one aspect of my research project, outlined above. Accordingly, Chapters 2 through 5 in this thesis each reflect the content of one of these papers. The published paper has been reproduced here with the permission of the publisher. Chapter 6 is a summary of my research project, and a discussion of future applications. Following is a brief description of Chapters 2 though 5.

1.5.1 A computerised system for MR exam analysis

Chapter 2 is the basis for a paper entitled "Computer-assisted Identification and Quantification of Multiple Sclerosis Lesions in MR Imaging Volumes in the Brain" by Mitchell, Karlik, Lee, and Fenster which was published in the March/ April 1994 Journal of Magnetic Resonance Imaging (pp. 197-208). D.H. Lee, a neuroradiologist, read and diagnosed the MR exams of the MS patient, and identified MS lesions for analysis. Karlik, Lee and Fenster aided the preparation and revision of the manuscript. Portions of this chapter have been presented at: the Annual Meeting of the Radiological Society of North America (RSNA) (1992); the Annual Meeting of the Society of Magnetic Resonance in Medicine (1992); the Annual Meeting of the American Association of Physicists in Medicine (1992); SPIE Medical Imaging (1992 and 1993); and, The Canadian Congress of Neurological Sciences (1993). The work in this chapter was awarded the University of Western Ontario's Dean's Award for Research Excellence, in 1993.

Chapter 2 is a description of the computerised system I developed for visualization and quantification of multi-echo MR exams. This chapter presents an overview of this system along with a brief description of its procedures and data structures. Also presented is a description of a lesion phantom I constructed and its application to determine system accuracy and precision when quantifying lesions. I used the techniques described in this chapter to quantify the ventricle volume, and the volumes of four large MS lesions, over time in a single chronic-progressive patient. 3-D volume visualizations of the lesion changes over time, as well as quantitative measures of their volume changes are also presented.

A number of interesting results were obtained from this work. First, I discovered that accurate and precise measurements of "small" objects (less than 7 cm^3) is possible. In addition, I showed that precision in volume measurement improves as volume increases and that variability in volume quantification is related to MR slice thickness. I also discovered that MS lesions have a distinct "signature" in the

intensity distribution of the dual echo MR exams of MS patients. Finally, visualization of quantified lesion volumes revealed the dynamic nature of MS, with some lesions expanding while others are contracting.

1.5.2 Quantification Reliability and Variability

Chapter 3 is the basis for a paper entitled "The Variability of Manual and Computer-assisted Quantification of Multiple Sclerosis Lesion Volumes" by Mitchell, Karlik, Lee, Eliasziw, Rice, and Fenster which has been submitted for publication in the journal Medical Physics. M. Eliasziw, a biostatistician, developed the statistical methods that I applied, and helped analyze the operator measurements. D.H. Lee read and diagnosed the MR exams of the MS patient and identified MS lesions for analysis. Portions of this chapter have been presented at: the Annual Meeting of the RSNA (1993); the World Congress on Medical Physics and Biomedical Engineering (1994); SPIE Visualization in Biomedical Computing (1994); and, the Joint Symposium of the Canadian Organization of Medical Physics and The Canadian College of Physicists in Medicine (1994). The work in this chapter was awarded a Scientific Exhibit Award of Merit by the RSNA (1993), and the University of Western Ontario's Dean's Award for Research Excellence (1995).

Chapter 3 presents work I performed to analyze the variability and reliability of MS lesion quantification using two quantification techniques provided by my system. In the first technique the operator manually defines the lesion border and the system counts the voxels contained within the border, and calculates the lesion volume. In the second technique the operator initially trains the system about non-lesion regions, then identifies one or more locations inside a single lesion. The system then uses the training information to estimate the lesion's voxels, and determine the lesion volume. Although both techniques technically are computer assisted, throughout this thesis the former is referred to as "manual quantification" and the latter as "assisted quantification".

The accuracy of both techniques was determined using exams of the lesion phantom described in Chapter 2. The reliability and variability were determined from exams of single MS patient. In both instances, 6 operators made repeated measures of phantom or patient lesions using both quantification techniques. The phantom study revealed that accuracy is improved by assisted quantification. The variability study revealed that assisted quantification increased operator reliability, and reduced operator variability. These results, in turn, mean that the minimum change in lesion volume between two single successive measurements that can be confidently detected is reduced by 1/3 for assisted quantification. Also the number of lesions which must be monitored to detect a change in mean volume is reduced by a factor of 2 to 4 for assisted quantification. These results suggest that assisted quantification may have important application in clinical trials.

1.5.3 The Effects of Field Strength and a Noise Reduction Filter on Quantification

Chapter 4 is the basis for a paper entitled "Quantification of Multiple Sclerosis Lesion Volumes in 1.5T and 0.5T Anisotropically Filtered and Unfiltered MR Exams" by Mitchell, Karlik, Lee, Eliasziw, Rice, and Fenster which has been submitted for publication in the journal Medical Physics. M. Eliasziw developed the statistical methods that I applied, and helped analyze the operator measurements. D.H. Lee read and diagnosed the MR exams of the MS patient and identified MS lesions for analysis. Like Chapter 3, portions of this chapter have been presented at: the Annual Meeting of the RSNA (1993); the World Congress on Medical Physics and Biomedical Engineering (1994); SPIE Visualization in Biomedical Computing (1994); and, the Joint Symposium of the Canadian Organization of Medical Physics and The Canadian College of Physicists in Medicine (1994). The work in this chapter was awarded a Scientific Exhibit Award of Merit by the RSNA (1993), and the University of Western Ontario's Dean's Award for Research Excellence (1995).

Chapter 4 presents work I performed to determine the impact of magnetic field strength, and the application of post-processing algorithms to reduce noise, on the variability and reliability of lesion quantification. The filter algorithm, implemented as part of my system, is modeled upon anisotropic diffusion, and reduces image noise without moving or blurring object boundaries.[50] Operator variability and reliability were determined from repeated quantifications of MS lesions in 1.5T and 0.5T filtered and unfiltered MR exams. The patient and the lesions studied in this experiment were the same as those studied in Chapter 3.

Results from this experiment indicate that operator variability is greatest in 1.5T unfiltered exams. Similarly, the minimum significant change in lesion volume between quantifications is larger, and the number of lesions which must be monitored to detect a change is greater, in 1.5T unfiltered exams. The increased variability in 1.5T unfiltered exams may be due to reduced detail signal-to-noise ratios, defined as:

$$dSNR = \frac{|S_A - S_B|}{\sigma} \tag{1.7}$$

where S_A and S_B are the mean image intensities measured from homogenous regions inside tissue A and tissue B, and σ is an estimate of image noise. Further studies revealed that filtering improved detail signal-to-noise ratios for all tissues. In addition, the detail signal-to-noise ratio between CSF and lesion was higher at the lower field strength. Since many lesions are periventricular, this may account for improved operator performance at the lower field strength.

1.5.4 Analysis of Lesion Intensity Composition

Chapter 5 is the basis for a paper entitled "Ma_ ..c Resonance Multispectral Analysis of Multiple Sclerosis Lesions" by Mitchell, Jones, Karlik, Lee, Rutt, and Fenster which has been submitted for publication in the journal Magnetic Resonance in Medicine. C. Jones and B. Rutt constructed the phantoms used in this

experiment, and helped with the imaging and analysis of the phantom data. S. Karlik helped analyze the results of the patient experiments. D. Lee read and diagnosed the MR exams of the MS patient, and assisted with the physiological interpretation of patient results. Portions of this chapter have been presented at The Canadian Congress of Neurological Sciences (1995).

Chapter 5 presents a new technique I developed to extract information about MS lesion composition from standard spin-echo MR exams. The new information is based on the multispectral intensity distribution of these lesions. In this chapter I derive expressions which show that this distribution reflects the underlying proton density and T2 of the lesion regions. MR exams of materials with known T1, T2 and proton densities were performed, and used to test the expressions I derived. I also used my technique to analyze lesions in standard spin-echo MR exams of three MS patients. Analysis involves constructing intensity histograms by reducing the multidimensional intensity information available in MR exams down to a single dimension. Information loss was minimized by projecting onto the principle component of the WM-GM cluster. The 1-D histograms enable better visualization and quantification of lesion intensities. Lesion analyses were compared over time and between patients by first calibrating the intensity information in the MR exams. Calibration is based upon the intensity characteristics of CSF, an internal reference tissue.

Results reveal that calibration eliminated variability in the intensity distributions of CSF, as expected. Calibration also greatly reduced variability in the intensity distributions of normal appearing white matter. Many of the lesion distributions had a distinctive peak at low intensity corresponding to normal appearing WM. Within the lesion distributions, increases in high intensity peaks were accompanied generally by reductions in the WM peak. Increases at high intensities may correspond to increased inflammation, while regrowth of the WM peak may represent reduced inflammation, or remyelination at lesion sites. Serial analysis of the lesion distributions revealed some dramatic fluctuations. In some lesions, intensity

fluctuations occurred even while the quantified lesion volume remained constant. Changes in the lesion intensity distribution may be a more sensitive indicator of disease activity than changes in lesion volume.

1.6 References

[1] D.H. Miller, F. Barkhof, I. Berry, L. Kappos, G. Scotti, A.J. Thompson, "Magnetic Resonance Imaging in Monitoring the Treatment of Multiple Sclerosis: Concerted Action Guidelines." Journal of Neurology Neurosurgery and Psychiatry, **54**, 683-688, (1991).

[2] A.E. Miller, "Clinical Features" in *The Handbook of Multiple Sclerosis*, S.D. Cook editor. New York: Marcel Dekker Inc., 171-177 (1990).

[3] J.W. Prineas, "Multiple Sclerosis: Pathology of the Early Lesion." In *Multiple Sclerosis: Current Status of Research and Treatment*, R.M. Herndon, F.J. Seil, editors. New York: Demos Publications, (1994).

[4] W.B. Matthews, A. Compston, I.V. Allen, C.N. Martyn, *McAlpine's Multiple Sclerosis*. Edinburgh, Scotland: Churchill Livingstone, 2nd edition, (1991).

[5] L. Steinman, "Autoimmune Disease." Scientific American Special Issue: Life, Death and the Immune System. **269**, 106-114, (1993).

[6] J.F. Kurtzke, "Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS)". Neurology, **33**, 1444-1452 (1983).

[7] C.J. Wallace, T.P. Seland, and T.C. Fong, "Multiple Sclerosis: The Impact of MR Imaging." American Journal of Radiology, **158**, 849-857, (1992).

[8] D.W. Paty, "Neuroimaging in Multiple Sclerosis" in *The Handbook of Multiple Sclerosis*, S.D. Cook editor. New York: Marcel Dekker Inc., 291-316 (1990).

[9] S. C. Bushong, *Magnetic Resonance Imaging: Physical and Biological Principles*. St. Louis, Missouri: Mosby, (1988).

[10] S. Balter, "An Introduction to the Physics of Magnetic Resonance Imaging". RadioGraphics, 7, 371-383 (1989).

[11] P. Sprawls, "The Magnetic Resonance Image: A Physical Perspective" in *The Physics of MRI: 1992 AAPM Summer School Proceedings*, Sprawls, P., and Bronskill, M. J., editors. Woodbury NY, AIP, 1-14 (1993).

[12] D. B. Plewes, and J. Bishop, "Spin-Echo MR Imaging" in *The Physics of MRI: 1992 AAPM Summer School Proceedings*, Sprawls, P., and Bronskill, M. J., editors. Woodbury NY, AIP, 166-187 (1993).

[13] S.J. Riederer, "Spatial Encoding and Image Reconstruction" in *The Physics of MRI: 1992 AAPM Summer School Proceedings*, Sprawls, P., and Bronskill, M. J., editors. Woodbury NY, AIP, 135-165 (1993).

[14] P.S. Allen, "Some Fundamental Principles of Nuclear Magnetic Resonance" in *The Physics of MRI: 1992 AAPM Summer School Proceedings*, Sprawls, P., and Bronskill, M. J., editors. Woodbury NY, AIP, 15-31 (1993).

[15] H.W. Fischer, P.A. Rinck, Y.V. Haverbeke, R.N. Muller, "Nuclear Relaxation of Human Brain Gray and White Matter: Analysis of Field Dependence and Implications for MRI." Magnetic Resonance in Medicine, **16**, 317-334 (1990).

[16] B. Condon, J. Patterson, A. Jenkins, D. Wyper, D. Hadley, R. Grant, J. Rowan, G. Teasdale, "MR Relaxation Times of Cerebrospinal Fluid." Journal of Computer Assisted Tomography, **11(2)**, 203-207 (1987).

[17] W. Kucharczyk, "Neurological MRI" in *The Physics of MRI: 1992 AAPM Summer School Proceedings*, Sprawls, P., and Bronskill, M. J., editors. Woodbury NY, AIP, 412-429 (1993).

[18] I.R. Young, A.S. Hall, C.A. Pallis, G.M. Bydder, N.J. Legg, R.E. Steiner, "Nuclear Magnetic Resonance Imaging of the Brain in Multiple Sclerosis". Lancet, **2**,1063-1066 (1981).

[19] I.E.C. Ormerod, D.H. Miller, W.I. McDonald, et al. "The role of NMR imaging in the assessment of multiple sclerosis and isolated neurological lesions; a quantitative study". Brain, **110**, 1579-1616 (1987).

[20] L. Kappos, D. Stadt, M. Ratzka, et al. "Magnetic resonance imaging in the evaluation of treatment in multiple sclerosis". Neuroradiology, **30**, 299-302 (1988).

[21] C. Isaac, D.K. Li, M. Genton, C. Jardine, E. Grochowski, M. Palmer, L.F. Oger, D.W. Paty, "Multiple sclerosis: A serial study using MRI in relapsing patients". Neurology, **38(10)**, 1511-1515 (1988).

[22] E.W. Willowghby, E. Grochowski, J. Oger, L.F. Kastrukoff, D.W. Paty, "Serial magnetic resonance scanning in multiple sclerosis: A second prospective study in relapsing patients". Annals of Neurology, **25**, 43-49 (1989).

[23] R.A. Koopmans, D.K.B. Li, J.J.F. Oger, L.F. Kastrukoff, C. Jardine, L. Costley, S. Hall, E.W. Growchowski, D.W. Paty, "Chronic Progressive Multiple Sclerosis: Serial Magnetic Resonance Brain Imaging Over Six Months". Annals of Neurology, **26**, 248-256 (1989).

[24] R.A. Koopmans, D.K.B. Li, E. Grochowski, P.J. Cutler, D.W. Paty, "Benign versus Chronic Progressive Multiple Sclerosis: Magnetic Resonance Imaging Features". Annals of Neurology, **25**, 74-81 (1989).

[25] J.H. Noseworthy, M.K. Vandervoort, C.J. Wong, G.C. Ebers, "Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial". Neurology, **40(6)**, 971-975 (1990).

[26] S.J. Karlik, M.K. Vandervoort, M. Hopkins, J.H. Noseworthy "Serial quantitative brain MRI in multiple sclerosis using visual and volumetric techniques". Neurology, 40, 142 (1990).

[27] A.J. Thompson, A.G. Kermode, D.G. MacManus, et al. "Patterns of disease activity in multiple sclerosis: clinical and magnetic resonance imaging study". British Medical Journal, **300**, 631-634 (1990).

[28] J.O. Harris, J.A. Frank, N. Pattronas, D.E. McFarlin, H.F. McFarland, "Serial Gadolinium-enhanced Magnetic Resonance Imaging Scans in Patients with Early, Relapsing-Remitting Multiple Sclerosis: Implications for Clinical Trials and Natural History". Annals of Neurology, **29**, 548-444 (1991).

[29] M.D. Weinshenker, R.N. Bass, S.J. Karlik, G.C. Ebers, G.P.A. Rice, "An open trial of OKT3 in patients with multiple sclerosis." Neurology, 41, 1047 (1991).

[30] D.E. Goodkin, J.S. Ross, S. Vanderbrug-Medendorp, J. Konecsni, R.A. Rudick, "Magnetic resonance imaging lesion enlargement in multiple sclerosis". Archives of Neurology, **49**, 261-263 (1992).

[31] S. Wiebe, D.H. Lee, S.J. Karlik, M. Hopkins, M.K. Vandervoort, C.J. Wong, L. Hewitt, G.P.A. Rice, G.C. Ebers, J.H. Noseworthy, "Serial Cranial and Spinal Cord Magnetic Resonance Imaging in Multiple Sclerosis." Annals of Neurology, **32**, 643-650 (1992).

[32] D.W. Paty, D.K.B. Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group "Interferon beta-1b is effective in relapsing-remitting multiple sclerosis". Neurology, **43**, 662-667 (1993).

[33] S.P. Morrissey, D.H. Miller, B.E. Kendall, D.P.E. Kingsley, M.A. Kelly, D.A. Francis, D.G. MacManus, W.I. McDonald, "The significance of brain magnetic resonance imaging abnormalities at presentation with clinically isolated syndromes suggestive of multiple sclerosis." Brain, **116**, 135-146 (1993).

[34] A. Feinstein, M. Ron, A. Thompson, "A serial study of psychometric and magnetic resonance imaging changes in multiple sclerosis." Brain, **116**, 569-602 (1993).

[35] J.P. Armspach, D. Gounot, I.J. Namer, H.H. Ohlenbusch, L. Rumbach, J. Chambron, "Quantitative Cerebral Magnetic Resonance Imaging during ACTH Treatment of Multiple Sclerosis". Magnetic Resonance Imaging, **11**, 1147-1153 (1993).

[36] S.J. Khoury, C.R.G. Guttman, E.J. Orav, M.J. Hohol, S.S. Ahn, L. Hsu, R. Kikinis, G.A. Mackin, F.A. Jolesz, H.L. Weiner, "Longitudinal MRI in multiple sclerosis: Correlation between disability and lesion burden." Neurology, **44**, 2120-2124, (1994).

[37] F. Pannizzo, M.J.B. Stallmeyer, J. Friedman, R.J. Jennis, J. Zabriskie, C. Plank, R. Zimmerman, J.P. Whalen, P.T. Cahill, "Quantitative MRI Studies for Assessment of Multiple Sclerosis". Magnetic Resonance in Medicine, **24**, 90-99 (1992).

[38] M.W. Vannier, R.L. Butterfield, D. Jordan, W.A. Murphy, R.G. Levitt, M. Gado, "Multispectral Analysis of Magnetic Resonance Images". Radiology, **154**, 221-224 (1985).

[39] R. Kikinis, F.A. Jolesz, G. Gerig, T. Sandor, H.E. Cline, W.E. Lorensen, M. Halle, S.A. Benton, "3D Morphometric and Morphologic Information Derived from Clinical Brain MR Images". In: *NATO ASI Series, Vol. F60: 3D Imaging in Medicine*. Eds. K.H. Hohne *et. al.*, Berlin, Springer-Verlag, 441-454 (1990).

[40] E.F. Jackson, P.A. Narayana, J.S. Wolinsky, T.J. Doyle, "Accuracy and Reproducibility in Volumetric Analysis of Multiple Sclerosis Lesions". Journal of Computer Assisted Tomography, **17(2)**, 200-205 (1993).

[41] E.F. Jackson, P.A. Narayana, J.C. Falconer, "Reproducibility of Nonparametric Feature Map Segmentation for Determination of Normal Human Intracranial Volumes with MR Imaging Data". Journal of Magnetic Resonance Imaging, **4**, 692-700 (1994).

[42] M. Kamber, D.L. Collins, R. Shinghai, G.S. Francis, A.C. Evans, "Model-based 3D segmentation of multiple sclerosis lesions in dual-echo MRI data". SPIE Visualization in Biomedical Computing, **1808**, 590-600 (1992).

[43] U. Raff, F.D. Newman, "Automated lesion detection and lesion quantification in MR images using autoassociative memory". Medical Physics, **19(1)**, 71-77 (1992).

[44] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, A.C. Palmer, "Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation". IEEE Transactions on Medical Imaging, **13(4)**, 716-724 (1994).

[45] S.P. Raya, "Low-Level Segmentation of 3-D Magnetic Resonance Brain Images - A Rule-Based System". IEEE Transactions on Medical Imaging, **9(3)**, 327-337 (1990).

[46] P. Armitage, G. Berry, *Statistical methods in medical research.* 2nd edition. Oxford: Blackwell Scientific Publications, 1-559 (1990).

[47] M. Eliasziw, S.L. Young, M.G. Woodbury, K. Fryday-Field, "Statistical Methodology for the Concurrent Assessment of Interrater and Intrarater Reliability: Using Goniometric Measurements as an Example". Physical Therapy, 74(8), 777-788 (1994).

[48] T.L. Richards, "Proton MR Spectroscopy in Multiple Sclerosis: Value in Establishing Diagnosis, Monitoring Progression, and Evaluating Therapy." American Journal of Radiology, **157**, 1073-1078, (1991).

[49] J.R. Mitchell, S.J. Karlik, D.H. Lee, A. Fenster, "Computer Assisted Identification and Quantification of Multiple Sclerosis Lesions in MR Volumes of the Brain". Journal of Magnetic Resonance Imaging, **4**, 197-208, (1994).

[50] G. Gerig, O. Kübler, R. Kikinis, F.A. Jolesz, "Nonlinear Anisotropic Filtering of MRI Data." IEEE Transactions on Medical Imaging, **11 (2)**, 221-232, (1992).

CHAPTER 2 COMPUTER ASSISTED IDENTIFICATION AND QUANTIFICATION OF MULTIPLE SCLEROSIS LESIONS IN MR VOLUMES OF THE BRAIN

2.1 Introduction

Magnetic Resonance Imaging (MRI) provides excellent soft tissue differentiation and is often used to assist in confirming the clinical diagnosis of MS. Demyelination and edema in MS lesions results in increased unbound water at lesions sites. This increase causes increased intensities at lesion locations compared to normal white and grey matter in spin-echo MR images. Qualitative assessment of the changes in the appearance and extent of high intensity regions in spin-echo images of MS patients has been used to evaluate disease activity and progression.[1-4] However, quantitative measurement of these changes is now considered essential for accurate assessment of disease state.[1,6] This approach to disease assessment has been taken in our laboratory,[7-9] and other laboratories.[10-17] However, the vast quantities of data produced make routine manual analysis arduous and difficult. Since the imaging modality is inherently digital, MR exams for MS are well suited for computer aided analysis. However, the wide range of required knowledge, complicated physiological processes, and the multifaceted appearance of MS lesions in MR images makes the automated identification of MS lesions difficult to describe deterministically. Nevertheless, attempts at deterministic classification of brain anatomy have been carried out, and analysis techniques employing varying degrees of automation have been proposed.[18-22]

In this Chapter I describe a supervised approach to the detection and analysis of MS lesions in proton density (PD) and T2 weighted 3-D MR images. I have designed a semi-automated system utilizing multispectral classification techniques[23-27] which is closely coupled with a graphical computer workstation to provide interactive assistance to an experienced, knowledgable operator. In this

way, the burden of routine analysis of many MR exams can be alleviated while still relying on human intelligence for detection and diagnosis.

2.2 Materials and Methods

2.2.1 System Overview

The system has been implemented on a Sun Sparc II Workstation (Sun Microsystems, Mountain View, California) with 32 megabytes of memory running SunOs, Sun Network File System (NFS) and Openwindows (an X11 based window manager). It is written in C using the Xview X11 Toolkit, the Graphical User Interface Development Environment (DevGuide) for user interface development and the Revision Control System (RCS) for source code control. This environment allows interactive response to most functions.

System organization is a result of four design goals: 1) to provide simple, direct manipulation of the MR images; 2) to make all operations interactive; 3) to make operations intuitive and simple to use; and, 4) to make the system easy to extend and modify. Figure 2.1 is a block diagram showing the major components of my system and the possible flow of information between them.

3-D MR image. - Images were acquired on a 1.5T GE Signa imaging system (GE Medical Systems, Milwaukee) using a multiple spin-echo technique with TR/TE_1/TE_2 = 2000/35/70 ms, 1 NEX and gradient moment nulling for velocity compensation. The two echoes had bandwidths of 16 and 8 kHz, respectively. Exams consisted of between 26 and 49 adjacent, non overlapping slices which were either 5mm or 3mm thick. At each slice, images with a 20cm field-of-view were produced for each echo. Exams were transferred via ethernet to the Sparc workstations for processing. PD and T2 weighted 3-D images were created from each exam by combining all early echo images and all later echo images, respectively.

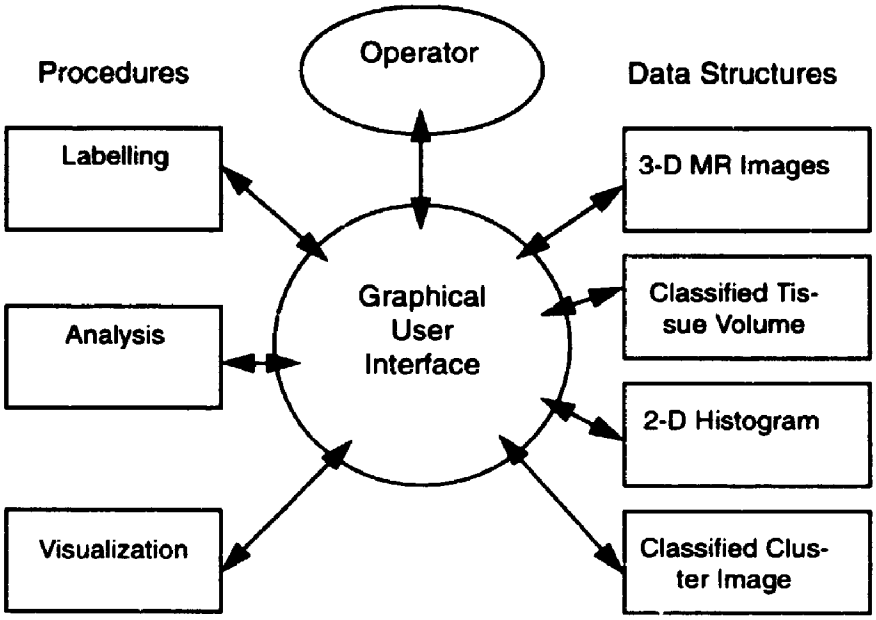


Figure 2.1 System procedures and data structures.

My system is based on four Data Structures: 1) a 3-D MR image containing PD and T2 weighted intensities for each voxel; 2) classified tissues in the 3-D MR image; 3) a 2-D histogram showing the frequency of PD/T2 intensity combinations in the MR data; 4) classified clusters in the 2-D histogram. The interactive nature of the system facilitates the flow of data between each of these structures. System Procedures fall into one of three general categories: labelling, analysis or visualization. Labelling Procedures are used to identify regions-of-interest and connected components within the data structures. Analysis Procedures are used to determine morphological and statistical parameters for labelled regions. 3-D visualization is provided by exporting the data structures to commercially available packages.

2-D Histogram. - Information about the various tissues in the 3-D images was obtained by examining a two dimensional (2-D) histogram showing the frequency of voxels with specific PD and T2 weighted intensities.[24,26,27] The intensity at each location in the histogram is proportional to the number of voxels with the PD weighted intensity given on the horizontal axis and the T2 weighted intensities given on the vertical axis. Figure 2.2 shows a pair of PD and T2 images from a single slice in the brain of an MS patient, and their corresponding 2-D histogram. The histogram has a number of distinct clusters obliquely oriented and extending outward from the origin. Ideally, each cluster would correspond to a distinct feature or tissue in the images. In practice, blurring between clusters due to partial volume effects, and radio frequency (RF) nonlinearities occurred.[23,26] To increase cluster separability, anisotropic diffusion filtering[26,28] may be performed on the 3-

D images. Although this technique has been beneficial in some tissue quantification applications[13,26] its effect upon MS lesions has not been established. Therefore, it was not applied to the data analyzed in this experiment.

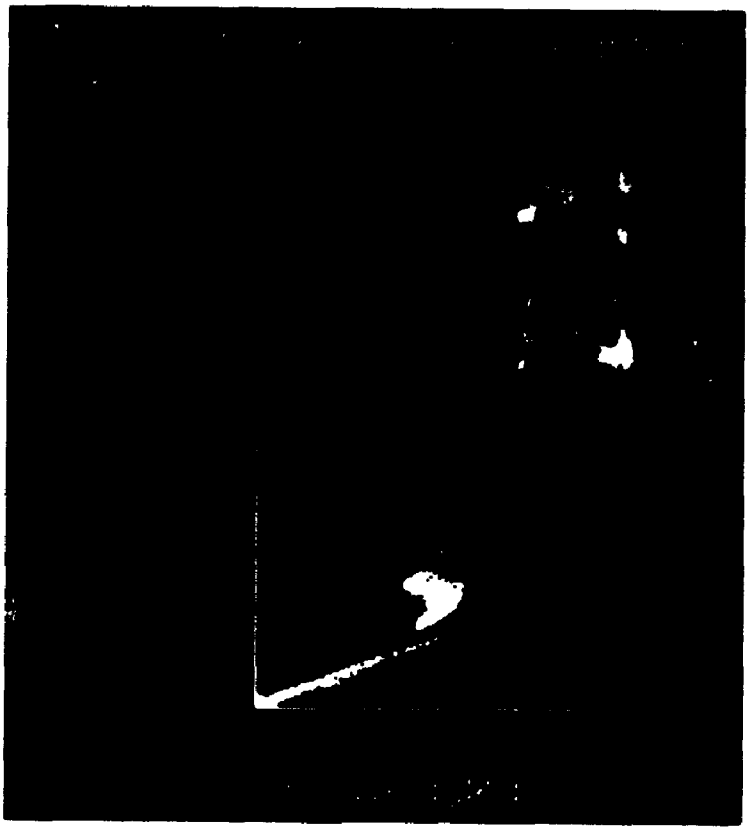


Figure 2.2 A pair of spin-echo images, and their 2-D histogram.

The 2-D histogram corresponding to a pair of PD and T2 weighted images showing the frequency of all possible PD/T2 intensity combinations. The vertical axis represents the range of possible T2 weighted intensities. The horizontal axis represents the range of PD weighted intensities. The intensity at each location in the histogram is proportional to the frequency of a particular PD/T2 intensity combination. This histogram is calculated and displayed for manipulation by the system operator.

Classified Cluster Image. - The classified cluster image represents the classification of regions-of-interest in the 2-D histogram. Histogram regions-of-interest were identified using either an interactive, iterative k-nearest-neighbour classifier or a maximum-likelihood classifier.[26,29,30] Initially, the operator identified and labelled tissue regions-of-interest in either or both of the 3-D MR images. Regions-of-interest were selected from multiple locations to sample the intensity variations caused by RF nonlinearities (for example, the operator may have identified a number of small regions inside each of the ventricles, and labelled them as CSF, then identified and labelled a number of WM and GM regions in the frontal and occipital lobes and the corpus callosum. A 2-D histogram of the tissue regions-of-interest *only* were then calculated ar d displayed. For maximum-likeli-

hood classification, the mean and covariance of the region histogram were used to estimate the mean and covariance of the tissue distribution (for example, the mean and covariance of the ventricle region-of-interest were used to estimate the mean and covariance of the CSF cluster). Principle components[31] of the distribution estimate were calculated and used to define an elliptical region centered about the estimated mean. Prior knowledge about the tissue cluster was then incorporated into the estimate by an experienced operator who selected a confidence interval, and thus symmetrically extended or contracted the elliptical region threshold. For k-nearest-neighbour classification, the operator selected two thresholds. The first determined the percentage of "close" samples in each cluster for consideration when calculating distances between histogram locations and sample clusters. The second threshold defined the maximum distance between a histogram location and sample cluster beyond which the histogram location would not be considered for classification to the cluster.

Adjustment of the 2-D histogram classifications was possible by changing interactively either the original tissue regions-of-interest, the maximum-likelihood confidence interval, or the k-nearest-neighbour thresholds. Finally, the histogram regions-of-interest were assigned unique labels and colors. This procedure is illustrated in Figure 2.3 for classifying CSF using the k-nearest-neighbour classifier.

Classified Tissue Volume. - The classified tissue volume is a result of classifying tissues and regions-of-interest in the 3-D MR images. Tissues were classified by using the 2-D histogram regions-of-interest described above to color and label the corresponding MR voxels. Regions-of-interest within classified tissues were identified by labelling physiologically related connected components.[32] The classified tissue volume allows regions to overlap within a single voxel permitting partial volume[33] representation. Simultaneous display of both the 3-D MR image and the classified tissue volume is accomplished by assigning unique colors to each

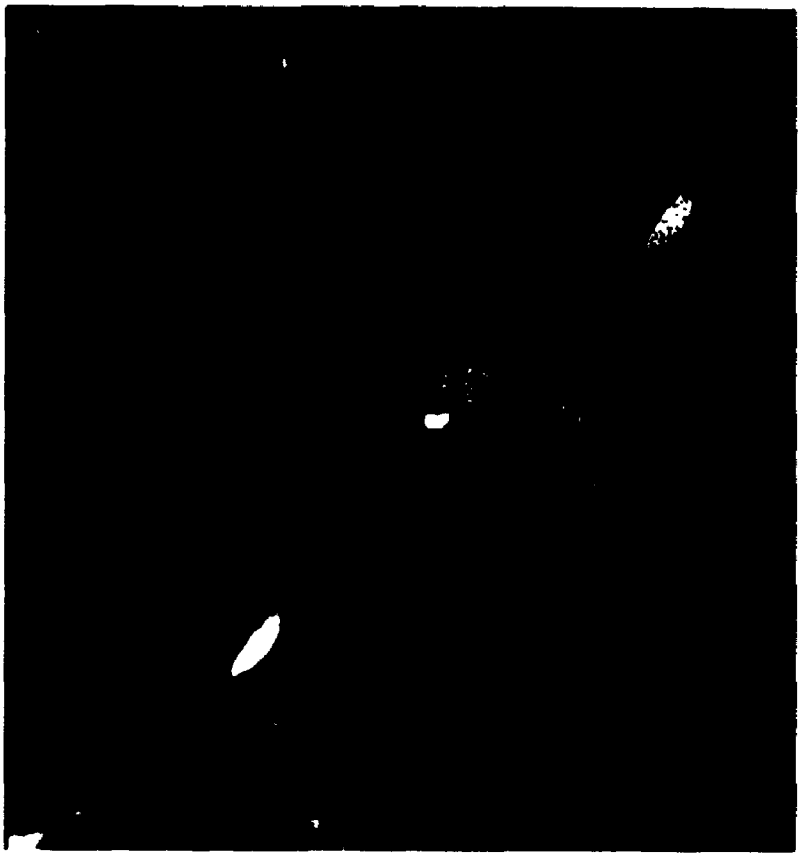


Figure 2.3 Classification of CSF

Classification of CSF involves 4 steps: a) The operator uses the labelling function to identify two sets of sample regions. The first is located inside the ventricles, and the second in WM and GM. Each set of samples is assigned a unique color, and named respectively "CSF" and "brain"; b) the 2-D histogram distributions of each set of samples is computed, labelled, and displayed; c) the operator selects the k-nearest-neighbour cluster classification algorithm, which then partitions the 2-D histogram into regions corresponding to "CSF" and "brain"; d) the classified 2-D histogram regions are used to color and name corresponding voxels in the 3-D MR image.

classified tissue and regions-of-interest and then mapping them onto the 3-D images. An example of this procedure is illustrated in Figure 2.3.

Labelling. - Regions-of-interest within anatomical structures were identified using connected component labelling.[32] This was accomplished by selecting a seed location, and initiating 2-D or 3-D region growing based on a prior classification (for example, finding all connected voxels which belong to CSF), or based on a combination of PD and T2 intensities (for example, finding all connected voxels within an elliptical region centered about the seed point's intensity combination in the 2-D histogram).

Analysis. - Once image regions-of-interest were identified, a number of morphological and statistical parameters for each, such as centre of mass, average intensity, intensity variance, and volume, were determined.[34] An ellipse bounding the 2-sigma confidence interval of each tissue region-of-interest in the histogram

was plotted and recorded. Additionally, each 2-D histogram region-of-interest was analyzed to determine its mean, major axis, minor axis, and variance. In some situations English-like relational queries were used to relate MR characteristics, spatial location, labelling, and histogram distribution of different MR voxels. For example, the simple query "CSF and Lesion" was constructed to highlight voxels where the two regions-of-interest respectively labelled "CSF" and "Lesion" overlapped. Iterative modification of these regions was then performed to eliminate the overlap.

Visualization. - The 3-D MR images and the tissue classification volumes were exported to commercially available software for 3-D visualization (SunVision visualization package, Advanced Visual Systems, Waltham, MA.). Multiple renderings from different views were obtained to view the classified volume and 3-D MR image from different orientations. A sequence of these were then played back in a cine loop to provide animated views of the data. This dynamic visualization of the data revealed spatial and anatomical information not readily apparent in static displays.

2.2.2 System Validation

Since the slice thickness can range from 3mm to 5mm, partial volume effects may cause errors in tissue classification and volume measurement. Therefore, I constructed a phantom to determine the effect of slice positioning and thickness on the accuracy and precision of my system in measuring the volumes of "small" objects (0.4 to 6.8 cm^3) under ideal conditions. My phantom consists of 1% weight per volume agarose "lesions" doped with 0.5% graphite surrounded by 1% weight per volume agarose "brain".[14] Thirty seven "lesions" were constructed by forming a layer of uniform thickness doped agarose and then cutting it to create regular "lesions" with similar in-plane dimension (1.2 cm by 1.2 cm) but inter-plane thickness varying from 0.2 to 5.3 cm. To determine if systematic errors occurred in measuring volume, additional "lesions" with differing in-plane and inter-plane

dimensions were constructed with volumes near those of a number of regular "lesions". The volume of each "lesion" was determined by measuring its water displacement to a precision of ±0.05 cm^3.

My phantom was imaged 11 times using the image acquisition protocol described above. During the first 5 exams, 29 adjacent, non overlapping 5mm thick sagittal slices were acquired. Slice positioning was incremented by 1 mm between examinations to cover the range of possible positioning. During the last 6 exams, 49 adjacent, non overlapping 3mm thick sagittal slices were acquired. Slice positioning was incremented by 0.5 mm between examinations to cover the range of possible positioning.

All the resulting 3-D images were classified using my system. For each exam, the operator selected sample "lesion" voxels, and "brain" voxels. These samples were then used to estimate the means of the "lesion" and "brain" 2-D histogram distributions. A nearest-neighbour classifier[29] was then used to differentiate the "lesion" and "brain" distributions in the histogram and create corresponding histogram regions-of-interest. These were then used to classify the 3-D MR images after which individual "lesions" were labelled, and their volumes determined automatically and compared to their actual volumes for both 3mm and 5mm slices. The error, ε_i, in the volume measurement of "lesion" i was determined according to Equation 2.1 and plotted for both 3mm and 5mm slices.

$$\varepsilon_i = \bar{M}_i - A_i \tag{2.1}$$

where $\bar{M}_i$ is the average measured volume of "lesion" i, and A_i is the actual volume of "lesion" i. The average absolute error of all "lesion" volume measurements, ε_{abs}, was determined according to:

$$\varepsilon_{abs} = \frac{\sum_{i=1}^{N_L} \sum_{j=1}^{N_E} |M_{ij} - A_i|}{N_L \cdot N_E} \qquad (2.2)$$

where N_L is the number of "lesions", N_E is the number of exams, M_{ij} is the measured volume of "lesion" i in exam j, and A_i is the actual volume of "lesion" i.

The "tissue" classification volume was exported for visualization using the techniques described above. Figure 2.4 shows two 3-D views of the rendered "tissue" classification.

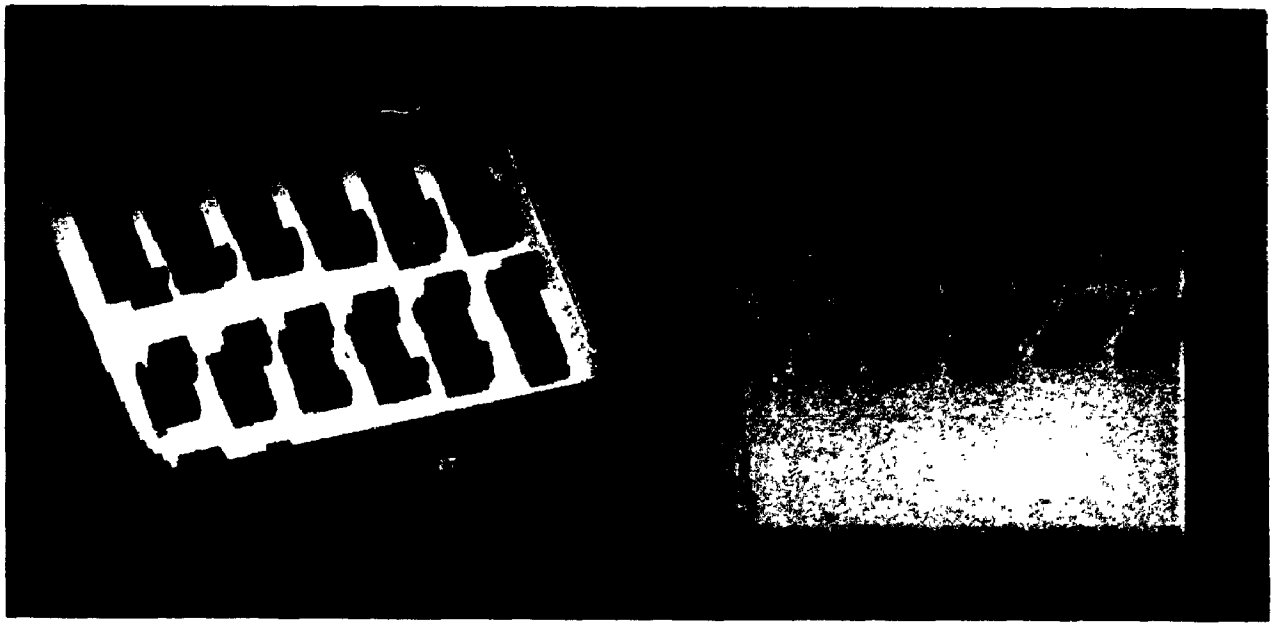


Figure 2.4 The volume rendered "lesion" phantom

Two views of the volume rendered "lesion" phantom constructed to determine accuracy and precision of volume measurement of small objects under ideal conditions. The phantom consists of agarose "brain" (rendered grey and translucent) imbedded with 37 doped agarose "lesions" (rendered red and opaque). This phantom was imaged 11 times using a standard MS protocol. Slice positioning was incremented between exams to cover the range of possible positioning.

2.2.3 Quantitative MS Lesion Volume Analysis

My system was used to classify and analyze four MR exams of a clinically definite, chronic-progressive MS patient taken over an 18 month period. Each exam was obtained using the image acquisition protocol described above, except that 26 oblique axial 5mm thick slices were obtained in each instance. Slice positioning between exams was maintained to within one slice thickness by first obtaining a sagittal "scout" view, and then identifying the anterior and posterior extent of the corpus callosum, and using these to establish slice position, and orientation.[35]

As an internal control, measurements of the ventricular volume in this patient from the four exams over 18 months were obtained by using a procedure similar to that described by Kikinis *et. al.*[13] For each exam, sample regions-of-interest inside the third and lateral ventricles were identified, and labelled as CSF. In addition, sample regions-of-interest inside cerebral white and grey matter were identified and labelled as WM/GM. Samples were used by the k-nearest-neighbour classifier described above to estimate the 2-D histogram distributions of CSF and WM/GM. Exam voxels with PD/T2 combinations falling within the estimated CSF histogram distribution were then classified as CSF in the 3-D MR images, as shown in Figure 2.3. The third and lateral ventricles were isolated, labelled, and their combined volume calculated. To determine the effect of the choice of sample regions-of-interest, the k-nearest-neighbour classification was repeated 5 times, each with different initial regions-of-interest.

Analysis of the 2-D histogram from an entire patient exam revealed six distinct clusters, five of which were also observed by Fletcher et. al.[23] Each was identified, labelled and then used to label the corresponding tissue structures in the PD and T2 volumes so that anatomical correspondence, if any, could be established. In Figure 2.5a I show the classified 2-D histogram and in Figure 2.5b I show a single slice through the resulting classification.

Figure 2.5 2-D histogram clusters and the corresponding regions in the MR slice.

The histogram and segmentation of the PD and T2 images shown in Figure 2.2. a) is the 2-D histogram of these slices. The colored regions in this histogram are defined interactively by the user and correspond to tissues in the images with similar PD and T2 weighted intensities. The yellow region corresponds to muscle and skin, the purple region to adipose tissue, the blue region to CSF, the green region to GM, the pink region to WM and the red region to MS lesion. b) shows the resulting segmentation of the slice. My system allows the user to change interactively the size, shape and location of the regions of interest until the resulting segmentation accurately delineates tissues of interest in the input images.

To determine whether there were differences in the 2-D histograms of slices with and without lesions, I determined the 2-D histograms from slices which were identified by a neuroradiologist as containing lesions and also identified with no lesions. Two such 2-D histograms are shown in Figure 2.6. To determine if differences exist in the 2-D histogram distributions of CSF, WM, GM and MS lesion in these slices, approximately 200 sample voxels from each of these tissues were identified and labelled. The 2-D histogram distribution of each tissue sample was then calculated, and the mean, magnitude of the principle components and orientation of the first principle component determined in each case. To determine the effect of sample selection on these measurements, 5 different sets of sample vox-

els were identified and labelled for each tissue. The results of all these measurements are listed in Table 2.1

Tissue	Mean		2σ along 1st PC	2σ along 2nd PC	Angle of 1st PC
	PD Weighted	T2 Weighted			
Lesion	172.4 ± 1.5	178.8 ± 2.4	50.8 ± 8.4	9.8 ± 2.8	59.0 ± 5.4
CSF	132.6 ± 1.3	176.4 ± 2.1	26.0 ± 5.1	8.2 ± 0.8	52.2 ± 4.7
GM	151.8 ± 2.1	137.6 ± 2.9	31.8 ± 9.0	8.6 ± 1.8	44.4 ± 3.2
WM	134.8 ± 0.8	108.6 ± 1.5	25.4 ± 8.0	9.6 ± 0.6	44.2 ± 4.6

Table 2.1 Data from the 2-D histogram distributions of CSF, MS Lesion, White Matter and Grey Matter.

Values given (in arbitrary units) are the mean, two standard deviations in the direction of the principle components, and orientation of the 1st principle component. Values are the average of 5 sets of samples, each consisting of approximately 200 sample voxels from each tissue. PC = principle component, GM = gray matter, WM = white matter.

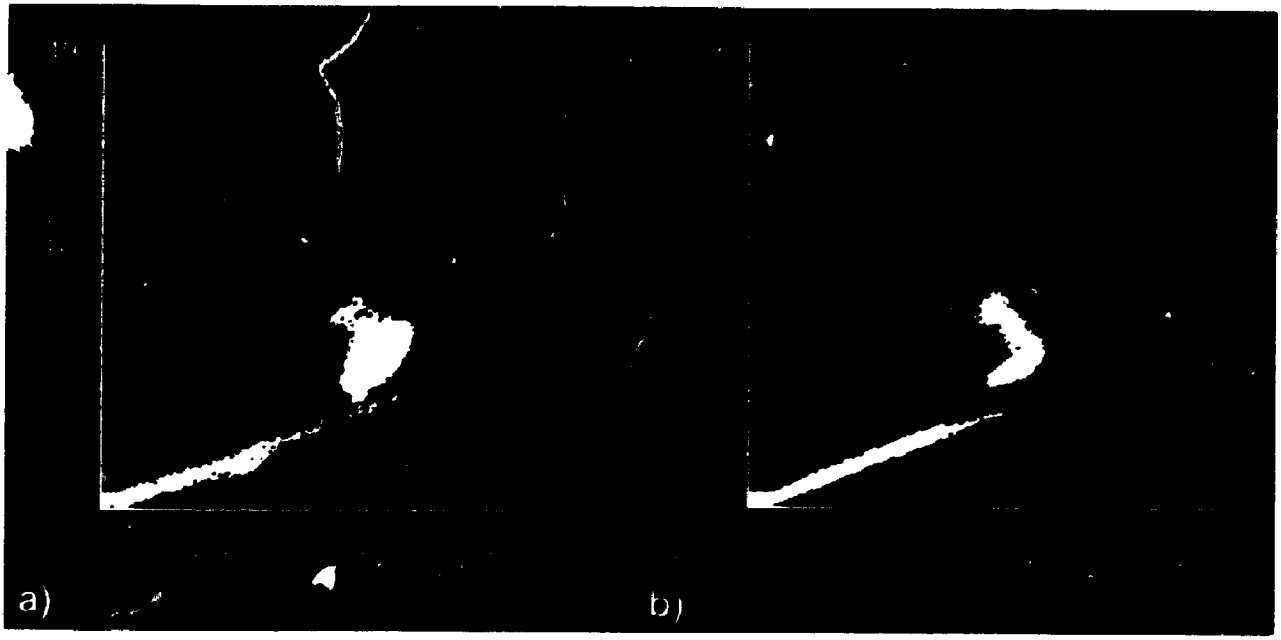


Figure 2.6 2-D histograms from slices with and without MS lesions.

a) Histogram of the PD and T2 weighted images from a slice with lesions. b) Histogram of a slice from the same exam, but without lesions. A distinctive peak indicated in the histogram of the slice with MS lesions is missing in the histogram of the slice without lesions. The location of this peak suggests that the corresponding image voxels represent edema.

My system was used to measure volume changes in MS lesions observed in the 4 exams over the 18 month period. A neuroradiologist identified the location and extent of every lesion in each exam along with regions of typical WM, GM and CSF. In addition, the histogram cluster corresponding to all identified lesions, WM/GM, and CSF were calculated, and principle component analysis performed to determine the ellipses bounding the 2-sigma confidence interval of each distribution. Lesion boundaries were then outlined manually using my system, and four "large" lesions (greater than 0.5 cm^3) were selected for analysis. These lesions were isolated, labelled, and their volumes calculated in each of the 4 patient exams. 3-D MR images and classified tissue volumes of the four exams were exported for visualization using the visualization techniques described above.

2.3 Results

2.3.1 System Validation

Figure 2.7a shows the accuracy of volume measurements of the "lesions" in my phantom from 0.4cm^3 to 6.8cm^3 made from exams with 5mm thick slices, and Figure 2.7b shows the accuracy of volume measurements made from exams with 3mm thick slices. Measurements from exams with 5mm thick slices had an average absolute error of 0.21cm^3 (Equation 2.2), while those from exams with 3mm thick slices had an average absolute error of 0.15cm^3. However, acquisition time for exams with 3mm thick slices were approximately 18 minutes each, almost double that of exams with 5mm thick slices.

Figure 2.8a (on page 41)shows the error, ε_i, given by Equation 2.1 for measuring "lesion" volumes from exams with 5mm thick slices. Figure 2.8b shows the error of the measurements made from exams with 3mm thick slices. These graphs show that the error in the volume measurement is essentially constant for all "lesion" volumes and that the greatest deviation from the actual volume is

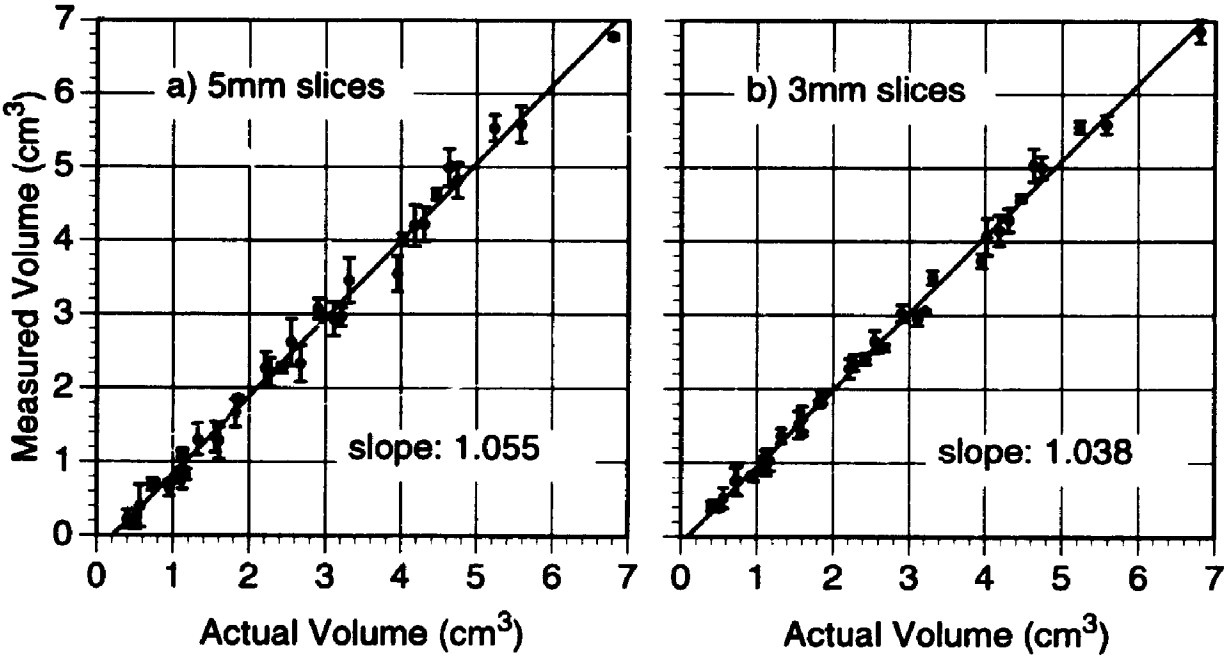


Figure 2.7 The accuracy of phantom "lesion" volume measurement.

a) a graph showing the accuracy of "lesion" volume measurements made from exams with 5mm thick slices. b) a graph showing the accuracy of "lesion" volume measurements made from exams with 3mm thick slices. The points in each graph represent the average measured volume for a particular "lesion" over all slice positioning for a particular slice thickness, and the error bars represent ± one standard deviation. In each graph, the line of best fit has slope very near to 1.

about 0.3 cm^3. This implies that the relative error of "lesion" volume measurement will decrease as "lesion" volume increases, that is, the precision of measurement will improve as the "lesions" get larger.

Figure 2.9a (on page 42)shows the precision of volume measurement (defined as the standard deviation in the volume measurements divided by the actual volume) of "lesions" from exams with 5mm thick slices, while Figure 2.9b shows the precision of volume measurements from exams with 3mm thick slices, each plotted on a log-log graph. Generally, the precision improves as "lesion" volume increases. Exams with 5mm thick slices provided "lesion" volume measurements with precision ranging from 0.5% (actual "lesion" volume 6.8 cm^3) to 51.4% (actual "lesion" volume 0.56 cm^3), with 10% precision or better for "lesion" volumes greater than 1.05 cm^3. Exams with 3mm thick slices provided "lesion" vol-

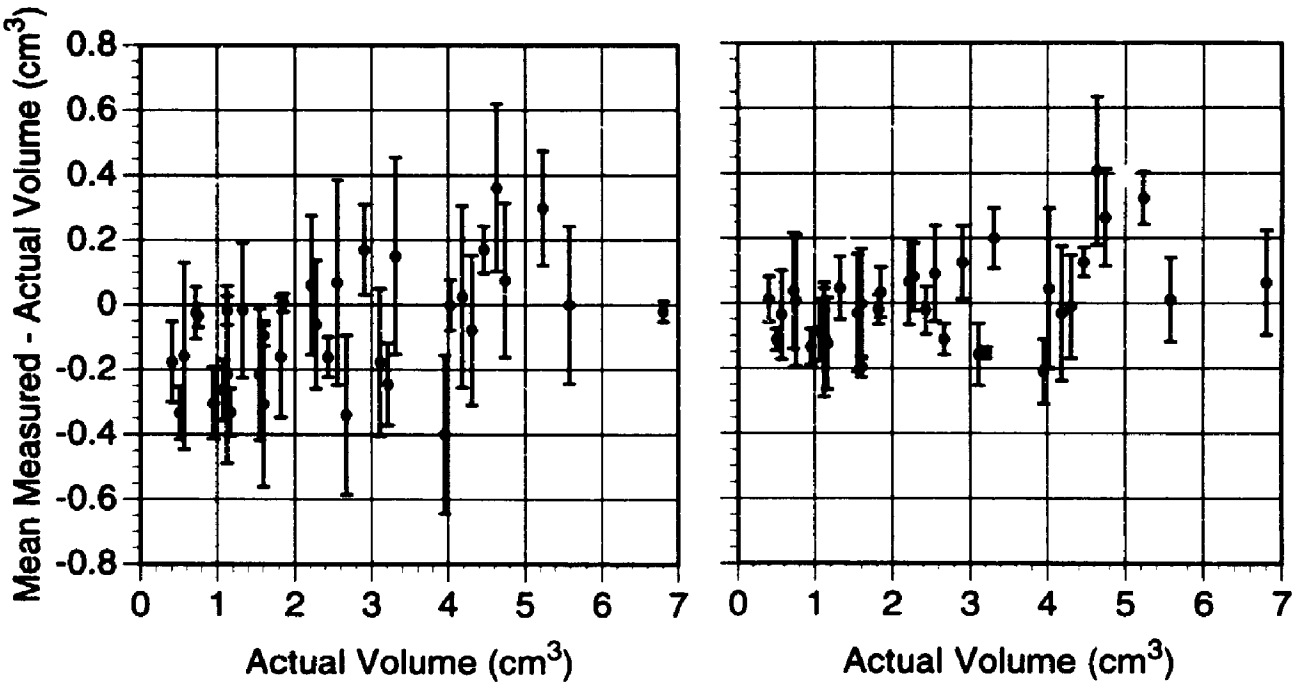


Figure 2.8 The errors in phantom "lesion" volume measurement.

a) the errors, defined as the mean measured "lesion" volume less the actual "lesion" volume, made from exams with 5mm thick slices. b) the errors made from exams with 3mm thick slices. The error bars in these graphs represent ± one standard deviation in the volume measurements. Comparison of these two graphs reveals that volume estimations made from exams with 3mm thick slices were more accurate than those made from exams with 5mm thick slices.

ume measurements with precision ranging from 0.5% (actual "lesion" volume 3.2 cm^3) to 27.0% (actual "lesion" volume 0.75 cm^3), with 10% precision or better for "lesion" volumes greater than 0.9cm^3.

2.3.2 Patient Studies

Ventricular volume. - Figure 2.10 (on page 42)shows the average measured volume of the third and lateral ventricles from the 4 MR exams over the 18 month period and the line of best fit through the data. The slope of the line of best fit through the data is 0.0004, indicating that the ventricle volume in this patient remained constant through the exam period.

2-D Histogram analysis. - Figure 2.5 (on page 37)shows the classification of 6 clusters in the 2-D histogram, each labelled with a unique name and color. In par-

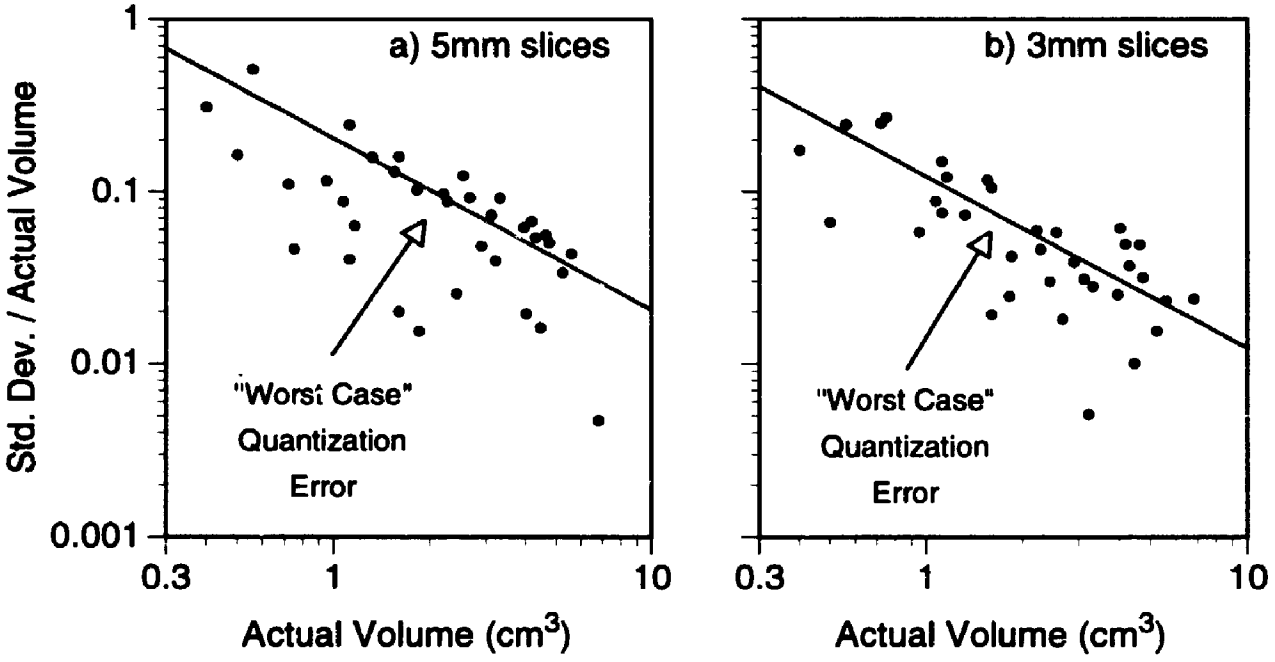
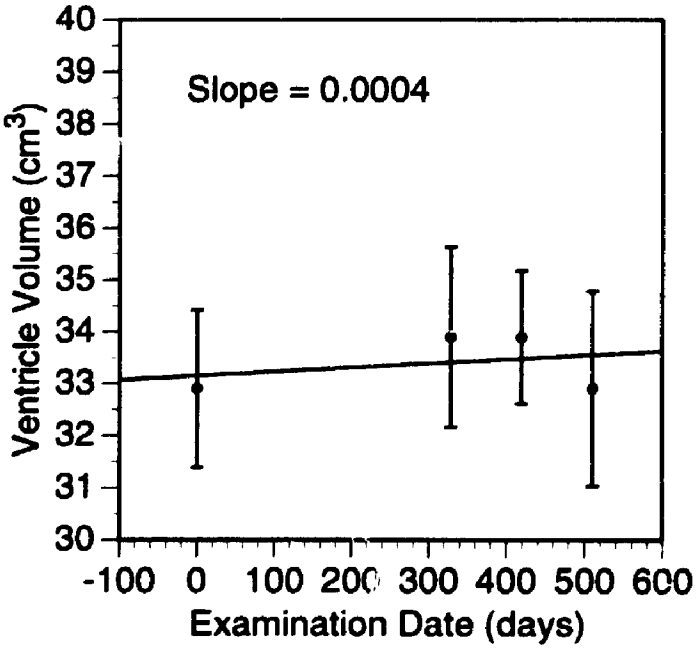


Figure 2.9 Precision as a function of phantom "lesion" volume.

a) the relationship between precision of "lesion" volume measurements made from exams with 5mm thick slices, and actual "lesion" volume. b) the relationship between precision of "lesion" volume measurements made from exams with 3mm thick slices, and actual "lesion" volume. Generally, precision improves as "lesion" volume increases. The variation in "lesion" volume measurement can be explained as a variation in "lesion" quantization due to slice positioning (see Discussion). In each graph the solid line represents the "worst case" quantization error.

Figure 2.10 The average ventricle volume of an MS patient over an 18 month period.

A plot of the average ventricle volume in the MS patient over the exam period. The points represent the average measured ventricle volume for each exam, and the error bars represent ± one standard deviation. The slope of the line of best fit is very nearly 0, (R^2 = 0.97) indicating that the ventricle volume remained essentially constant during this period.



ticular, muscle and skin are colored in yellow, adipose tissue in purple, CSF in blue, WM in pink, GM in green, and lesion in red. This visualization is useful for determining the anatomical significance of various clusters in the 2-D histogram. In addition, the narrow, highly oriented nature of various clusters suggests an interdependence between the ordinate and abscissa.

Figure 2.6a (on page 38)shows a histogram of a slice which contained neuroradiologist identified lesions, while Figure 2.6b shows the histogram from a slice without lesions. A distinctive cluster extending from the WM/GM cluster, but with a narrower, more linear distribution, is observed *only* in histograms of slices with lesions (Figure 2.6a). Table 2.1 (on page 38) lists the average values for the distribution means, 2 standard deviations in the direction of each principle component, and orientations of the first principle component from 5 sets of samples of WM, GM, CSF, and MS lesion. Figure 2.11 (on page 44) shows the 2 standard deviation elliptical regions determined from the values in Table 2.1. The location and shape of the MS lesion distribution in Figure 2.11 indicates that the lesions are responsible for the distinctive peak in the 2-D histogram. This observation has been confirmed using the labelling and analysis functions of my system to label the MR image locations corresponding to the distinctive peak in the 2-D histogram, and then comparing those to the lesions indicated by the neuroradiologist.

Lesion volume analysis. - Figure 2.12 (on page 45) shows axial views of a volume rendering of the classified MS patient exams. Figure 2.12a shows four lesions, labelled A, B, C, and D, along with CSF and WM/GM at day 0, while Figures 2.12b- d show the changes which have occurred by days 297, 388 and 486, respectively.

Although these volume renderings are informative and useful in visualizing and monitoring changes in lesions, they are not quantitative. Measurement of lesion volumes at each exam date are shown in Figure 2.13 (on page 46). At Day

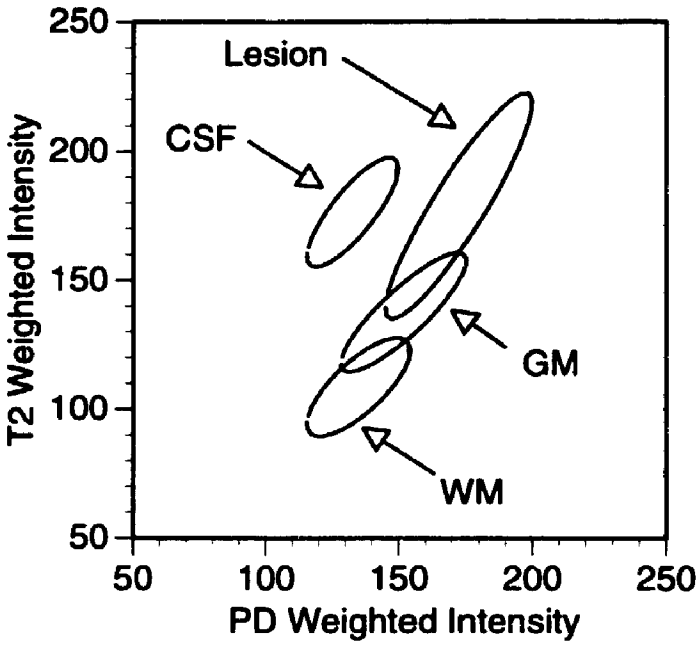


Figure 2.11 The 2-D histogram distributions of CSF, white matter, grey matter and MS lesion.

A plot of the 2 sigma ellipses bounding the 2-D histogram distributions of WM, GM, CSF and MS lesion, determined from the values in Table 2.1. These 4 tissues may be categorized into 2 groups based on the angle of their principle axis with the horizontal axis: 1) WM and GM; and 2) CSF and MS lesion.

0 each lesion is distinct. However, by Day 297 lesions A and C have merged forming one large lesion which is slightly larger than the combined volumes of lesions A and C on Day 0. The volume of lesion D has remained constant, while the volume of lesion B has increased. By Day 388 the combined lesion A/C has increased in volume again. Lesion B has decreased slightly in volume, while lesion D has increased slightly. By the last exam date (Day 486) lesions A and C are distinct once more. Their separation is likely the result of a dramatic decrease in the volume of lesion A, which is much smaller at Day 486 than at Day 0. Lesions B and D have combined to form one large lesion, whose volume is larger than the combined volumes of the distinct lesior ˛ at Day 388. From Figure 2.12d it is apparent that a large increase in lesion B has resulted in the merger of lesions B and D.

Both the rendered tissue classification and volume measurements reveal the dynamic nature of many MS lesions, which can increase, decrease or remain static. These results in turn support observations from other investigators who have reported variation in lesion volumes.[2-6,10,12]

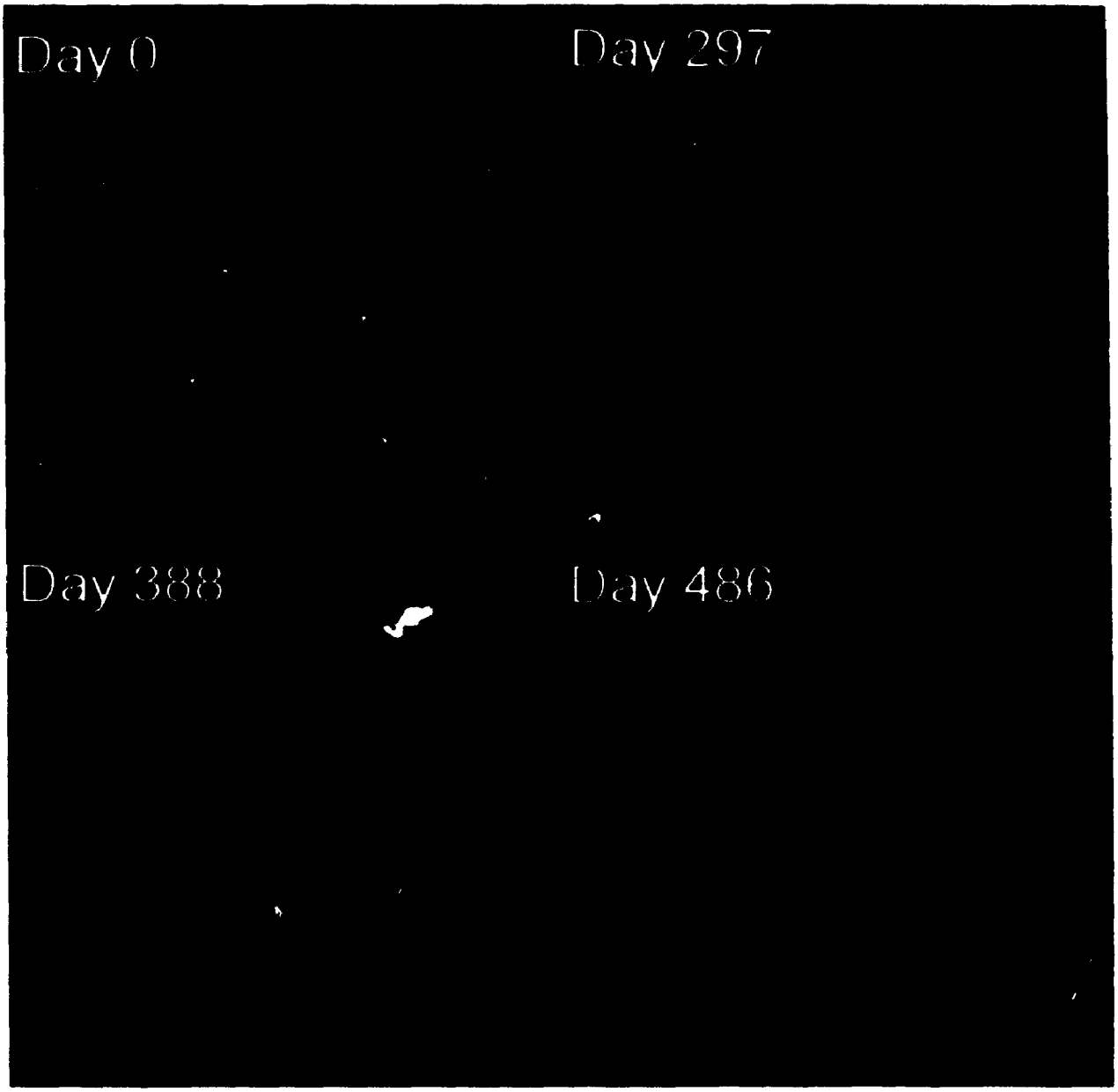


Figure 2.12 Volume renderings of the changes in 4 lesions from a chronic-progressive MS patient over a 486 day period.

The view is axial, looking from inferior towards superior, with the patient facing up. a) The four lesions, labelled A,B,C and D are opaque and rendered in green, yellow, white and red respectively. Brain and CSF are highly translucent and rendered in grey and blue respectively. Initially, all four lesions are isolated. b) Day 297. Lesions A and C have become connected by a thin bridge on the superior side of the ventricle. Lesion B has increased in size, particularly in the extreme posterior region. Lesion D appears relatively unchanged. c) Day 388. Lesion A appears to have decreased in size slightly. The bridge connecting lesions A and C has also thinned. Lesions B and D have similar appearance to day 297. d) Day 486. The bridge connecting lesions A and C has disappeared. Lesion A appears to have decreased dramatically in the immediate periventricular region. Lesion C has a similar appearance to day 0. Lesion B meanwhile, has increased in the posterior and periventricular region. The latter increase has caused a merger with lesion D.

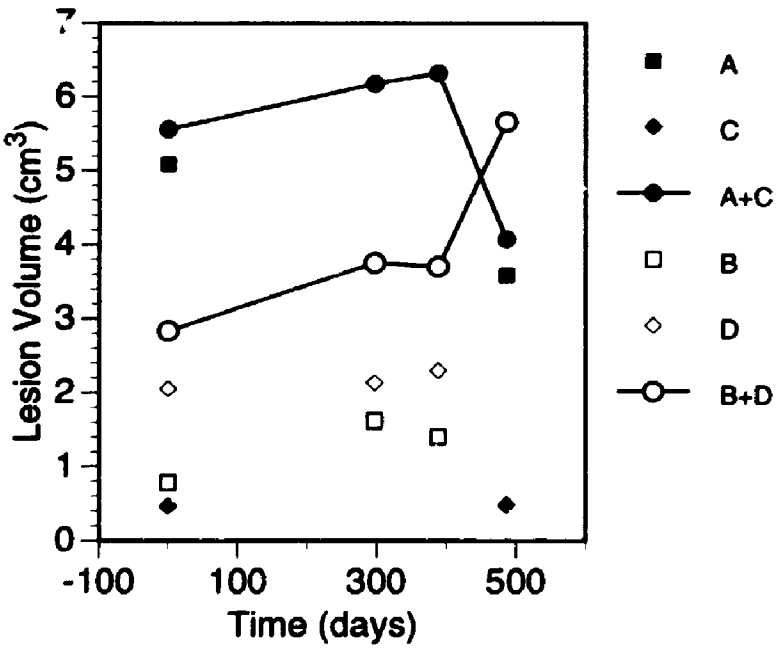
2.4 Discussion

2.4.1 System Validation

A number of investigators have measured the accuracy of volume estimation from MR images of phantoms and human subjects.[11,13,14] The accuracy of volume estimation shown in Figure 2.7 and Figure 2.8 reveals that under ideal conditions, volumes of regular, high contrast objects can be estimated using my system with an accuracy of better than 0.3 cm^3. This accuracy is independent of lesion size (Figure 2.8) and thus a plot of the precision of the measurements (Figure 2.9) reveals that the precision improves as the volume of the object being measured increases. Additionally, this improvement in precision occurs for measurements made from exams with both 5mm and 3mm thick slices.

Figure 2.13 Volume changes of the four lesions shown in Figure 2.12 over the 486 day period.

The upper solid line shows the combined volume for lesions A and C, while the lower line shows the combined volume of lesions B and D. Individual lesion volume measurements are shown with symbols. Volume measurements of lesions A and C are represented with closed symbols, while those of lesions B and D are represented with open symbols. This graph reveals the dynamic nature of some MS lesions, which may expand and contract over time.



In order to explain the relationship between precision and "lesion" volume, I consider that the "lesion" volume is composed of columns, each with cross-sectional area given by the in-plane pixel dimensions, and height equal to the "lesion" thickness. In the examination procedure, the "lesion" columns are imaged with

contiguous slices which are either 3mm or 5mm thick. The volume of each "lesion" is then estimated by summing the volumes of the exam voxels which cover the "lesion" and are classified as containing "lesion". Slice positioning will affect the manner in which the "lesion columns" are sampled, and thus the estimation of "lesion" volume. The variance in "lesion" volume estimation will therefore depend on the "quantization" of the top and bottom of each "lesion". Thus, the variance in "lesion" volume measurement will be given by the sum of the variances due to quantization of the top and bottom of each "lesion", less the covariance between these two quantizations, i.e.:

$$\sigma^2 = \sigma_t^2 + \sigma_b^2 - C[\sigma_t^2,\sigma_b^2] \tag{2.3}$$

where σ_t^2 and σ_b^2 are, respectively, the variances in the quantization at the top and bottom of a "lesion", and $C[\sigma_t^2,\sigma_b^2]$ is the covariance between these quantizations.

The quantization variance for a given slice thickness (bin size), T, is given by[36]:

$$\sigma_t^2 = \sigma_b^2 = \frac{T^2}{12} \tag{2.4}$$

If we assume the worst case, i.e. that there is no correlation between the quantization error at the top and bottom of each "lesion", then:

$$C[\sigma_t^2,\sigma_b^2] = 0 \tag{2.5}$$

and

$$\sigma^2 = \frac{2T^2}{12} \tag{2.6}$$

Therefore, the precision of volume estimation is:

$$\frac{\sigma}{V} = \sqrt{\frac{2}{12}}\frac{T}{V} \tag{2.7}$$

The solid lines in Figure 2.9 are a plot of $\frac{\sigma}{V}$, given in Equation 2.7, as a function of V. From Figure 2.9 it is apparent that the data generally follow the behavior predicted in Equation 2.7. However, noise in the examination procedure degrades the precision of volume measurements, causing some points to fall above each line. There will also, in general, be some degree of correlation between the quantization errors at the top and bottom of each "lesion" since the top and bottom are from the same "lesion". Thus, $C[\sigma_t^2,\sigma_b^2]$ in Equation 2.4 may be greater than 0. In these situations, the precision of volume measurement will be improved, resulting in points below the lines in Figure 2.9.

Equation 2.7 suggests that the precision of "lesion" volume measurement should improve linearly with decreasing slice thickness and that thinner slices should be used. However, physical limitations in the design and construction of the MR scanner prevent significantly thinner slices being used with this imaging protocol. Furthermore, if all other factors remain constant, then a decrease in slice thickness will cause a corresponding decrease in signal-to-noise ratio. One manner to overcome these limitations is to use 3 dimensional fast spin-echo imaging techniques. With these techniques, an increase in the volume from which signal is measured allows smaller, isotropic voxels to be acquired in the same time as 2-D techniques, while maintaining the signal-to-noise ratio. Currently protocols are being developed in the Imaging Research Laboratories which utilize 3-D fast spin-echo imaging techniques, for use in the examination of MS patients.[37]

2.4.2 Patient Studies

Ventricular volume. In the 4 MS patient exams obtained with 5mm slices, the ventricles are large and fairly "regular". However, the relatively coarse quantization in the slice direction means that they are present in only 8 slices in each

exam. From the results of the phantom experiment shown in Figure 2.9, I expect that the coarse quantization of the ventricles in the slice direction is a major source of the variation in ventricle volume measurement. Nevertheless, the standard deviation in these measurements, as shown in Figure 2.10, ranged from 3.9% to 5.7% of the mean measured ventricle volume, with an average of 4.8%. These values are in accordance with those reported by other researchers,[13,14,16,17] and suggest that coarse quantization in the slice direction is the dominant factor determining the variability of ventricle volume measurement over the exam period.

Variability in volume measurement may also result from other factors. In particular, hardware and software changes, along with recalibration of the imaging system will affect image intensities, and may therefore impact tissue classification. However, as long as sufficient contrast is maintained between tissues of interest, manual delineation of tissue boundaries and identification of corresponding clusters in the 2-D histogram should still be possible, though the variability of tissue quantification will increase as tissue contrast decreases. Therefore, an important consideration in the design of any clinical trial based on the identification and quantification of regions-of-interest in MR images will be the careful calibration of MR images over time to provide improved accuracy and precision of region quantification.

Calibration may be accomplished by imaging a phantom composed of materials which mimic the MR characteristics of the tissues-of-interest before each patient exam, then correcting for phantom image variations. Alternatively, it may be possible to employ some internal standard as the basis for calibration. Neither of these techniques were employed on the patient exams analyzed for these experiments since my objective was to determine the system's potential clinical application, and not its performance as the standard in a clinical trial over an extended period involving many patients.

2-D Histogram analysis. The narrow, linear distribution of some clusters (adipose tissue, muscle and lesion) shown in Figure 2.5 suggests that it may be possible to characterize the corresponding tissues by a single parameter, which relates in a linear manner the ordinate (PD weighted intensity) and abscissa (T2 weighted intensity). Comparison of the 2-D histogram distributions shown in Figure 2.6 suggests that MS lesions have a higher PD and longer T2 than WM or GM, and a higher PD than CSF. These observations in turn suggest that some lesions have higher levels of unbound water,[38] and may represent regions of edema. Examination of the angles of the first principle components of the WM, GM, and lesion distributions suggests that although their intensities may be similar they represent different linear combinations of PD and T2 weighted intensities. However, comparison of these angles for the CSF and lesion distributions suggests that although their intensities are different, they represent similar linear combinations of PD and T2 weighted intensities.

Lesion volume analysis. Figure 2.12 shows a 3-D volume rendering of 4 MS lesions at each exam date, while Figure 2.13 shows the volume of these lesion over the exam period. These visualizations and measurements are a result of manually outlining previously identified lesions in multiple slices, in each exam. Manual outlining is a time consuming procedure (requiring approximately 30 minutes per exam) and is prone to inter and intra-observer variability. A solution to this problem is to employ purely automated techniques for lesion classification. However, I believe that the heterogenous composition of the lesions, highly irregular shape, small size, wide range of intensity distribution in the 2-D histogram, and overlap of intensity distribution with WM and GM (Figure 2.11) make a purely automated classification technique difficult. Another approach is to use computer assisted techniques, which have the advantage of employing human expertise while identifying lesion boundaries, but which use automated techniques to provide lesion volume measurements, and thus improve the precision of those measurements. The effect of computer assisted techniques on the accuracy of lesion

volume measurements is difficult to establish however, since the lesion volumes are rarely known. Furthermore, my phantom experiments indicate that measurements of the volume of regular, high contrast objects under ideal conditions using a standard MS protocol can have significant errors, especially for objects less than 0.5 cm^3 (Figure 2.9). Volume measurements of actual MS lesions with lower contrast and irregular borders are likely to be less accurate.

I believe that two approaches are needed to improve the accuracy and precision of the analysis of MS lesions in MR exams. The first is continued development of computer assisted techniques for MR exam analysis, with special emphasis on the identification of MS lesions. The second is to employ improved MR imaging protocols, 3-D spin-echo techniques in particular, to provide better fundamental data for analysis. Additional information may also be gained by acquiring three or more echoes, perhaps weighted by other parameters such as T1, diffusion or magnetization transfer. The additional echoes may be used then to extend the 2-D histogram to higher dimensions, and thus potentially increase cluster separability.

2.5 Conclusions

I have developed a computer assisted system to aid in the identification and analysis of MS lesions in MR PD and T2 weighted volumes of the head. My system is based upon four data structures: 1) a 3-D MR image; 2) classified tissues in the MR image; 3) a 2-D histogram of the PD and T2 weighted intensity combinations in the MR image: and 4) classified clusters in the 2-D histogram. It provides functions to label, analyze and visualize each of these data structures.

Application of my system to the classification and analysis of MR exams of phantom containing regular, high contrast "lesions" reveals that: a) accurate measurements of objects between 0.4 cm^3 and 6.8 cm^3 are possible; b) precision

improves as "lesion" volume increases, generally being better than 10% for "lesions" larger than 1 cm^3; c) variation in "lesion" volume measurement is largely due to coarse quantization of the "lesions" in the slice direction; and, d) an estimate of the quantization error predicts the uncertainty in the volume estimation.

My system was also used to classify and analyze four MR exams of a clinically definite, chronic-progressive MS patient taken over an 18 month period. My results indicate: a) the ventricle volume in this patient remained constant during the exam period; b) six distinct clusters corresponding to white matter, grey matter, cerebrospinal fluid, MS lesion, muscle and adipose tissue, are present in the 2-D histograms of the MR exams; c) a distinctive peak is present only in the 2-D histograms of slices with MS lesions; d) tissues may be differentiated based on their intensity distributions in the 2-D histogram, and the slope of the first principle component of their histogram distribution; e) my system can be used to visualize and quantify the 3-D changes of MS lesions over time; f) lesions in the chronic-progressive MS patient I examined have a dynamic nature, growing and shrinking over an 18 month period.

2.6 References

[1] S. Wiebe, D.H. Lee, S.J. Karlik, et. al., "Serial cranial and spinal cord magnetic resonance imaging in multiple sclerosis". Annals of Neurology, 32, 643-650 (1992).

[2] A.J. Thompson, A.G. Kermode, D.G. MacManus, et al., "Patterns of disease activity in multiple sclerosis: clinical and magnetic resonance imaging study". British Medical Journal, 300, 631-634 (1990).

[3] J.O. Harris, J.A. Frank, N. Patronas, et al., "Serial gadolinium-enhanced magnetic resonance imaging scans in patients with early, relapsing-remitting multiple sclerosis: implications for clinical trials and natural history". Annals of Neurology, 29, 548-555 (1991).

[4] L. Kappos, D. Stadt, M. Ratzka, et al., "Magnetic resonance imaging in the evaluation of treatment in multiple sclerosis". Neuroradiology, 30, 299-302 (1988).

[5] C.J. Wallace, P. Seland, T. Chen Fong, "Multiple sclerosis: the impact of MR imaging". American Journal of Radiology, 158, 849-857 (1992).

[6] D.H. Miller, F. Barkhof, I. Berry, et. al., "Magnetic resonance imaging in monitoring the treatment of multiple sclerosis: concerted action guidelines". Journal of Neurology, Neurosurgurgery and Psychiatry, 54, 683-688 (1991).

[7] B.G. Weinshenker, B. Bass, S. Karlik, C.G. Ebers, G.P.A. Rice, "An open trial of OKT3 in patients with multiple sclerosis". Neurology, 41, 1047-1052 (1991).

[8] J.R. Mitchell, S.J. Karlik, D.H. Lee, A. Fenster, "Automated detection and quantification of multiple sclerosis lesions in MR volumes of the brain". SPIE Medical Imaging, 1652, 99-106 (1992).

[9] J.R. Mitchell, S.J. Karlik, D.H. Lee, A. Fenster, "Multi-spectral analysis and visualization of multiple sclerosis lesions in MR volumes of the brain". SPIE Medical Imaging, 1898, 442-452. (1993).

[10] C. Isaac, D.K. Li, M. Genton, et. al., "Multiple sclerosis: a serial study using MRI in relapsing patients". Neurology, 38, 1511-1515 (1988).

[11] E.F. Jackson, P.A. Narayana, J.S. Wolinsky, T.J. Doyle, "Accuracy and reproducibility in volumetric analysis of multiple sclerosis lesions". Journal of Computer Assisted Tomography, 17(2), 200-205 (1993).

[12] R.A. Koopmans, D.K.B. Li, J.J.F. Oger, L.F. Kastrukoff, C. Jardine, L. Costley, S. Hall, E.W. Growchowski, D.W. Paty, "Chronic Progressive Multiple Sclerosis: Serial Magnetic Resonance Brain Imaging over Six Months". Annals of Neurology, 26, 248-256 (1989).

[13] R. Kikinis, M.E. Shenton, G. Gerig, J. Martin, M. Anderson, D. Metcalf, C.R.G. Guttmann, R.W. McCarley, W. Lorensen, H. Cline, F.A. Jolesz, "Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging". Journal of Magnetic Resonance Imaging, 2, 619-629 (1992).

[14] M.I. Kohn, N.K. Tanna, G.T. Herman, S.M. Resnick, P.D. Mozely, R.E. Gur, A. Alavi, R.A. Zimmerman, R.C. Gur, "Analysis of Brain and Cerebrospinal Fluid Volumes with MR Imaging". Radiology, 178, 115-122 (1991).

[15] O. Kübler, G. Gerig, "Segmentation and analysis of multidimensional datasets in medicine". NATO ASI Series, F60, 63-79 (1990).

[16] M. Ashtari, J.L. Zito, B.I. Gold, J.A. Lieberman, M.T. Borenstein, P.G. Herman, "Computerized volume measurement of brain structure". Investigative Radiology, 25, 798-805 (1990).

[17] M.S. van der Knaap, C.J. Bakker, J.A.J. Faber, J. Valk, W.P.T.M. Mali, J. Willemse, R.H.J.M. Gooskens, "Comparison of skull circumference and linear measurements with CSF volume MR measurements in hydrocephalus". Journal of Computer Assisted Tomography, 16(5), 737-743. (1992).

[18] M. Kamber, D.L. Collins, R. Shinghal, G.S. Francis, A.C. Evans, "Model-based 3D segmentation of multiple sclerosis lesions in dual-echo MRI data". SPIE Medical Imaging, 1808, 590-598 (1992).

[19] S.P. Raya, "Low-Level Segmentation of 3-D Magnetic Resonance Brain Images - A Rule Based System". IEEE Transactions on Medical Imaging, 9, 327-337 (1990).

[20] I. Kapouleas, "Automatic Detection of Multiple Sclerosis Lesions in MR Brain Images". Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care, 739-745 (1989).

[21] R.P. Velthuizen, L.P. Clarke, L.O. Hall, A.M. Bensaid, "Multispectral 3D MRI segmentation using knowledge based systems". Medical Physics, 18, 622 (1991).

[22] U. Raff, F.D. Newman, "Automated lesion detection and lesion quantization in MR images using autoassociative memory". Medical Physics, 19(1), 71-77 (1992).

[23] L.M. Fletcher, J.B. Barsotti, J.P. Hornak, "A multispectral analysis of brain tissues". Magnetic Resonance in Medicine, 29, 623-630 (1993).

[24] M.W. Vannier, R.L. Butterfield, D. Jordan, W.A. Murphy, R.G. Levitt, M. Gado, "Multispectral Analysis of Magnetic Resonance Images". Radiology, 154, 221-224 (1985).

[25] M.W. Vannier, C.M. Speidel, D.L. Rickman, "Magnetic Resonance Imaging Multispectral Tissue Classification". News in Physiological Sciences, 3, 148-154 (1988).

[26] G. Gerig, J. Martin, R. Kikinis, O. Kübler, M. Shenton, F.A. Jolesz, "Automating segmentation of dual-echo MR head data". In: *Information processing in medical imaging*, Colchester ACF, Hawkes DJ, eds. Berlin: Springer-Verlag, 175-185 (1991).

[27] M. Jungke, W. von Seelen, G. Bielke, S. Meindl, G. Krone, M. Grigat, P. Higer, P. Pfannenstiel, "Information processing in nuclear magnetic resonance imaging". Magnetic Resonance Imaging, 6, 683-693 (1988).

[28] P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion". IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 629-639. (1990).

[29] R.O. Duda, P.E. Hart, *Pattern classification and scene analysis.* New York: John Wiley & Sons, 45-49. (1973).

[30] L.P. Clarke, R.P. Velthuizen, S. Phuphanich, J.D. Schellenberg, J.A. Arrington, M. Silbiger, "MRI Stability of three supervised segmentation techniques". Magnetic Resonance Imaging, 11, 95-106 (1993).

[31] J.C. Davis, "Analysis of multivariate data". In: *Statistics and data analysis in geology.* New York: Wiley, 527-546 (1986).

[32] H.E. Cline, W.E. Lorensen, R. Kikinis, F.A. Jolesz, "Three-dimensional segmentation of MR images of the head using probability and connectivity". Journal of Computer Assisted Tomography, 14, 1037-1-45. (1990).

[33] T.S. Curry, J.E. Dowdey, R.C. Murry, *Christensen's introduction to the physics of diagnostic radiology.* Philadelphia: Lea & Febiger, 343-345 (1984).

[34] R. Kikinis, F.A. Jolesz, G. Gerig, T. Sandor, H.E. Cline, W.E. Lorensen, M. Halle, S.A. Benton, "3D Morphometric and Morphologic Information Derived from Clinical Brain MR Images". NATO ASI Series, F60, 441-454 (1990).

[35] S.J. Karlik, M.K. Vandervoort, M. Hopkins, J.H. Noseworthy, "Serial quantitative brain MRI in multiple sclerosis using visual and volumetric techniques". Presented at the American Academy of Neurology. Neurology, 40(suppl 7):142. (1990).

[36] P.F. Panter, *Modulation, noise, and spectral analysis, applied to information transmission.* New York: McGraw-Hill, 634-635. (1965).

[37] B.K. Rutt, J.D. Mitchell, "Myelin-selective Imaging with 3-D Fast Spin-Echo MRI". Book of Abstracts, Proceedings of the SMRM, 137, (1993).

[38] W. Kucharczyk, "Neurological MRI". In: *The physics of magnetic resonance imaging*, Proceedings of the 1992 AAPM Summer School, Sprawls P, Bronskill MJ, eds. New York: AIP, 1992; 413-429.

CHAPTER 3 THE VARIABILITY OF MANUAL AND COMPUTER ASSISTED QUANTIFICATION OF MULTIPLE SCLEROSIS LESION VOLUMES

3.1 Introduction

It is generally agreed that Magnetic Resonance Imaging (MRI) has revolutionized the diagnosis and evaluation of demyelinating diseases such as Multiple Sclerosis (MS), since it was the first technique which permitted the direct visualization of MS lesions in vivo.[1] MR images reveal clearly the multiple, primarily periventricular, lesions that grow and shrink in size at different rates in separate areas of the brain[1-9]. The traditional measure of disease progression and therapeutic outcome in MS has been the clinical disability score.[10] Since the patient's disabilities progress slowly and their measure is subjective, considerable inter- and intra-raters variations occur.[11] Thus, soon after demonstrating the utility of MRI in assessing MS, a number of investigators began to develop quantitative tools for the analysis of MS plaques in MR images. These included estimates of lesion area from its largest diameter,[12] and estimates of total lesion area (or volume) from computer assisted outlining of individual lesions in multiple MR slices.[8,13] However, a number of studies have shown that the total lesion volume obtained from MRI does not correlate well with clinical status.[14,15] One factor explaining this lack of correlation may be the variability and subjectivity of clinical scores.[11] A second factor may be that lesions with similar volumes and appearance can cause different symptoms. Other factors may be that lesions detected by MRI can be asymptomatic[2,3] or may have little to do with the primary disease process. Also, patient symptoms may be largely due to spinal cord disease, which is not well imaged with MRI.

An additional factor explaining the lack of correlation between total lesion volume and clinical status may be the arduous and operator dependent nature of

lesion quantification techniques.[16] Although MR images provide an accurate view of the existence of the disease, ambiguities often arise in evaluating the size or extent of lesions. This results in less than optimal interpretation and monitoring of the progression, arrest, or reversal of the disease, particularly in patients participating in clinical trials.[2,3,6,7,12,17,18] Several groups have developed various quantitative techniques which permit a trained operator to view a 2-D MR image and then to use a track ball or similar device to outline lesions and other regions of interest.[13,19-28] In obtaining total area or volume, the outlining is repeated in multiple slices. These techniques generally rely on subjective user judgement, and data-driven segmentation (differentiation between lesion and normal tissue) such as edge detection,[29,30] thresholding,[31] and 2-D region growing.[32,33] Although these techniques are based on sound image processing fundamentals, they still suffer from inter- and intra-operator variations in computing volumes.[30,17,16] In addition, these techniques are variable because they are based on 2-dimensional images, yet the lesions are 3-dimensional. These variations make it difficult to monitor changes in individual lesions, especially small ones which may be of special interest during primary progressive MS.

In this chapter I present a new algorithm, which I have developed to overcome several difficulties in lesion quantification, and evaluate its impact on operator reliability and variability. The algorithm allows three-dimensional isolation of MS lesions in MR exams and reduces the difficulty of lesion volume measurement by providing assisted lesion quantification.

3.2 Methods

3.2.1 Multi-spectral Analysis

Many MR imaging protocols are inherently multi-spectral because they can provide multiple images with different contrast characteristics of the same anatomy. The multi-spectral nature of the data acquired in many MR exams provides

extra information which may be utilized to improve MR image segmentation.[34-42] My technique is based on multi-spectral analysis[34] and cluster classification,[43] which are common techniques used to exploit the additional information. These techniques are based on manipulation of a 2-D histogram showing the frequency of voxels with specific spectral characteristics such as proton density (PD) and T2-weighted intensities.[37-40] Thus, a typical 2-D histogram shows a 2-D distribution of intensities proportional to the number of pixels in the MR images with the PD-weighted intensity given on the horizontal axis and the T2-weighted intensities given on the vertical axis. Figure 2.2 on page 30 shows a typical pair of PD- and T2-weighted images from the same slice in the brain of an MS patient, and the corresponding 2-D histogram. The histogram in Figure 2.2 has a number of distinct clusters obliquely oriented and extending outward from the origin. Ideally, each cluster corresponds to a distinct feature or tissue in the cross-section.

3.2.2 Computer Assisted Lesion Quantification

I have developed a new algorithm based on multi-spectral analysis which provides interactive assistance to a knowledgeable operator and acts locally upon individual lesions. Nearest-neighbour cluster classification[43] is used interactively to differentiate the intensity characteristics of a single lesion from that of surrounding tissues, after which the lesion voxels are identified. The lesion quantification procedure, illustrated in Figure 3.1 required three steps:

1) Identification of white matter and cerebrospinal fluid. The operator identified regions of "pure" cerebrospinal fluid (CSF) and white matter (WM) in multiple locations in a number of slices, as shown in Figure 3.1a. Sample regions were selected across each slice to incorporate intensity variations due to RF inhomogeneities. The sample regions were then used to estimate the PD and T2 intensity distributions in the 2-D histogram of WM and CSF and to "train" the system about the typical intensity characteristics of both WM and CSF.

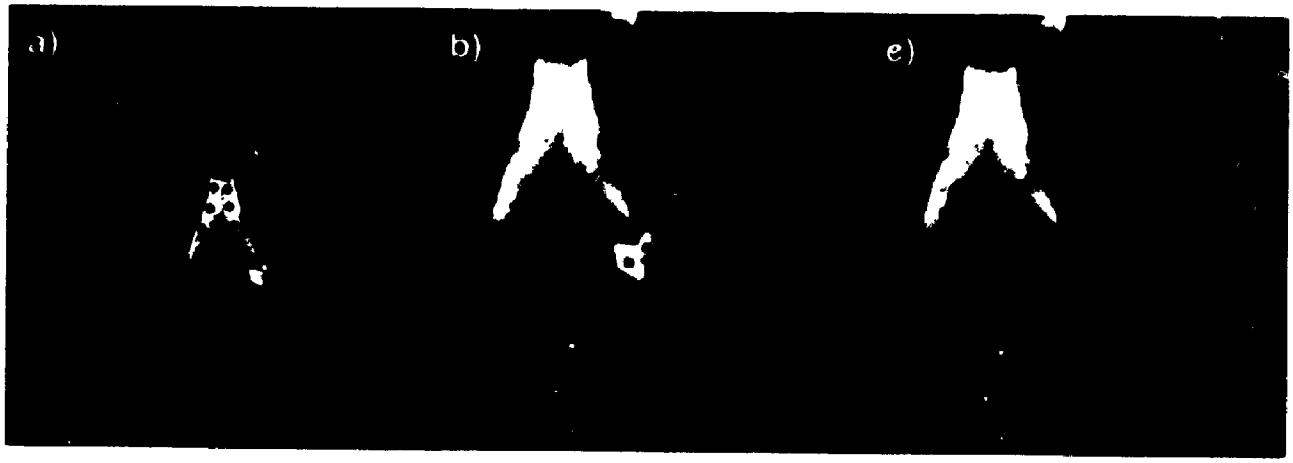


Figure 3.1 The three steps of computer assisted quantification.

Computer assisted lesion quantification requires the following steps: a) Background identification - many small regions of "pure" CSF and white matter from various regions are identified and labelled non-lesion, or "background"; b) Lesion identification - the operator selects one or more locations inside a single lesion; c) Volume calculation - multi-spectral classification is performed to determine the intensity characteristics of the lesion which is then identified using connected-component labelling.

2) Lesion identification. The operator identified one or more voxels inside a single lesion as shown in Figure 3.1b. The 3x3 regions centered about each identified voxel were used to estimate the PD and T2 intensity distribution of the entire lesion. This information, along with estimates of WM and CSF intensities obtained from "training" in the previous step, were used to identify the specific PD and T2 intensities which occurred in the lesion.

3) Lesion Quantification. Connected-component labelling[44] was performed beginning at each of the lesion voxels identified by the operator. The labelling was constrained to include only those voxels with PD and T2 intensities which were identified as valid lesion intensities in the previous step. All lesion-labelled voxels were then highlighted in the MR images for visual verification by the operator as shown in Figure 3.1c.

Step 1 typically required less than 5 minutes to complete, and was performed once before all lesions in an exam were quantified. Steps 2 and 3 were repeated interactively in each slice for each lesion until the operator was satisfied with the

labelling of lesion voxels. Often, only one iteration of steps 2 and 3 was necessary to satisfactorily identify all voxels in a lesion. Occasionally however, it was difficult to obtain a satisfactory labelling or the initial labelling contained obvious errors. In these situations the algorithm allowed the operator to identify improperly labelled voxels, and "train" the system about errors for a particular lesion. This information was then used in the next iteration of steps 2 and 3 to further constrain the voxel labelling and correct the lesion quantification.

Although connected-component labelling of lesion voxels can proceed in 2-D or 3-D, the large slice thickness and interslice gap reduced the effectiveness of 3-D labelling in this experiment. Therefore 2-D labelling was employed in each slice in which the lesion was visible. The volume of the lesion was determined by summing the number of lesion-labelled voxels in each slice, then multiplying by the voxel volume.

This algorithm was implemented as part of a more extensive analysis program I developed[45] which includes 3D multi-spectral analysis and visualization. The system was implemented on Sun SparcStation 2 computers (Sun Microsystems, Mountainview Calif.) running SunOS, and Openwindows (Sun Microsystems, Mountainview Calif.). The system was written in the C language and uses the Xview Toolkit (Sun Microsystems, Mountainview Calif.).

3.2.3 Manual Lesion Quantification

Operators optionally adjusted the contrast of either or both of the PD and T2 weighted images (using window and level operations), zoomed the images (up to ten times magnification), and then used the mouse to interactively outline lesion boundaries in all the slices in which the lesion was visible. Lesion outlining occured in either image, while the lesion boundary was displayed in both images. Detailed adjustment of lesion boundaries was possible using the editing procedures provided by the system. The volume of each lesion was determined by

identifying its boundary in multiple slices, and summing the voxels entirely contained within each boundary and then multiplying by the voxel volume.

3.2.4 Operator performance: Phantom Study

Operator accuracy while performing manual and assisted quantification was measured under ideal conditions using a test phantom containing high contrast "lesions" with regular shapes and known volumes. A detailed description of this phantom and its construction is presented elsewhere.[39,45] Briefly, the phantom consisted of agarose "lesions" doped with 0.5% graphite surrounded by 1% weight per volume agarose "brain". Thirty seven "lesions" were constructed to have similar in-plane dimensions (1.2cm x 1.2cm) and out-of-plane dimensions varying from 0.3 cm to 5.3 cm. The resulting "lesions" varied in volume from 0.4 cm^3 to 6.8 cm^3. The phantom was imaged six times using the imaging protocol listed in Table 3.1. 3mm thick slices were employed to allow accurate quantification of "lesion" volumes.[45] Slice positioning was incremented by 0.5 mm between each examination to cover the range of possible slice positionings.

Image Type	T2 Weighted
Image Plane	Sagittal
Field-of-view	20 cm
TR	2000 ms
TE1/TE2	35/70 ms
Number of slices	49
Slice thickness	3 mm
Slice gap	0 mm
Matrix size	256 x 128
NEX	1
Bandwidth	16/8
Scan time	9:20

Table 3.1 MR imaging parameters for the phantom exams

Proton density (TE1) and T2 weighted (TE2) images were obtained at each slice.

From the 37 "lesions", five were selected for repeated measurements by the operators. The five "lesions" varied in volume from 0.4 cm^3 to 4.02 cm^3. Six operators performed both computer assisted, and manual quantification (described above) of the 5 "lesions". This process was repeated for each of the 6 exams, after which the standard error and mean volume of each lesion was computed for both manual and assisted quantification. These measurements are listed in Tables 3.2 and 3.3. The mean "lesion" volumes were plotted against the actual

Lesion	Actual Volume	Operator 1	Operator 2	Operator 3	Operator 4	Operator 5	Operator 6
A	0.40	0.60 ± .05	0.91 ± .12	0.95 ± .04	0.93 ± .03	0.93 ± .04	0.95 ± .05
B	1.16	1.09 ± .19	1.67 ± .18	1.93 ± .23	1.82 ± .25	1.76 ± .20	1.89 ± .23
C	1.82	1.73 ± .10	2.40 ± .14	2.72 ± .09	2.71 ± .09	2.67 ± .03	2.64 ± .16
D	2.90	2.86 ± .15	3.64 ± .31	4.08 ± .17	4.12 ± .14	4.06 ± .03	3.97 ± .15
E	4.02	3.89 ± .12	4.89 ± .41	5.21 ± .18	5.16 ± .15	5.23 ± .18	5.15 ± .18

Table 3.2 Average test phantom "lesion" volumes measured using manual quantification.

Values are the average test phantom "lesion" volumes (± 2 standard errors, determined from 6 repeated measurements) measured by each operator from the six phantom exams, using manual quantification. The actual "lesion" volumes are shown in the first column

Lesion	Actual Volume	Operator 1	Operator 2	Operator 3	Operator 4	Operator 5	Operator 6
A	0.40	0.46 ± .04	0.63 ± .05	0.77 ± .04	0.80 ± .06	0.84 ± .06	0.65 ± .07
B	1.16	0.87 ± .04	1.38 ± .09	1.52 ± .15	1.54 ± .19	1.67 ± .18	1.41 ± .15
C	1.82	1.58 ± .13	2.04 ± .21	2.31 ± .20	2.18 ± .13	2.36 ± .11	2.09 ± .19
D	2.90	2.66 ± .16	3.16 ± .25	3.31 ± .21	3.32 ± .22	3.78 ± .13	3.20 ± .27
E	4.02	3.52 ± .07	4.13 ± .17	4.59 ± .14	4.29 ± .14	4.77 ± .25	4.18 ± .23

Table 3.3 Average test phantom "lesion" volumes measured using assisted quantification.

Values are the average test phantom "lesion" volumes (± 2 standard errors, determined from 6 repeated measurements) measured by each operator from the six phantom exams, using assisted quantification. The actual "lesion" volumes are shown in the first column

"lesion" volumes in Figure 3.2 to determine operator accuracy for both manual and assisted quantification. Finally, the residual error, defined as the magnitude of the actual "lesion" volume minus the mean measured "lesion" volume, is plotted in Figure 3.3 for manual and assisted quantification of each "lesion" by each operator.

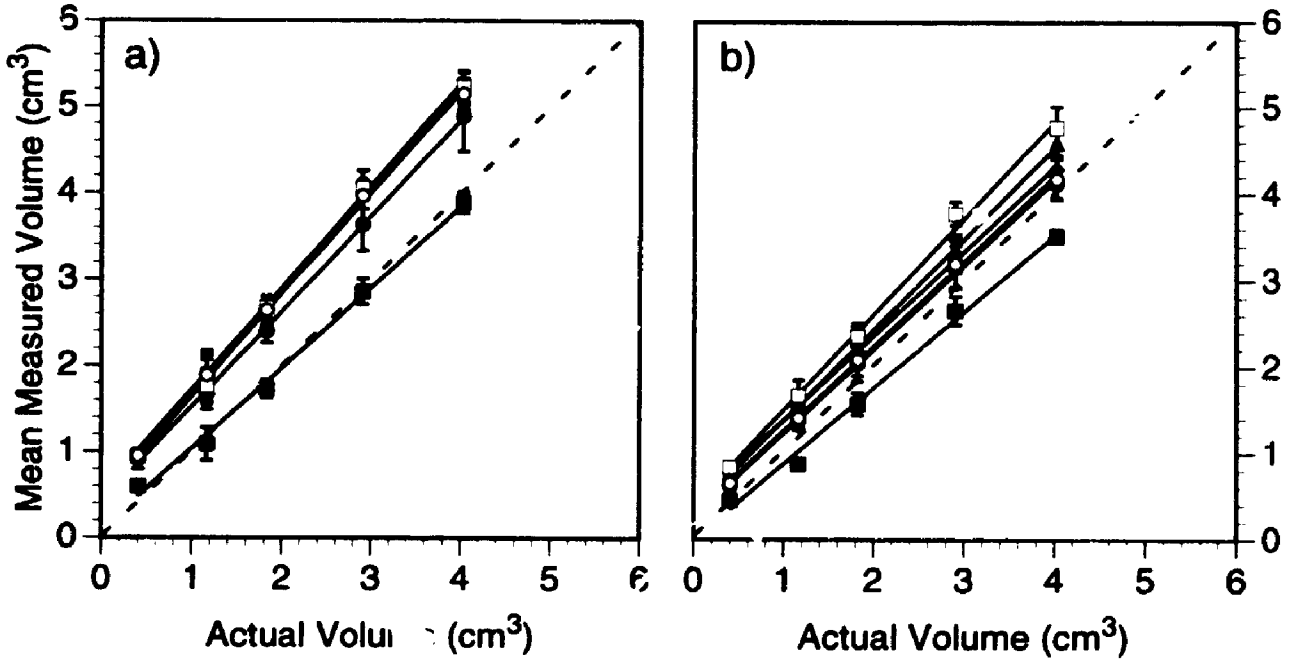


Figure 3.2 The accuracy of volume measurement of test phantom "lesions".

(a) and (b) are plots of the results from manual and assisted quantification, respectively. Each point represents the mean "lesion" volume determined from 6 repeated measurements of 5 "lesions" by 6 operators, and the error bars represent two standard errors in each mean volume measurement. The dashed line indicates ideal operator performance.

3.2.5 Operator Performance: Patient Study

Operator variability and reliability under clinical conditions was measured using MR exams of a single MS patient. Experiments were performed to measure the inter- and intra-operator variability of manual and assisted lesion quantifica-

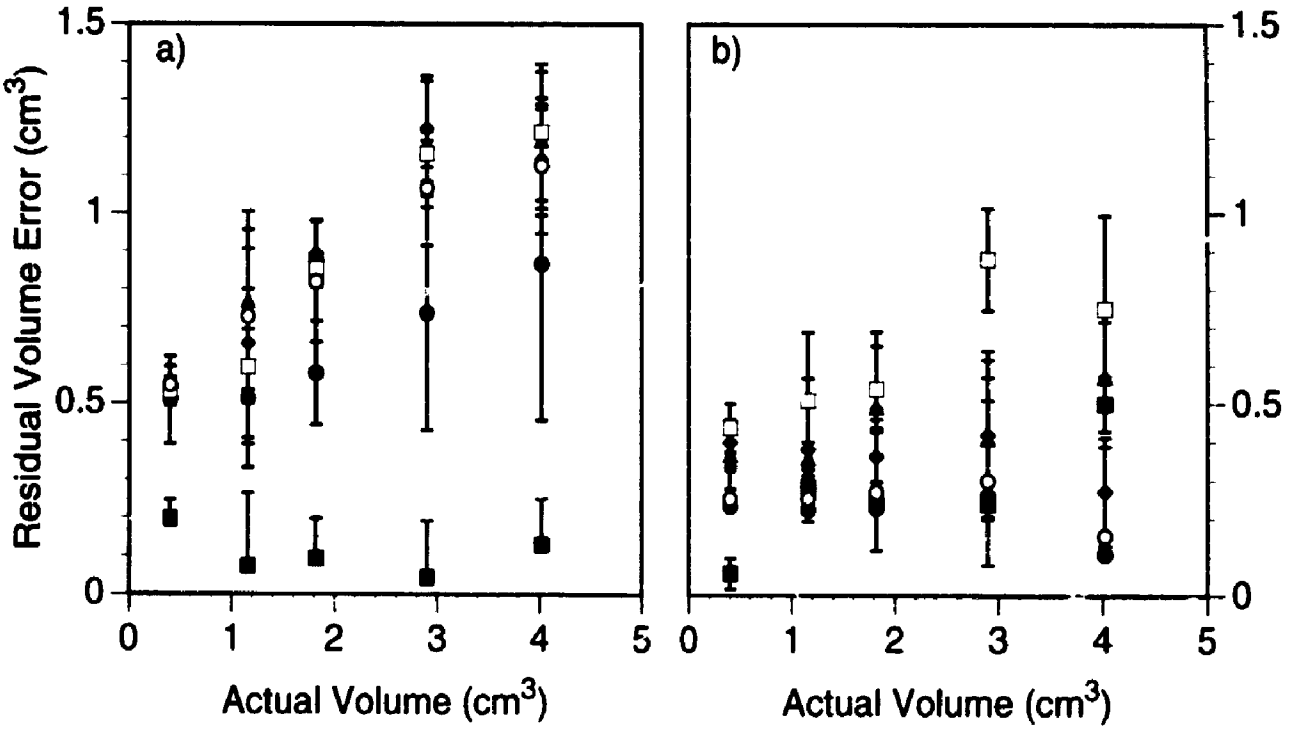


Figure 3.3 The residual errors in volume measurement of test phantom "lesions".

(a) and (b) are plots of the results from manual and assisted quantification, respectively.The residual errors are defined as the magnitude of the actual "lesion" volume minus the mean measured "lesion" volume. Each point represents the mean "lesion" volume determined from 6 repeated measurements of 5 "lesions" by 6 operators, and the error bars represent two standard errors in each mean volume measurement.

tion. Operator accuracy was not measured, since the actual lesion volumes were not available.

MR Exam Protocol. Exams were acquired on a 1.5T GE Signa imaging system (GE Medical Systems, Milwaukee) using the multi-echo imaging protocol listed in Table 3.4. This protocol was developed for a study evaluating the accuracy of MR at 1.5T and 0.5T for the diagnosis of MS.[46] Initially, a series of sagittal images were acquired using the "localizer" protocol listed in Table 3.4. Next, the inferior margins of the genu and splenium of the corpus callosum were identified

in these images, and used to establish the oblique axial slice orientation and position.[46] Finally, the "multi-echo" protocol listed in Table 3.4 was used to acquire PD and T2 weighted images at each of the oblique axial slices in the patient's brain. All exams were transferred via ethernet to Sparc workstations (Sun Computer Corporation, Mountainview Calif.) for processing. PD and T2 weighted 3-D images were created from each exam by combining all early echo images and all later echo images, respectively.

	Localizer	Mulit-echo
Image Type	T1 Weighted	T2 Weighted
Image Plane	Sagittal	Axial-oblique
Field-of-view (cm)	22	22
TR (ms)	400	2800
Te1/Te2 (ms)	20	30/80
Number of slices	9	36
Slice thickness (mm)	5	5
Slice gap (mm)	0	2.5
Matrix size	256 x 128	256 x 192
NEX	1	1
Bandwidth (kHz)	16	16/8
Scan time	0.56	9:21

Table 3.4 MR imaging parameters used in the patient experiments.

Study Design. Five lesions were selected for the operator studies. Figure 3.4 shows detailed views from the MR exam of each of the five lesions, labelled A through E. The lesions were deliberately selected to vary in volume (from a minimum of approximately 0.2 cm^3 to a maximum of approximately 2.7 cm^3) and vary from focal (Lesion D) to diffuse (Lesion E), and thus more completely cover the range of possible lesions. Six operators made 5 repeated measurements of each of the 5 lesions using both manual and assisted quantification over a 10 month period. In total, 300 lesion volumes were quantified (i.e. 6 operators x 5 repeated measurements x 5 lesions x 2 techniques). All manual quantifications were com-

pleted before any assisted quantification was performed, to avoid any operator bias introduced by the results of the assisted quantification.

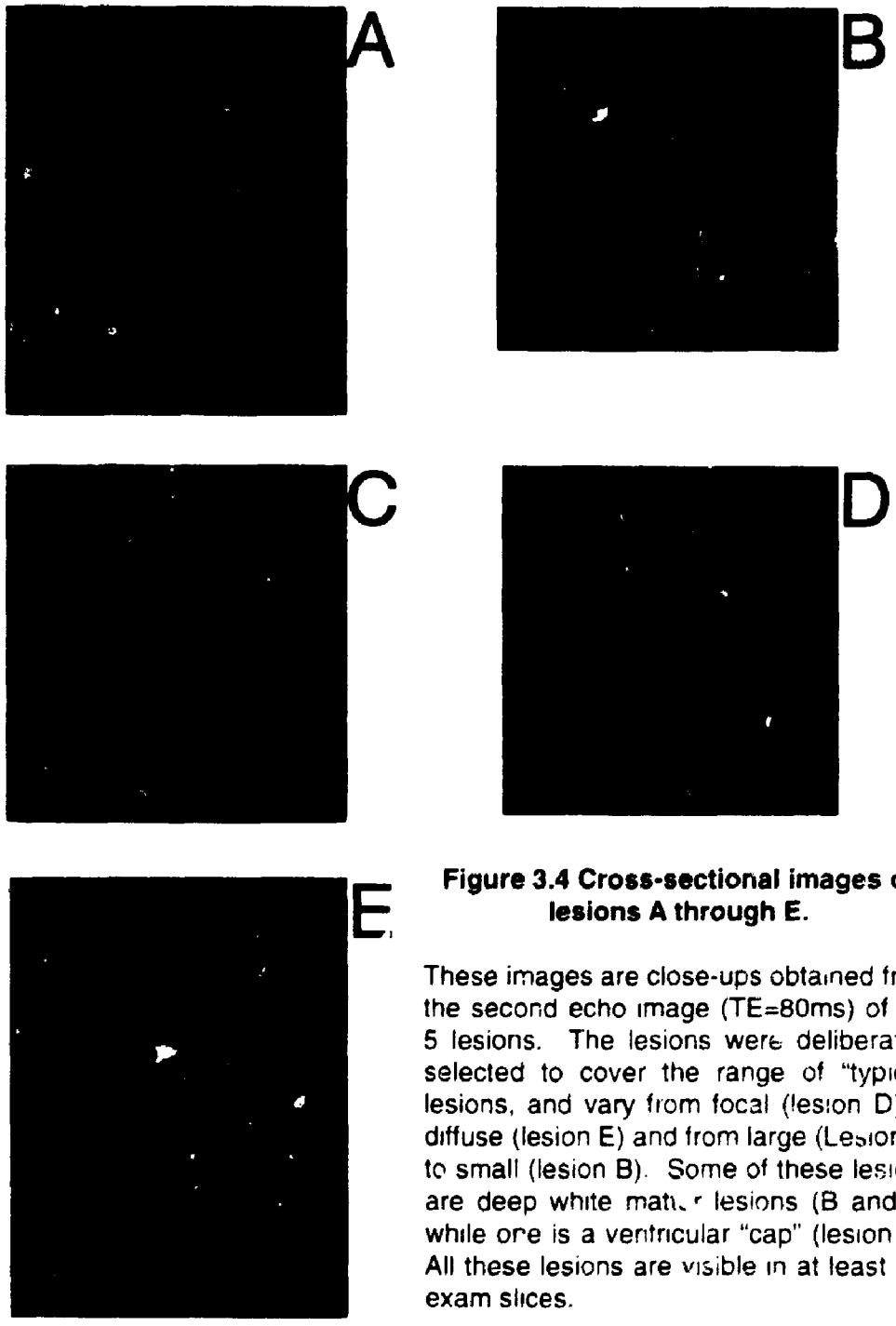


Figure 3.4 Cross-sectional images of lesions A through E.

These images are close-ups obtained from the second echo image (TE=80ms) of the 5 lesions. The lesions were deliberately selected to cover the range of "typical" lesions, and vary from focal (lesion D) to diffuse (lesion E) and from large (Lesion A) to small (lesion B). Some of these lesions are deep white matter lesions (B and D) while one is a ventricular "cap" (lesion C). All these lesions are visible in at least two exam slices.

The mean lesion volumes measured by each operator for both manual and assisted quantification are plotted in Figure 3.5. A two-way analysis of variance (ANOVA) was performed on the manual and assisted quantifications to determine the total variance (σ_T^2) and its components due to: differences between lesions (σ_L^2); differences between operators (σ_O^2); the interaction between operators and lesions (σ_{LO}^2); and, the random error within the typical operator (σ_ϵ^2). These variance components were used to calculate inter- and intra-operator reliabilities, ($\hat{\rho}_{inter}$ and $\hat{\rho}_{intra}$), and standard errors of measurement (SEM_{inter} and SEM_{intra}), using methods described by Eliasziw et. al.[47] SEM's from different experiments were also tested to determine if they were statistically different (for more information see the Appendix, Section 3.5 on page 82). Finally, absolute differences between manual and assisted quantification were compared using paired t-tests[48] across measurements from both experiments.

Spatial Error distribution. The impact of computer assisted lesion quantification on the variability of lesion boundary determination was examined by analyzing the spatial distribution of the operator variabilities. My system was used to mark and record voxels labelled as lesion in each exam for all trials described above. For both manual and assisted quantification, the selection frequency, F(x,y,z), for a single voxel with coordinates (x,y,z) was determined using the following formula:

$$F(x, y, z) = \frac{N(x, y, z)}{N_T} \qquad 1 \le N(x, y, z) \le N_T \tag{3.1}$$

where N(x,y,z) was the number of times the voxel with coordinates (x,y,z) was determined to be lesion, and N_T was the total number of manual measurements or assisted measurements. The spatial distribution of voxel selection frequencies was coded by a color-scale where white represents 100% selection and black represents 3% selection. The spatial distribution is plotted in Figure 3.6 for all manual and assisted quantifications of all five lesions.

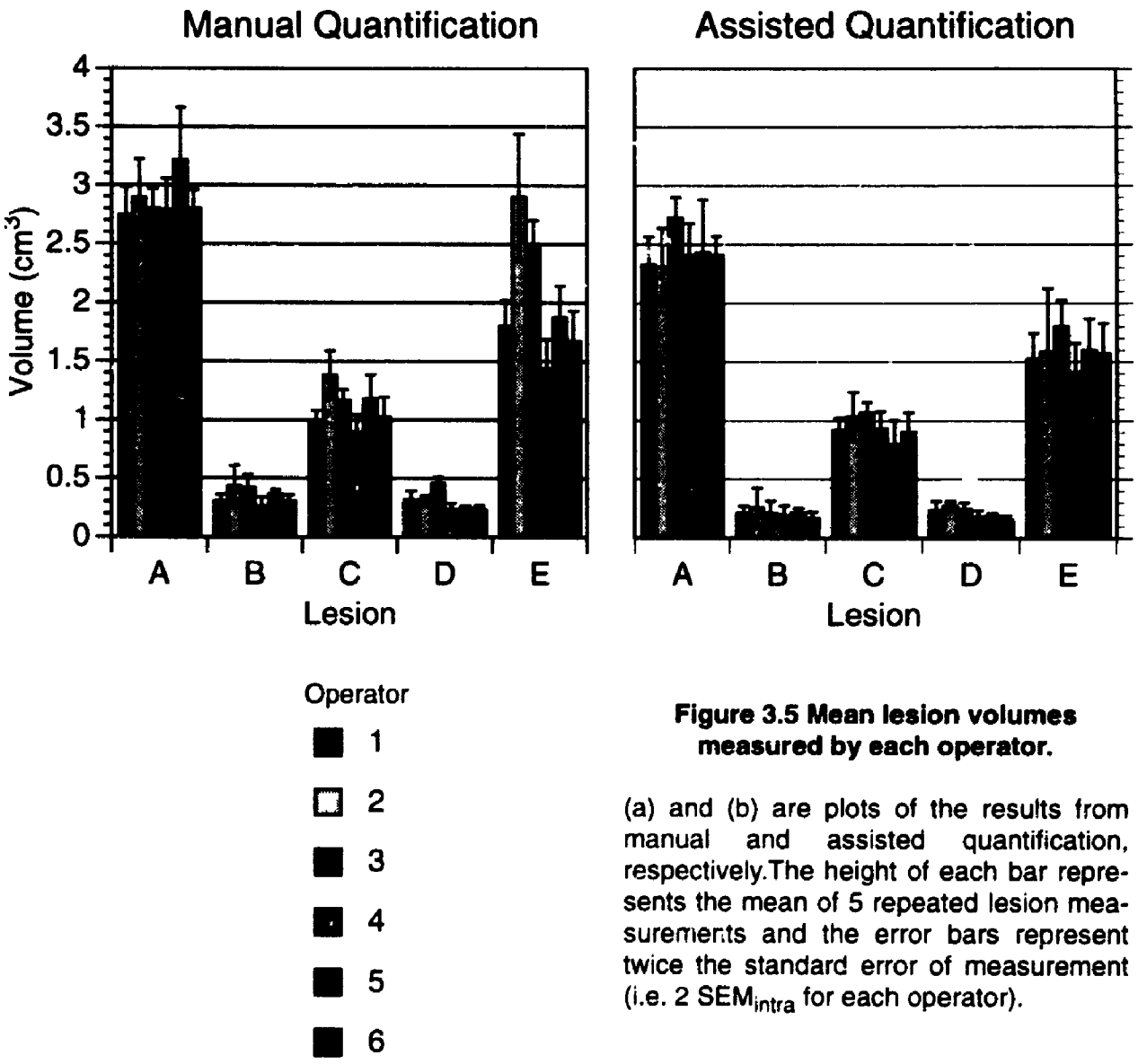
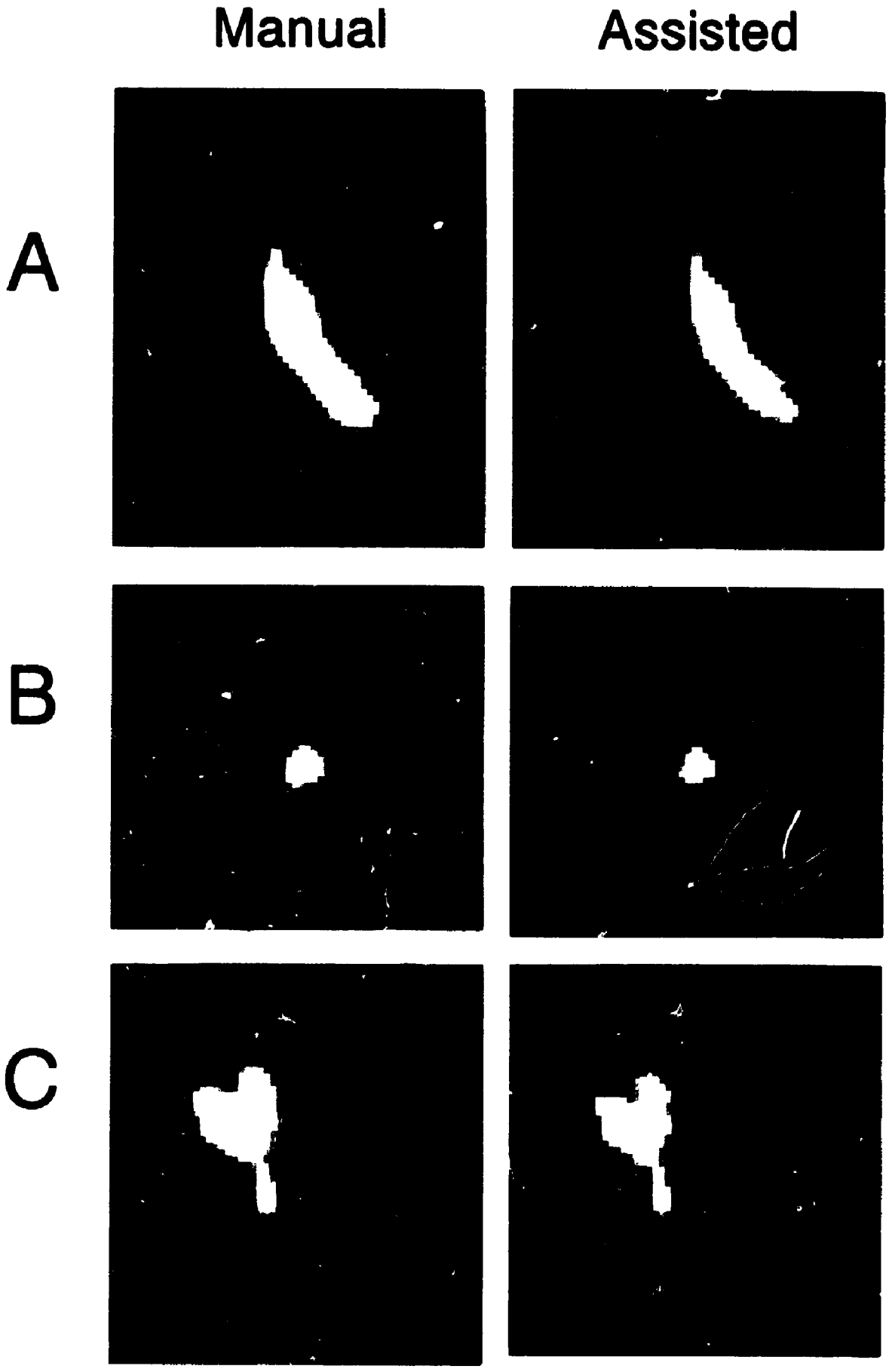


Figure 3.5 Mean lesion volumes measured by each operator.

(a) and (b) are plots of the results from manual and assisted quantification, respectively.The height of each bar represents the mean of 5 repeated lesion measurements and the error bars represent twice the standard error of measurement (i.e. 2 SEM_{intra} for each operator).

To examine and compare the distribution of voxel selection frequencies for manual and assisted quantification, I used these data to calculate an operator error probability distribution by summing the number of voxels at each selection frequency F. Histograms of these results are shown in Figure 3.7.


Manual
Assisted
A
B
C

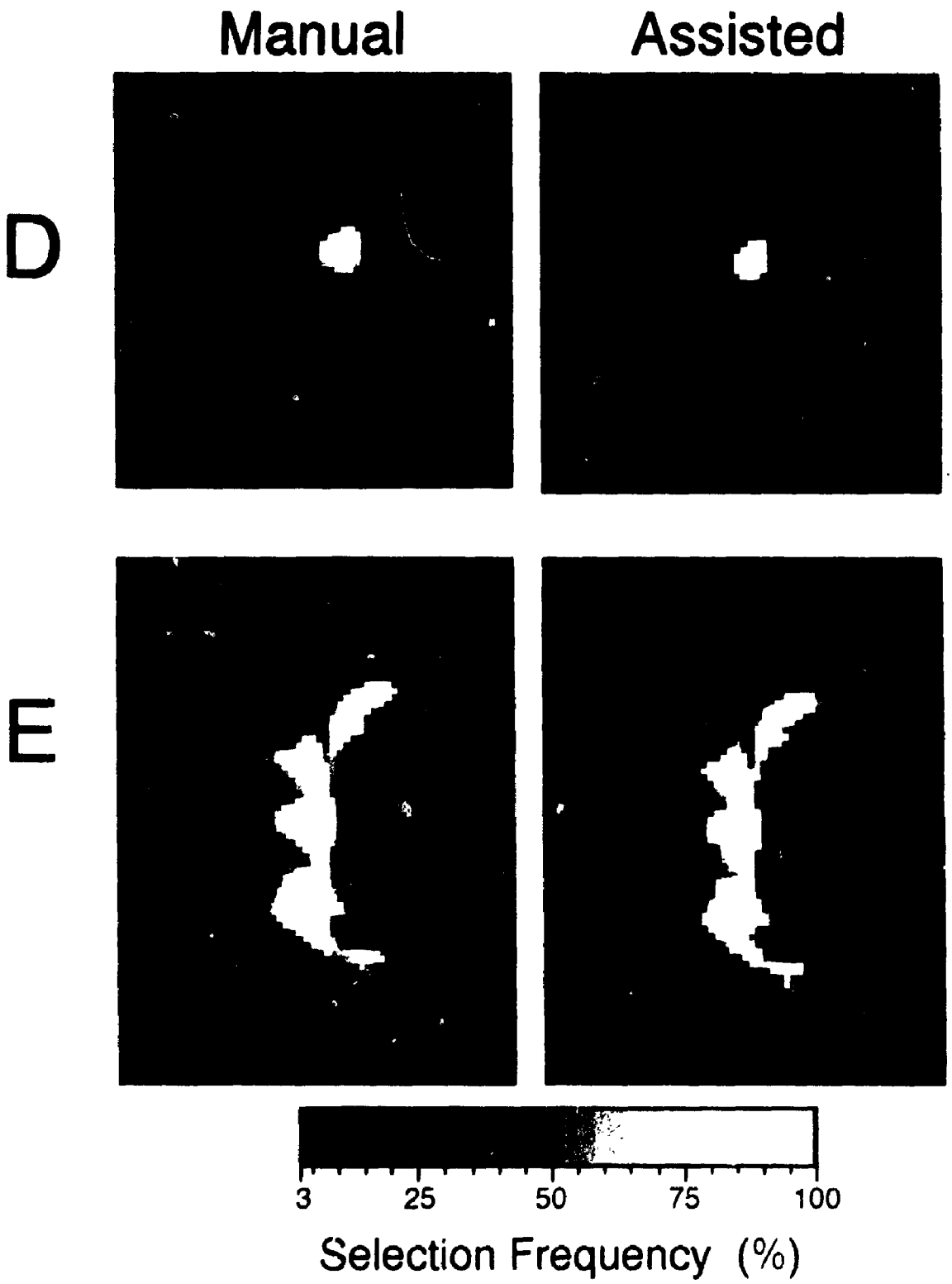


Figure 3.6 The frequency of voxel selection during lesion quantification.

The left and right columns show results from manual and assisted quantification, respectively. The color-scale overlay indicates the number of times each voxel was labelled as a lesion by the 6 operators performing 5 trials each. Dark red regions correspond to voxels which were rarely labelled lesion (3%), while white regions correspond to regions always labelled as a lesion (100%).

3.3 Results

3.3.1 Operator Performance: Phantom Study

Tables 3.2 and 3.3 (on page 63) show the average "lesion" volumes measured by each operator using manual and assisted quantification, respectively. "Lesion" volume estimates provided by manual quantification were on average 0.42 ± 0.05 cm^3 larger than those provided by assisted quantification. This difference is significant at better than the 1% level. The mean measured "lesion" volumes are plotted against the actual "lesion" volumes in Figure 3.2 (on page 64) to reveal the accuracy of "lesion" quantification by each operator for manual (Figure 3.2a) and assisted (Figure 3.2b) quantification. Five of six operators overestimated "lesion" volumes using manual quantification, while assisted quantification resulted in more ideal performance by operators 2 through 5, but less than ideal performance by Operator 1. Figure 3.3 (on page 65) shows the operators' residual errors, defined as the magnitude of the differences between the actual and average measured "lesion" volumes, for manual (Figure 3.3a) and assisted (Figure 3.3b) quantification. During manual quantification, the errors of Operator 1 remain essentially constant for all "lesion" volumes, reaching a maximum of 0.13 cm^3 at "lesion" volume 4.02 cm^3. For Operators 2 through 6 residual errors increased with increasing "lesion" volume. The residual errors of Operators 3 through 6 increased from approximately 0.55 cm^3 at "lesion" volume 0.4 cm^3, to approximately 1.20 cm^3 at "lesion" volume 4.02 cm^3, while those of Operator 2 went from 0.50 cm^3 to 0.87 cm^3. Assisted quantification reduced the residual errors of Operators 2 through 6 for all "lesions". Their errors reached a maximum of 0.88 cm^3 (Operator 3, "lesion" volume 2.90 cm^3). However assisted quantification increased the residual errors of Operator 1 to a maximum of 0.50 cm^3 ("lesion" volume 4.02 cm^3).

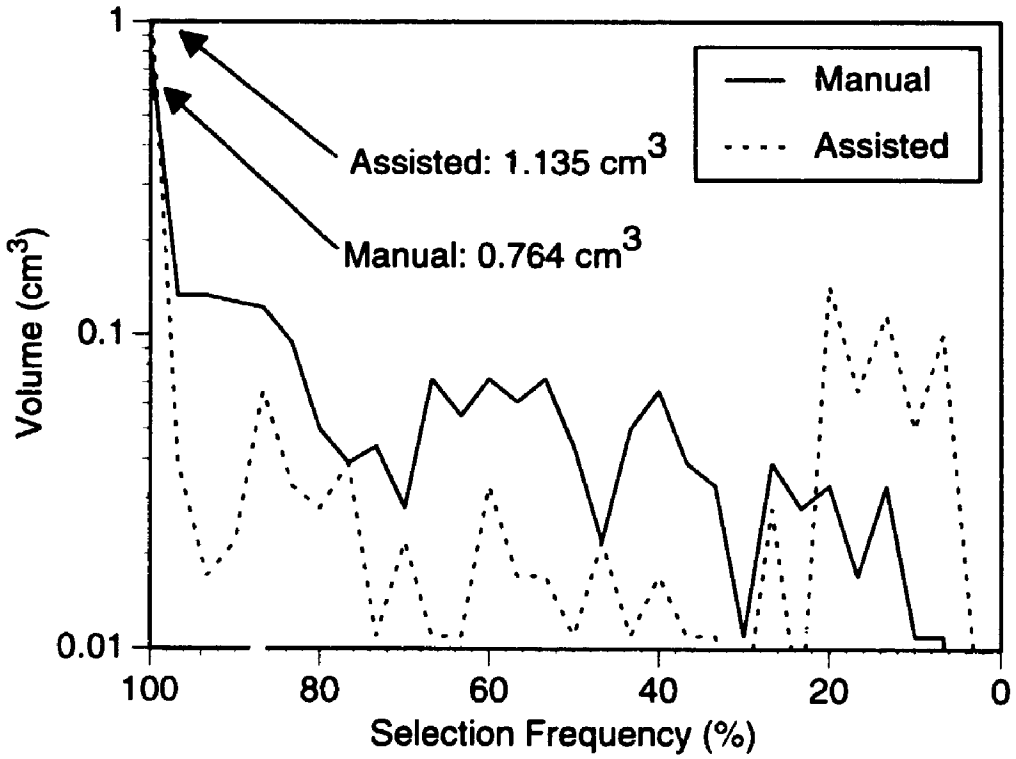


Figure 3.7 The volume selected at each frequency during quantification of lesion E.

Histograms showing the volume (ie. number of voxels) selected at each frequency for manual (solid line) and assisted (dashed line) quantification of lesion E.

3.3.2 Operator Performance: Patient Study

Figure 3.5 (on page 69) shows histograms of the average volume of each lesion measured by the six operators using both manual and assisted quantification. A comparison of results between lesions reveals that volume measurements ranged from a maximum of 3.21 cm^3 (Manual quantification, Lesion A, Operator 5) to a minimum of 0.16 cm^3 (Assisted quantification, Lesion D, Operator 6) with a mean of 1.20 cm^3. The individual standard errors of measurement ranged from a maximum of 0.27 cm^3 (Manual quantification, Lesion E, Operator 2) to a minimum of 0.01 cm^3 (Assisted quantification, Lesion D, Operator 6). Volume measurements of small focal lesions B and D were highly consistent between operators and within operators, while measurements of the large diffuse lesion E were highly variable both between and within the operators. The results also show that

assisted quantification provided lesion volume estimates that were smaller by 0.24 cm^3 on average than those by manual quantification. This difference represents 20% of the mean lesion volume, and is significant at better than the 1% level.

Table 3.5 lists the total standard deviation of lesion volume measurement for both manual and assisted quantification and its various contributing components. The total standard deviation was 1.16 cm^3 for manual quantification, but only 0.97 cm^3 for assisted quantification. The lesion component of standard deviation, $\hat{\sigma}_L$, was found to be 1.11 cm^3 for manual quantification and 0.95 cm^3 for assisted quantification. This component is the largest contributor to total standard deviation and reflects the large differences between lesions selected for analysis. Other components of standard deviation are small in comparison. However, in both experiments, the within operator variability ($\hat{\sigma}_e$) is larger than either the operator ($\hat{\sigma}_O$) or the operator-lesion ($\hat{\sigma}_{LO}$) components. This implies that the random variability within the operators was larger than the effect of selecting this particular group of operators or having these operators quantify these particular lesions.

The standard deviation components listed in Table 3.5 were used to calculate the inter- and intra-operator SEM's also listed in Table 3.5 (see the Appendix on page 82). The inter-operator SEM, a measure of the average variation between operators, was 0.34 cm^3 for manual quantification and 0.17 cm^3 for assisted quantification. The intra-operator SEM, equivalent to the intra-operator standard deviation ($\hat{\sigma}_e$), is a measure of the random error within the typical operator. SEM_{intra} was 0.23 cm^3 for manual quantification, and 0.15 cm^3 for assisted quantification. SEM_{inter} will always be equal to or larger than SEM_{intra} since within-operator variability is just one component of between-operator variability. From Table 3.5, it can be observed that assisted quantification reduced SEM_{inter} by 50% (0.17 cm^3) and reduced SEM_{intra} by 35% (0.08 cm^3) compared to manual quantification. These differences were tested (using the statistic d, described in the Appendix), and both found to be highly significant with p-values less than 0.0001.

	Manual	Assisted
$\hat{\sigma}_{Total}$ (cm^3)	1.16	0.97
$\hat{\sigma}_L$ (cm^3)	1.11	0.95
$\hat{\sigma}_O$ (cm^3)	0.14	0.06
$\hat{\sigma}_{LO}$ (cm^3)	0.21	0.04
$\hat{\sigma}_e$ (cm^3)	0.23	0.15
SEM_{Inter} (cm^3)	0.34	0.17
$\hat{\rho}_{Inter}$	0.91	0.97
SEM_{Intra} (cm^3)	0.23	0.15
$\hat{\rho}_{Intra}$	0.96	0.97

Table 3.5 The variability and reliability of manual and assisted quantification.

Estimates of standard deviations ($\hat{\sigma}$), reliabilities ($\hat{\rho}$), and standard errors of measurement (SEM) for manual and assisted quantification. Units are cm^3 except for the reliability coefficients (ρ), which are dimensionless.

The standard deviation components listed in Table 3.5 were also used to determine the reliability coefficients, $\hat{\rho}_{inter}$ and $\hat{\rho}_{intra}$, (see the Appendix). These coefficients can range from 0 to 1, with higher values corresponding to increased reliability.[49] In general, $\hat{\rho}_{intra}$ is higher than $\hat{\rho}_{inter}$ since there are more sources of variability contributing to $\hat{\rho}_{inter}$.

Figure 3.6 (on page 71) shows the voxel selection frequency for lesions A through E determined from all trials by all operators using both manual and assisted quantification. In these visualizations, a sharper transition between pure white (region always quantified) and the gray-scale MR image (region never quantified) indicates more consistent operator performance. All of the lesions are bordered by regions of low quantification frequency. The small focal lesions B and

D have a "smooth" transition between frequently and infrequently quantified regions with manual quantification, and a "sharp" transition with assisted quantification. The large lesion A has a sharp transition with both quantification techniques, except in the immediate periventricular region. In this region, manual quantification results in a smoother transition, indicating operator variability is higher there than with assisted quantification. Lesion C contains two focal regions quantified very frequently using both techniques and a third diffuse region quantified in less than half of the manual trials. Unlike lesion A, this region is far from the ventricle. The large, diffuse, multi-focal lesion E contains a number of regions with high quantification frequency. For manual quantification each of these regions has a smooth transition to low quantification frequency, indicating that in some of these trials, operators considered the focal regions of lesion E to be connected and in others to be disjoint. However, with assisted quantification these regions are consistently connected. In addition, with manual quantification, lesion E has large posterior and anterior "lobes" with low quantification frequency (< 50%), indicating operator variability is high in these regions. These "lobes" are greatly reduced by assisted quantification.

Although useful for visualizing where the operator variabilities are high (and low) within each lesion, the images of Figure 3.6 are not quantitative. Quantitative analysis of quantification frequencies for lesion E is shown in Figure 3.7. This semi-log graph shows the volume at each quantification frequency for manual and assisted quantification. Under ideal conditions, every lesion voxel would be identified by all operators all of the time and therefore, these graphs would consist of a single point on the ordinate axis. However, in my experiments, variability within and between operators results in some voxels being selected with lower frequency, which in turn caused spreading of the volume distribution to lower selection frequencies. Thus, in these graphs high intersection with the ordinate axis, and small volumes at selection frequencies lower than 100% indicate better operator performance. Ordinate intersections, indicated in Figure 3.7, were 0.764 cm^3

and 1.135 cm^3 for manual and assisted quantification respectively. In addition, the volumes for most selection frequencies lower than 100% were smaller for assisted quantification than for manual quantification. However, a greater volume is selected at very low selection frequency (<20%) by assisted quantification, yet this volume does not contribute significantly to the mean estimate of mean lesion volume because of its low selection frequency.

3.4 Discussion

3.4.1 Operator Performance: Phantom Study

Examination of the residual errors in "lesion" volume measurement in Figure 3.3 reveals that they tend to increase with increasing "lesion" volume. For the high-contrast "lesions" quantified in the phantom experiment, most of the variability, and thus error, in volume measurement occurred along the "lesion" edge. If we consider each "lesion" to be approximately circular in cross-section, then in each slice the length of the "lesion" edge, and thus the error, will vary as the square root of the "lesion" area. In my experiment, each "lesion" volume was determined by summing the cross-sectional areas identified as "lesion" in each slice in which the "lesion" was visible. Since the "lesions" are highly regular, their contrast and area remain essentially constant in each slice. Thus, the errors in each slice are highly correlated and sum together when calculating total "lesion" volume. Therefore, I expect the residual error in "lesion" volume measurement to vary with the square-root of the "lesion" volume.

Indeed, the data in Figure 3.3 generally follow this trend, both for manual and assisted quantification. From Figure 3.3a, the average residual error at "lesion" volume of 1.0 cm^3 is approximately 0.5 cm^3. This average residual error approximately doubles to about 1.0 cm^3 with a quadrupling of "lesion" volume to 4.0 cm^3. Similarly, for assisted quantification the average residual error doubles from about

0.2 cm^3 to nearly 0.4 cm^3 with a quadrupling of "lesion" volume from 1.0 cm^3 to 4.0 cm^3.

3.4.2 Operator Performance: Patient Study

The patient study presented in this chapter has examined the reliability and variability of manual and assisted lesion quantification using typical MR exams of an MS patient. The statistical and experimental methods presented have addressed the precision, but not the accuracy, of quantification since true lesion volumes are very difficult to determine. Nevertheless, knowledge of the precision of lesion quantification is essential for clinical trials concerned with changes in individual lesions. In these situations, the minimum volume change that can be detected confidently beyond measurement variability depends upon the desired confidence level, and the operator variability. In particular, the change, Δv, in volume between two successive measurements that can be detected at a given confidence level is:[47]

$$\Delta v \geq z_{\alpha} \cdot \sqrt{2} \cdot SEM \qquad (3.2)$$

where z_{α} is the standard normal deviate exceeded in either direction with probability α (ie. for $\alpha = 0.05$ then $z_{\alpha} = 1.96$). If the same operator makes both measurements then SEM_{intra} is used in equation 3.2; however, if different operators make the measurements then SEM_{inter} is used. Table 3.6 lists Δv for manual and assisted quantification with $\alpha = 0.05$. The minimum change when different operators make successive measurements (Δv_{inter}) is 0.94 cm^3 for manual quantification, and 0.47 cm^3 for assisted quantification. The minimum change when the same operator makes both measurements (Δv_{intra}) is 0.64 cm^3 for manual quantification, and 0.42 cm^3 for assisted quantification. Thus, with multiple operators, assisted quantification reduces the minimum detectable change by one half, compared to manual quantification. With only one operator, assisted quantification reduces the minimum detectable change by one third, compared to manual

quantification. These reductions in Δv are a result of the lower SEM's for assisted quantification. Δv may also be reduced by averaging multiple initial and multiple subsequent measurements, then comparing their means. If there are 'n' measurements in each group then Δv will be reduced by $\sqrt{n}$.

	Manual	Assisted
Δv_{inter} (cm3)	0.94	0.47
Δv_{intra} (cm3)	0.64	0.42
n_{inter}	8	2
n_{intra}	4	2

Table 3.6 Minimum lesion volume changes (Δv), and the number of lesions which must be monitored to detect an average volume change of 0.5 cm^3 (n).

Δv_{inter} is the minimum detectable change in lesion volume from two successive quantifications by different operators which are significantly different at the 95% confidence level. Δv_{intra} is the minimum detectable change in lesion volume from two successive quantifications by the same operator which are significantly different at the 95% confidence level. n_{inter} is the number of lesions which must be monitored to detect an average change in volume of 0.5 cm^3 with a power of 80% and a two-sided confidence of 95% when different operators make lesion volume measurements. n_{intra} is the number of lesions which must be monitored when the same operator makes all volume measurements.

Examination of Table 3.6 also reveals that unlike manual quantification, there is little difference between Δv_{inter} and Δv_{intra} for assisted quantification. Thus, the penalty for having different operators make the successive measurements is comparatively small for assisted quantification. Indeed, from Table 3.6 we see that Δv_{inter} for assisted quantification is smaller than Δv_{intra} for manual quantification. This implies that it is better to have "before" and "after" measurements made by two different operators using assisted quantification than it is to have them made by the same operator using manual quantification. These results suggest that assisted quantification may have practical applications in clinical trials, especially those that are multi-center, large or extended over time, and may therefore require lesion measurements by multiple operators.

My results can also be used to estimate the sample size required to detect a desired average change in lesion volume, at a desired power and significance level, using:[48 (p. 182)]

$$n \geq 2 \cdot \frac{(z_\alpha + z_\beta) \cdot SEM}{d} \qquad (3.3)$$

where n is the number of lesions to be monitored; d is the desired mean difference in lesion volume; z_α is the standard normal deviate exceeded in either direction with probability α, where α is the desired significance level (ie. for $\alpha = 0.05$, $z_\alpha = 1.96$); and, z_β is the standard normal deviate exceeded in one direction with probability β where $1-\beta$ is the desired power (ie. for a power of 0.80, $\beta = 0.20$, and $z_\beta = 0.842$). If the same operator makes both measurements, then SEM_{intra} is used in equation 3.3; however, if different operators make the measurements then SEM_{inter} is used. For d = 0.5 cm^3, Table 3.6 lists values for n using $\alpha = 0.05$ and $\beta = 0.20$. It can be observed that substantially fewer lesions must be monitored using assisted quantification to achieve the same significance and power when measuring a 0.5 cm^3 average volume change, compared to manual quantification.

My results indicate that assisted quantification improves operator performance, compared to manual quantification. However, assisted quantification is not intended to replace the operator, rather it is designed to reduce the time and ease the arduous nature of lesion quantification. Assisted quantification still requires an operator with knowledge of normal and pathological anatomy, and to a lesser extent, of tissue contrasts in MR exams.

The operators involved in this experiment were considered members of a random population, and have different levels of experience quantifying MS lesions. Operators 1 and 2 are "experienced" lesion quantifiers, with several years experience, each. Operators 3 and 4 are "intermediate" quantifiers with approximately 1 year of experience each. Operators 5 and 6 are "novice" quantifiers, with a few months experience each. All operators have knowledge of tissue contrasts in

spin-echo MR exams and can identify major tissues such as white matter, grey matter, and CSF. In addition, all operators are "experienced" computer users who easily adapt to new computer programs. Indeed, very little instruction was required before all operators were comfortable with assisted quantification. Most considered it "easier" and "faster" than manual outlining of lesion boundaries.

Detailed comparisons between operators were not made in this chapter, nevertheless, it is interesting to compare the volume measurements plotted in Figure 3.5 between operators in terms of their experience. For manual quantification of Lesion E, the "experienced" operators 2 and 3 made volume measurements that were much larger than those of the other "experienced" and "novice" operators. However, during assisted quantification of lesion E, there was good agreement between all operators, both "experienced" and "novice". In addition, in these experiments, the largest within-operator variability was recorded for an "experienced" operator making manual quantifications (operator 2, Lesion E), while the smallest within-operator variability was recorded for a "novice" operator making assisted quantifications (operator 5, Lesion D). These observations suggest that assisted quantification may help reduce differences between "experienced" and "novice" operators, and may have further benefit for clinical trials involving operators with varying experience.

3.5 Appendix

Here, using terminology presented by Eliasziw et. al.,[47] I present the statistical methods used to estimate inter- and intra-operator reliabilities, ($\hat{\rho}_{inter}$ and $\hat{\rho}_{intra}$), and standard errors of measurement (SEM_{inter} and SEM_{intra}).

Each measured lesion volume can be described using the following additive model:

$$x_{ijk} = \mu + L_i + O_j + LO_{ij} + e_{ijk} \tag{3.4}$$

where x_{ijk} is the k^{th} measured lesion volume by the j^{th} operator on the i^{th} lesion. μ is the overall mean lesion volume; L_i is the effect of lesion i on volume measurement; O_j is the effect of operator j on volume measurement; LO_{ij} is the effect of the interaction between operator j and lesion i on volume measurement; and e_{ijk} is the random error within operator j making the k^{th} measurement of lesion i. These effects are assumed to be independent of each other, have a mean of 0 and variances of σ_L^2, σ_O^2, σ_{LO}^2, and σ_e^2 respectively. A two-way random analysis of variance (ANOVA) was performed on these data. Table 3.7 shows the ANOVA table used to estimate the variance components.[47] By equating the expected mean square expressions to the observed mean squares, it is possible to solve for estimates of each variance component. The resulting equations are listed in Table 3.8. From the estimated variance components, the inter-operator reliability coefficient, $\hat{\rho}_{inter}$, was calculated according to:[47]

$$\hat{\rho}_{inter} = \frac{\hat{\sigma}_L^2}{\hat{\sigma}_T^2} \tag{3.5}$$

where $\hat{\sigma}_T^2$ is the total variance. The intra-operator reliability coefficient, $\hat{\rho}_{intra}$, was calculated using:

$$\hat{\rho}_{intra} = \frac{\hat{\sigma}_L^2 + \hat{\sigma}_O^2 + \hat{\sigma}_{LO}^2}{\hat{\sigma}_T^2} \tag{3.6}$$

These reliability coefficients vary between 0 and 1, with higher values indicating greater reliability. In addition, the intra-operator standard error of measurement (SEM_{intra}) was calculated for each experiment according to:

$$SEM_{intra} = \sqrt{\hat{\sigma}_e^2} \tag{3.7}$$

and the inter-operator SEM (SEM_{inter})was calculated using:

$$SEM_{inter} = \sqrt{\hat{\sigma}_O^2 + \hat{\sigma}_{LO}^2 + \hat{\sigma}_e^2} \tag{3.8}$$

For the patient study described in this chapter, the total standard deviation, σ_T, its contributing components, along with SEM_{inter}, SEM_{intra}, $\hat{\rho}_{intra}$, and $\hat{\rho}_{inter}$ are listed for both manual and assisted quantification in Table 3.5.

The significance of the difference between the standard errors of measurement for manual and assisted quantification were tested using the following statistic:

$$d = \frac{SEM_{Manual}^2 - SEM_{Assisted}^2}{\sqrt{var\left(SEM_{Manual}^2\right) + var\left(SEM_{Assisted}^2\right)}} \tag{3.9}$$

where $var(SEM^2)$ represents the variance in the standard error of measurement. Assuming that the samples contributing to each SEM are from a normal distribution, then the variance in the SEM^2 is given by:[48]

$$var\left(SEM^2\right) = \frac{2 \cdot SEM^4}{DF(SEM)} \tag{3.10}$$

where DF(SEM) is the number of degrees of freedom in the SEM. For SEM_{intra} there are 120 degrees of freedom determined from the product of the number of lesions (6), the number of operators (5) and one fewer than the number of trials (4).[47]

However, determination of the degrees of freedom for SEM_{inter} is not straightforward. It is likely to be large and similar to that of SEM_{intra} since roughly the same number of measures contribute to the calculation of both SEM_{inter} and SEM_{intra}. In this chapter, I have approximated the degrees of freedom in SEM_{inter} by assuming $DF(SEM_{inter}) = DF(SEM_{intra}) = 120$ and have applied the statistic d, above, to determine the significance of the null hypothesis on the difference between inter-operator standard errors of measurement.

Source of Variation	Degrees of Freedom	Observed Mean Square	Expected Mean Square
Lesion	l-1	MS_L	$t \bullet o\sigma_L^2 + t\sigma_{LO}^2 + \sigma_e^2$
Operator	o-1	MS_O	$t \bullet l\sigma_O^2 + t\sigma_{LO}^2 + \sigma_e^2$
Operator x Lesion	(l-1)(o-1)	MS_{LO}	$t\sigma_{LO}^2 + \sigma_e^2$
Intra-operator	lo(t-1)	MS_e	σ_e^2

Table 3.7 Analysis of variance table for the repeated measures experiments

This table has been adapted from one presented by Eliasziw *et. al.* (1994). For the experiments described in this chapter, the number of lesions is l=5, the number of operators is o=6, and the number of trials is t=5. Each measurement contributed to the assessment of both inter- and intra-operator reliability.

Standard Deviation Estimates
$\hat{\sigma}_L = \sqrt{\frac{MS_L - MS_{LO}}{t \bullet o}}$
$\hat{\sigma}_O = \sqrt{\frac{MS_O - MS_{LO}}{t \bullet l}}$
$\hat{\sigma}_{LO} = \sqrt{\frac{MS_{LO} - MS_e}{t}}$
$\hat{\sigma}_e = \sqrt{MS_e}$

Table 3.8 Estimates of the standard deviation components in the repeated measures study.

These estimates were derived by Eliasziw *et. al.* (1994). For the experiments described in this chapter, the number of lesions is l=5, the number of operators is o=6, and the number of trials is t=5.

3.6 References

[1] J. Kesselring, I.E.C. Ormerod, D.H. Miller, E.P.G.H. du Boulay, W.I. McDonald *Magnetic Resonance Imaging in Multiple Sclerosis.* New York, Thieme Medical Publishers, 1-93 (1989).

[2] C. Isaac, D.K. Li, M. Genton, C. Jardine, E. Grochowski, M. Palmer, L.F. Oger, D.W. Paty, "Multiple sclerosis: A serial study using MRI in relapsing patients". Neurology, **38(10)**, 1511-1515 (1988).

[3] E.W. Willowghby, E. Grochowski, J. Oger, L.F. Kastrukoff, D.W. Paty, "Serial magnetic resonance scanning in multiple sclerosis: A second prospective study in relapsing patients". Annals of Neurology, **25**, 43-49 (1989).

[4] R.A. Koopmans, D.K.B. Li, J.J.F. Oger, L.F. Kastrukoff, C. Jardine, L. Costley, S. Hall, E.W. Growchowski, D.W. Paty, "Cronic Progressive Multiple Sclerosis: Serial Magnetic Resonance Brain Imaging Over Six Months". Annals of Neurology, **26**, 248-256 (1989).

[5] A.J. Thompson, A.G. Kermode, D.G. MacManus, et al. "Patterns of disease activity in multiple sclerosis: clinical and magnetic resonance imaging study". British Medical Journal, **300**, 631-634 (1990).

[6] J.O. Harris, J.A. Frank, N. Pattronas, D.E. McFarlin, H.F. McFarland, "Serial Gadolinium-enhanced Magnetic Resonance Imaging Scans in Patients with Early, Relapsing-Remitting Multiple Sclerosis: Implications for Clinical Trials and Natural History". Annals of Neurology **29**, 548-444 (1991).

[7] M.D. Weinshenker, R.N. Bass, S.J. Karlik, G.C. Ebers, G.P.A. Rice, "An oper trial of OKT3 in patients with multiple sclerosis." Neurology 41, 1047 (1991).

[8] D.W. Paty, D.K.B. Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group "Interferon beta-1b is effective in relapsing-remitting multiple sclerosis". Neurology, **43**, 662-667 (1993).

[9] L. Kappos, D Stadt, M. Ratzka, et al. "Magnetic resonance imaging in the evaluation of treatment in multiple sclerosis". Neuroradiology, **30**, 299-302 (1988).

[10] J.F. Kurtzke, "Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS)". Neurology **33**, 1444-1452 (1983).

[11] J.H. Noseworthy, M.K. Vandervoort, C.J. Wong, G.C. Ebers, "Interrater variability on MS scales in a blinded, prospective clinical trial" In: *Recent Advances in MS Therapy.* Eds. Gonsette R.E., and Dolmotte P., New York, Elsevier Scientific (1989).

[12] I.E.C. Ormerod, D.H. Miller, W.I. McDonald, et al. "The role of NMR imaging in the assessment of multiple sclerosis and isolated neurological lesions; a quantitative study". Brain, **110**, 1579-1616 (1987).

[13] S.J. Karlik, M.K. Vandervoort, M. Hopkins, J.H. Noseworthy "Serial quantitative brain MRI in multiple sclerosis using visual and volumetric techniques". Neurology, 40, 142 (1990).

[14] D.W. Paty, M. Bergstrom, M. Palmer, J. MacFadyen, D. Li, "A quantitative magnetic resonance image of the multiple sclerosis brain". Abstract. Neurology **35(1)**, 137 (1985).

[15] S.S. Gebarski, T.O. Gabrielsen, S. Gilman, J.E. Knake, J.T. Latack, A.M. Aisen, "The initial diagnosis of multiple sclerosis: clinical impact of magnetic resonance imaging". Annals of Neurology, **17**, 469-474 (1985).

[16] D.E. Goodkin, J.S. Ross, S. Vanderbrug-Medendorp, J. Konecsni, R.A. Rudick, "Magnetic resonance imaging lesion enlargement in multiple sclerosis". Archives of Neurology, **49**, 261-263 (1992).

[17] D.W. Paty, C.D. Isaac, E. Grochowski, *et al.* "Magnetic resonance imaging (MRI) in multiple sclerosis (MS): a serial study in relapsing and remitting patients with quantitative measurements of lesion size". Abstract. Neurology **36(suppl 1)**, 177 (1986).

[18] R.A. Koopmans, D.B.K. Li, J.J.F. Oger, J. Mayo, D.W. Paty, "The lesions of multiple sclerosis: Imaging of acute and chronic stages". Neurology, **39**, 959-963 (1989).

[19] D.W. Paty, "Multiple Sclerosis with an emphasis on MR imaging". Current Neurology, **11**, 169-198 (1991).

[20] G.D. Cramer, D.J. Allen, L.J.A. DiDio, W. Potvin, R. Brinker, "Evaluation of encephalic ventricular volume from the magnetic resonance imaging scans of thirty-eight human subjects". Sur. Radiol. Anat. **12**, 287-290 (1990).

[21] A.H. Young, D.H.R. Blackwood, H. Roxborough, J.K. McQueen, M.J. Martin, D. Kean, "A magnetic resonance imaging study of schizophrenia: brain structure and clinical symptoms" British Journal of Psychiatry, **158**, 158-164 (1991).

[22] J.P. Kesslak, O. Nalcioglu, C.W. Cotman, "Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease". Neurology, **41**, 51-54 (1991).

[23] G.W. Hynd, M. Semrud-Clikeman, A.R. Lorys, E.S. Novey, D. Eliopulos, H. Lyytinin, "Corpus callosum morphology in attention deficit-hyperactivity disorder: morphometric analysis of MRI". Journal of Learning Disabilities, **24**, 141-146 (1991).

[24] C.K. Liu, B.L. Miller, J.L. Cummings, et al. "A quantitative MRI study of vascular dementia". Neurology, **42**, 138-143 (1992).

[25] F. Pannizzo, M.J.B. Stallmeyer, J. Friedman, *et al.* "Quantitative MRI studies for assessment of multiple sclerosis". Magnetic Resonance in Medicine **24**, 90-99 (1992).

[26] C.R. Jack, M.D. Bentley, C.K. Twomey, A.R. Zinsmeister, . "MR imaging-based volume measurements of the hippocampal formation and anterior temporal lobe: validation studies'. Radiology, 176: 205-209 (1990).

[27] R.C. Gur, P.D. Mozley, S.M. Resnick, *et. al.* "Gender differences in age effect on brain atrophy measured by magnetic resonance imaging". *P*roceedings of the National Academy of Science, 88: 2845-2849 (1991).

[28] M.E. Shenton, R. Kikinis, F.A. Jolesz, *et. al.* "Abnormalities of the left temporal lobe and thought disorders in schizophrenia." The New England Journal of Medicine, 327: 604-612 (1992).

[29] M.Bomans, K.-H. Hohne, U.Tiede, M. Reimer, "3D segmentation of MR images of the head for 3D display". IEEE Transactions of Medical Imaging, **9(2)**, 177-183 (1990).

[30] D.N. Kennedy, P.A. Filipek, V.S. Caviness Jr., "Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging". IEEE Transactions of Medical Imaging, **8(1)**, 1-7, March (1989).

[31] K.O. Lim, A. Pfefferbaum, "Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter". Journal of Computer Assisted Tomography, **13(4)**, 588-593 (1989).

[32] T.L. Jernigan, G.A. Press, J.R. Hesselink, "Methods for measuring brain morphologic features on magnetic resonance images". Archives of Neurology, **47**, 27-32 (1990).

[33] R.T. Fan, S.S. Trivedi, L.L. Fellingham, A. Gamboa-Aldeco, "Soft tissue segmentation and 3D display from computerized tomography and magnetic resonance imaging". Proceedings SPIE: Medical Imaging, 494-504 (1987).

[34] M.W. Vannier, R.L. Butterfield, D. Jordan, W.A. Murphy, R.G. Levitt, M. Gado, "Multispectral analysis of magnetic resonance images". Radiology **154**, 221-224 (1985).

[35] M.W. Vannier, R.L. Butterfield, D.L. Rickman, D.M. Jordan, W.A. Murphy, P.R. Biondetti, "Multispectral magnetic resonce image analysis". CRC Critical Reviews in Biomedical Engineering, **15(2)**. 117-144 (1987).

[36] J.R. Mitchell, S.J. Karlik, D.H. Lee, A. Fenster, "Multi-spectral analysis and visualization of multiple sclerosis lesions in MR volumes of the brain." SPIE: Medical Imaging V **1898**, 442-452 (1993).

[37] G. Gerig, J. Martin, R. Kikinis, O. Kubler, M. Shenton, F.A. Jolesz, "Automating segmentation of dual-echo MR head data". *Information Processing in Medical Imaging - 12th International Conference*, Colchester, A.C.F., and Hawkes, D. J., editors. London, Springer-Verlag, 175-187 (1991).

[38] R. Kikinis, M.E. Shenton, G. Gerig, et. al., "Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging". *Journal of Magnetic Resonance Imaging*, 2, 619-629 (1992).

[39] M.I. Kohn, N.K. Tanna, G.T. Herman, *et. al.* "Analysis of brain and cerebrospinal fluid volumes with MR imaging". Radiology, 178, 115-122 (1991).

[40] O. Kubler, G. Gerig, "Segmentation and Analysis of Multidimensional Data-Sets in Medicine". NATO ADI, Series F60, 63 (1990).

[41] M. Ozkan, H.G. Sprenkels, B.M. Dawant, "Multi-spectral magnetic resonance image segmentation using neural networks". *Proceedings of the International Joint Conference on Neural Networks*, 429-434, June (1990).

[42] L.M. Fletcher, J.B. Barsotti, J.P. Hornak, "A multispectral analysis of brain tissues". Magnetic Resonance in Medicine, **29**, 623-630 (1993).

[43] R.O. Duda, P.E. Hart, *Pattern classification and scene analysis.* New York, NY: Wiley, 1973; 45-49 (1973).

[44] I.E. Cline, W.E. Lorensen, R. Kikinis, F. Jolesz, "Three-dimensional segmentation of MR images of the head using probability and connectivity". Journal of Computer Assisted Tomographpy, **14(6)**, 1037-1045, Nov./Dec (1990).

[45] J.R. Mitchell, S.J. Karlik, D.H. Lee, A. Fenster, "Computer Assisted Identification and Quantification of Multiple Sclerosis Lesions in MR Volumes of the Brain". Journal of Magnetic Resonance Imaging, 4, 197-208 (1994).

[46] D.H. Lee, A.D. Vellet, M. Eliasziw, L. Vidito, G.C. Ebers, G.P. Rice, L. Hewett, S. Dunlavy, "Magnetic Resonance Imaging Field Strength: Prospective Evaluation of the Diagnostic Accuracy of MR for the diagnosis of Multiple Sclerosis at 1.5T and 0.5T". Radiology, 194(1), 256-262, (1995).

[47] M. Eliasziw, S.L. Young, M.G. Woodbury. K. Fryday-Field, "Statistical Methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example." Physical Therapy, 74, 777-788 (1994).

[48] P. Armitage, G. Berry, *Statistical methods in medical research.* 2nd edition. Oxford: Blackwell Scientific Publications, 1-559 (1990).

[49] G.Bartzokis, J.Mintz, P.Marx, et. al. "Reliability of in vivo volume measures of hippocampus and other brain structures using MRI". Magnetic Resonance Imaging, **11**, 993-1006 (1993).

CHAPTER 4 QUANTIFICATION OF MULTIPLE SCLEROSIS LESION VOLUMES IN 1.5T AND 0.5T ANISOTROPICALLY FILTERED AND UNFILTERED MR EXAMS

4.1 Introduction

Soon after demonstrating the utility of MRI in the visualization and assessment of MS, a number of investigators used MRI to monitor disease progression.[1-15] Recently, guidelines for using MRI to monitor MS treatment trials have been established.[16] Those guidelines recommend use of imaging systems with mid field (0.5T to 1.0T) or high field (greater tnan 1.0T) strengths. Indeed, performing MR examinations of MS patients at high field strengths has many advantages. Primary among these is the higher inherent signal-to-noise ratio (SNR) at high field strengths. However, recent improvements in hardware and software have improved image quality at both high and mid field strengths. For example, application of post-processir _ algorithms, which reduce noise without blurring or moving object boundaries, has improved image SNR.[17,18]

These improvements, along with the reduced cost of installing and operating mid-field imagers prompted a prospective evaluation of the accuracy of MRI for the diagnosis of MS at high field and mid field strengths.[19] In that study, the authors found no difference in the accuracy, sensitivity, or specificity of the diagnosis of MS between current mid field and high field imagers. In this chapter, I use patient images acquired in the previous study to examine the impact of field strength and the application of post-processing noise reduction algorithms on the quantification of MS lesions. Specifically, I evaluate the impact on operator reliability and variability of: a) lesion quantification in high-field (1.5T) versus mid-field (0.5T) exams; and, b) a filter algorithm modelled after anisotropic diffusion which reduces image noise and increases contrast, without blurring or moving object boundaries.

4.2 Methods

Experiments were performed to measure the effect of magnetic field strength and anisotropic filtering on the inter- and intra-operator variability of lesion quantification in MR exams of a single MS patient. Operator accuracy was not measured, since the actual lesion volumes were not available.

4.2.1 Magnetic field strength.

Images were acquired on both 1.5T and 0.5T GE Signa imaging systems (GE Medical Systems, Milwaukee) within 24 hours using the multi-echo imaging protocols listed in Table 4.1. These protocols were constructed as part of a study assessing the accuracy of MR at 0.5T and 1.5T for diagnosis of MS.[19] The protocols were similar in the following sense: both were acquired on highly similar imaging systems, except for field strength; and, total imaging time was nearly identical in each case, with constant TR. The bandwidth was optimized for both the early and late echoes in the 0.5T exams. Slice positioning between imaging systems was maintained to within one slice thickness by first acquiring a series of sagittal images using the "localizer" protocol listed in Table 4.1. Next, the inferior margins of the genu and splenium of the corpus callosum were identified in these images, and used to establish the oblique axial slice orientation and position.[19] Finally, the "multi-echo" protocol listed in Table 4.1 was used to acquire proton density (PD) and T2 weighted images at each oblique axial slice in the patient's brain. Figure 4.1 shows PD and T2 weighted images acquired from the same slice using both 1.5T and 0.5T imaging systems. All images were transferred via Ethernet to Sparc workstations (Sun Microsystems, Mountainview Calif.) for processing. Images were then converted to 8 bits of dynamic range for display and further processing by dividing the original exam intensities by 4. Next, PD and T2 weighted 3-D images were created from each exam by stacking all early echo images and all later echo images, respectively.[20]

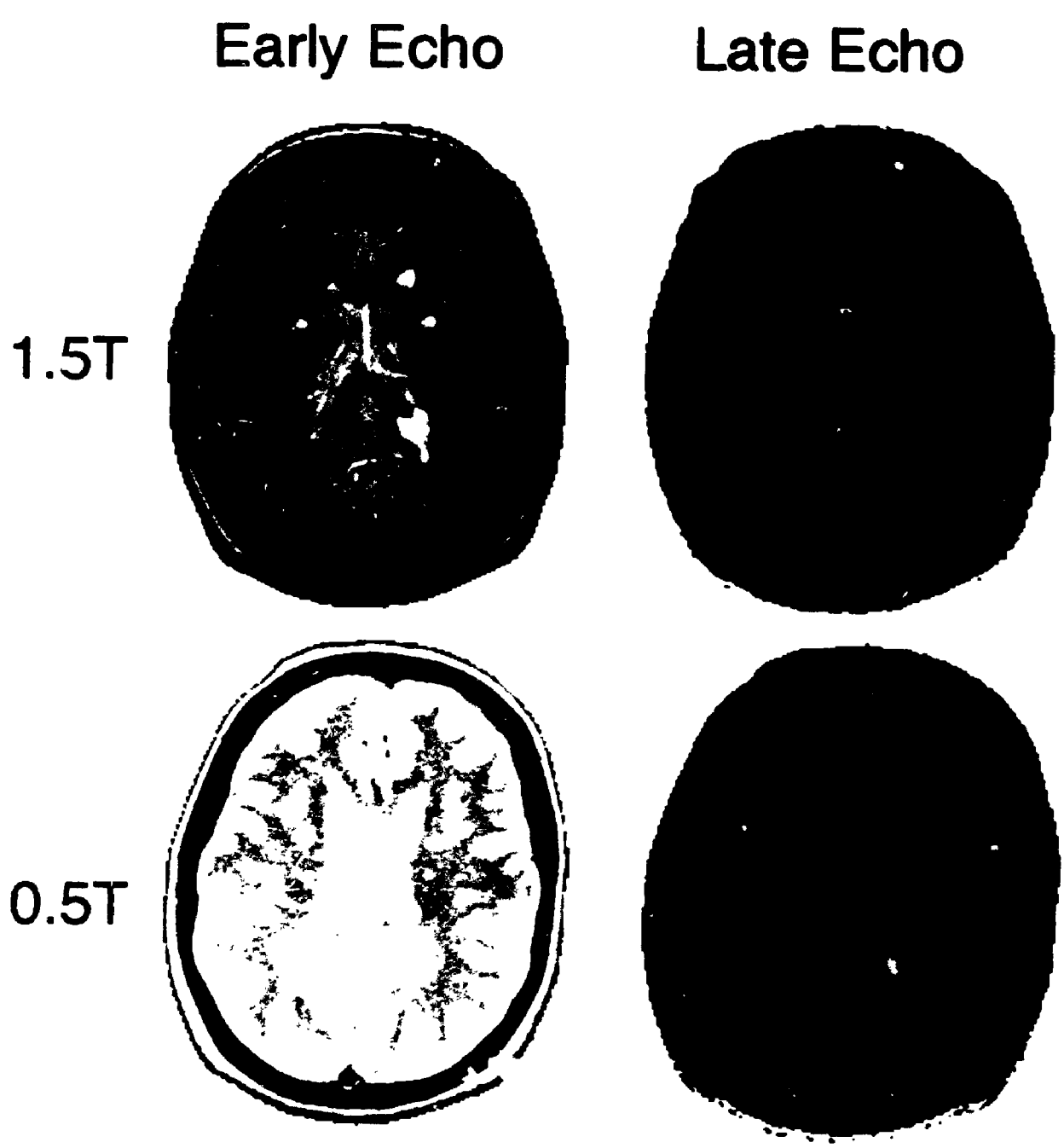


Figure 4.1 1.5T and 0.5T early echo (proton-density weighted) and late echo (T2 weighted) images from the same slice in the brain of an MS patient.

The 1.5T images were acquired at echo times of 30 and 80 ms, while the 0.5T images were acquired at echo times of 33 and 90 ms. In addition, the bandwidth was reduced for both 0.5T echoes. Other imaging parameters, including total scan time, were maintained between the two field strengths.

	1.5T Localizer	1.5T Multi-echo	0.5T Localizer	0.5T Multi-echo
Image Type	T1 Weighted	T2 Weighted	T1 Weighted	T2 Weighted
Image Plane	Sagittal	Axial-oblique	Sagittal	Axial-oblique
Field-of-view (cm)	22	22	22	22
TR (ms)	400	2800	400	2800
Te1/Te2 (ms)	20	30/80	20	33/90
Number of slices	9	36	9	36
Slice thickness (mm)	5	5	5	5
Slice gap (mm)	0	2.5	0	2.5
Matrix size	256 x 128	256 x 192	256 x 128	256 x 192
NEX	1	1	1	1
Bandwidth (kHz)	16	16/8	5.33	10.67/3.76
Scan time	0:56	9:21	0:58	9:31

Table 4.1 Imaging parameters used in the field strength experiments.

4.2.2 Anisotropic filtering.

Both the 1.5T and 0.5T images described above were filtered to reduce noise without blurring edges, using a process modelled on anisotropic diffusion. This filter was initially formulated by Perona and Malik[17], then extended by Gerig *et. al*[18] for smoothing MR images which are multi-spectral or 3D. Briefly, the filter uses edge detection and multi-scale noise reduction to smooth within regions, but not across boundaries. Edge detection and smoothing are controlled by the selection of 2D or 3D diffusion, the degree of connectedness (4 or 8 for 2D diffusion, 6 or 26 for 3D diffusion), the diffusion function, and the number of filter iterations.

For these experiments I used multi-channel diffusion since PD and T2 weighted images were available at each slice. Diffusion was 2D since the inter-plane voxel dimension (slice thickness) was large compared to the in-plane voxel dimension. Finally, 4-connectedness was used to simplify the diffusion calculations. The diffusion function, described elsewhere,[17] was:

$$c(\hat{x}, t) = \exp\left(-\left[\frac{|\nabla I(\hat{x}, t)|}{K}\right]^2\right) \tag{4.1}$$

where $\hat{x}$ is the spatial coordinate vector, t is the iteration number, $\nabla I(\hat{x}, t)$ is the image intensity gradient, and K is a user adjustable parameter which provides information about image noise levels. Gerig *et. al.*[18] suggest that:

$$1.5 \cdot \sigma_{noise} < K < 2.0 \cdot \sigma_{noise} \tag{4.2}$$

for MR images where σ_{noise} is an estimate of the image noise obtained in each exam from 5 regions of homogenous CSF in the third and lateral ventricles using the following equation:

$$\sigma_{noise} = \sqrt{\frac{\sum_{i=1}^{5}\left(\sigma_{PD,i}^2 + \sigma_{T2,i}^2\right)}{10}} \tag{4.3}$$

where $\sigma_{PD,i}$ and $\sigma_{T2,i}$ are the standard deviations in intensity measured in the ith region from the PD and T2 images, respectively. Based on these estimates, K was set to 6 for 1.5T exams, and to 17 for 0.5T exams. Finally, following the suggestion of Gerig *et. al.*,[18] the number of filter iterations was set to 3, as a compromise between optimal noise reduction, and execution time. Figure 4.2 shows filtered and unfiltered T2 weighted images acquired from the same slice using both 1.5T and 0.5T imaging systems.

To determine the impact of anisotropic filtering on exam signal and noise I measured the standard deviation and average signal intensity in unfiltered and filtered 1.5T and 0.5T exams. Measurements were made in regions of white matter (WM), grey matter (GM), cerebro-spinal fluid (CSF), MS lesion, and air from the same locations in each exam. Measurement regions were elliptical and consisted of 50 voxels in the case of WM, GM, CSF and air, and 34 voxels in lesion. Regions were, therefore, large enough to provide adequate samples for standard

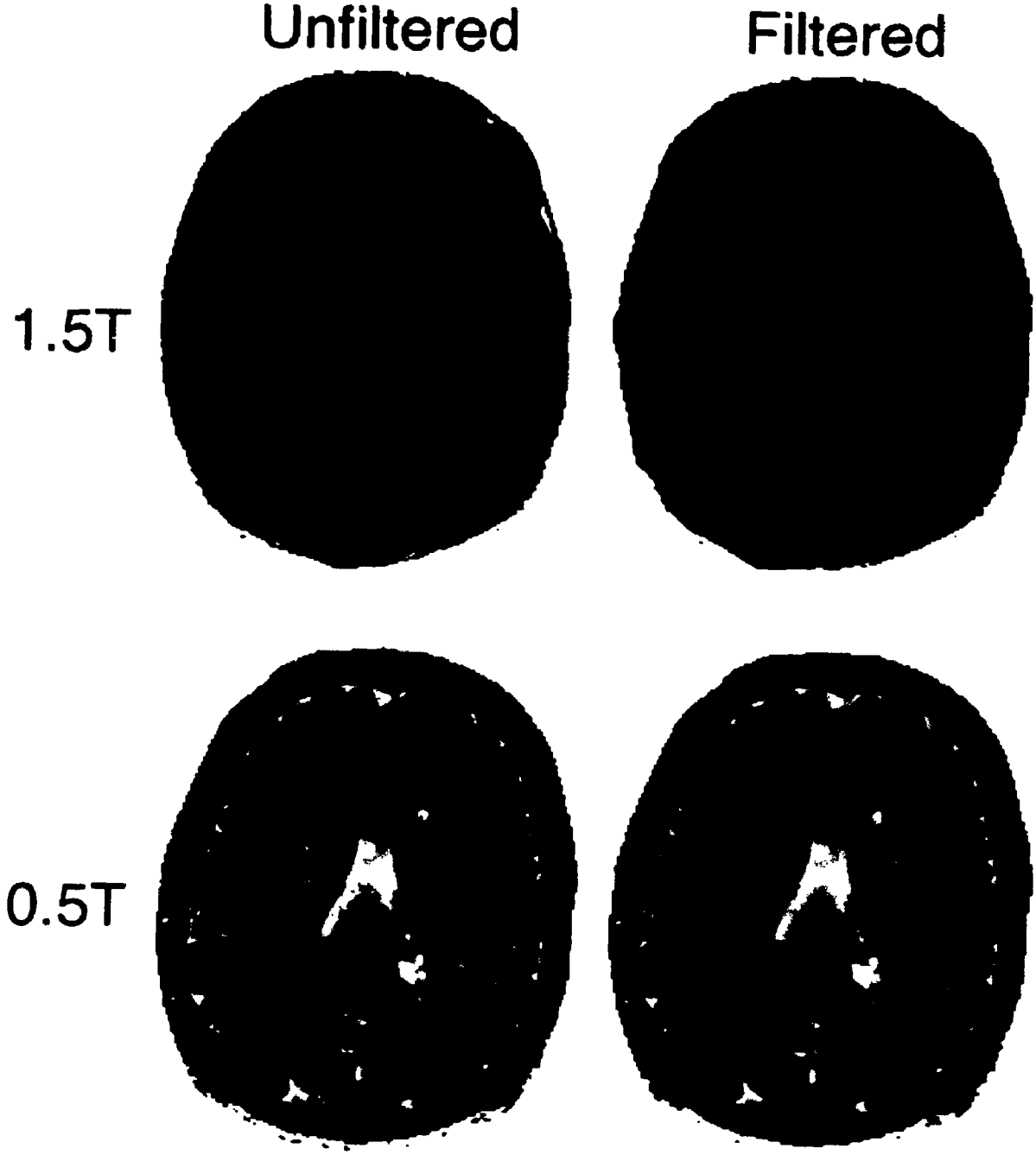


Figure 4.2 Filtered and unfiltered T2 weighted images acquired at 1.5T and 0.5T from the same slice in the brain of an MS patient.

Estimates of image noise were acquired in the ventricular CSF, then provided as a parameter to the anisotropic filter algorithm. The filter smooths within regions but does not move or blur object boundaries.

deviation calculation, but small enough to avoid non-uniformities due to radiofrequency inhomogeneities or normal tissue variation.[18] Results of signal and noise measurements are listed in Tables 4.2 and 4.3 for 1.5T and 0.5T exams, respectively.

	Early Echo				Late Echo			
	Unfiltered		Filtered		Unfiltered		Filtered	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ
WM	126.4	2.8	126.5	1.7	59.0	2.8	59.2	1.8
GM	152.3	2.1	152.1	1.6	81.0	3.2	80.8	3.0
CSF	149.4	3.7	149.5	2.8	128.8	3.0	128.9	2.4
Lesion	180.3	4.5	180.4	4.1	132.5	3.8	132.3	3.3
Air	2.0	1.4	1.8	0.6	2.1	1.3	2.1	0.6

Table 4.2 The mean signal and standard deviations of various brain tissues in 1.5T unfiltered and anisotropically filtered exams.

The mean signals were measured from small homogenous regions-of-interest in each exam. The same regions-of-interest were used in both filtered and unfiltered exams. Lesion signals were measured in the large lesion A (see Figure 4.3)

	Early Echo				Late Echo			
	Unfiltered		Filtered		Unfiltered		Filtered	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ
WM	151.8	5.7	151.8	2.0	69.7	4.3	69.9	2.4
GM	192.7	6.0	192.5	2.7	104.8	4.7	104.5	3.5
CSF	171.8	4.8	172.5	2.8	151.1	4.0	150.8	2.2
Lesion	243.8	4.8	243.3	2.7	182.8	7.1	181.5	4.0
Air	5.9	2.6	6.1	1.2	3.7	1.9	3.7	0.6

Table 4.3 The mean signal and standard deviations of various brain tissues in 0.5T unfiltered and anisotropically filtered exams.

The mean signals were measured from small homogenous regions-of-interest in each exam. The same regions-of-interest were used in both filtered and unfiltered exams. Lesion signals were measured in the large lesion A (see Figure 4.3).

Anisotropic filtering of MR data was implemented as part of a more extensive analysis program I developed which includes 3D multi-spectral analysis and visu-

alization.[20] The system was implemented on Sun SparcStation 2 computers (Sun Microsystems, Mountainview Calif.) running SunOS, and Openwindows (Sun Microsystems, Mountainview Calif.). The system was written in the C language and uses the Xview Toolkit (Sun Microsystems, Mountainview Calif.).

4.2.3 Manual Lesion Quantification.

Operators used my system to adjust the contrast of either image (using window and level operations), zoomed the images (up to ten times magnification), and then used the mouse interactively to outline lesion boundaries in all the slices in which the lesion was visible. Detailed adjustment of lesion boundaries was possible using the editing procedures provided by the system. The volume of each lesion was determined by identifying its boundary in multiple slices, and summing the voxels entirely contained within each boundary and then multiplying by the voxel volume.

4.2.4 Study Design.

The study design is based on one I developed to determine the variability of manual and computer assisted quantification of MS lesions.[21] Five lesions, shown in Figure 3.4 on page 67 and labelled A through E, were selected for the operator studies. Figure 4.3 shows detailed views from the 1.5T and the 0.5T unfiltered and filtered exams of each lesion. The lesions were deliberately selected to vary in volume (from a minimum of approximately 0.2 cm^3 to a maximum of approximately 2.7 cm^3) and vary from focal (Lesion D) to diffuse (Lesion E), and thus represent the range of lesions encountered in patient exams. Four experiments were performed to measure inter- and intra-operator reliability of MS lesion quantification. Each experiment represented a combination of the following two variables: 1) field strength (1.5T or 0.5T); and, 2) filter technique (anisotropically filtered, or not). In each experiment, six operators made 5 repeated measures of each of the 5 lesions over a 6 month period. In total, 600 lesion volumes were quantified (i.e. 6 operators x 5 repeated measurements x 5 lesions x 4

experiments). Measurements were obtained from exams in the following order: 1)1.5T; 2) 0.5T; 3) 1.5T filtered; and, 4) 0.5T filtered.

In each experiment, a 2-way analysis of variance (ANOVA) on the repeated measures was used to estimate the total variance (σ_T^2) and its components due to: differences between lesions (σ_L^2); differences between operators (σ_O^2); the interaction between operators and lesions (σ_{LO}^2); and, the random error within the typical operator (σ_e^2). Based on methods described by Eliasziw *et. al.*,[22] these variance components were used to estimate the inter-operator reliability, $\hat{\rho}_{inter}$, using:

$$\hat{\rho}_{inter} = \frac{\hat{\sigma}_L^2}{\hat{\sigma}_T^2} \tag{4.4}$$

while an estimate of the intra-operator reliability, $\hat{\rho}_{intra}$, was calculated using:

$$\hat{\rho}_{intra} = \frac{\hat{\sigma}_L^2 + \hat{\sigma}_O^2 + \hat{\sigma}_{LO}^2}{\hat{\sigma}_T^2} \tag{4.5}$$

The inter-operator standard error of measurement, SEM_{inter}, a measure of the variability between operators, was determined using:

$$SEM_{inter} = \sqrt{\hat{\sigma}_O^2 + \hat{\sigma}_{LO}^2 + \hat{\sigma}_e^2} \tag{4.6}$$

while the intra-operator standard error of measurement, SEM_{intra}, a measure of the variability within the typical operator, was determined using:

$$SEM_{intra} = \sqrt{\hat{\sigma}_e^2} \tag{4.7}$$

The inter- and intra-operator reliabilities and variabilities are listed in Table 4.4. SEM's from different experiments were also tested to determine if they were statistically different.

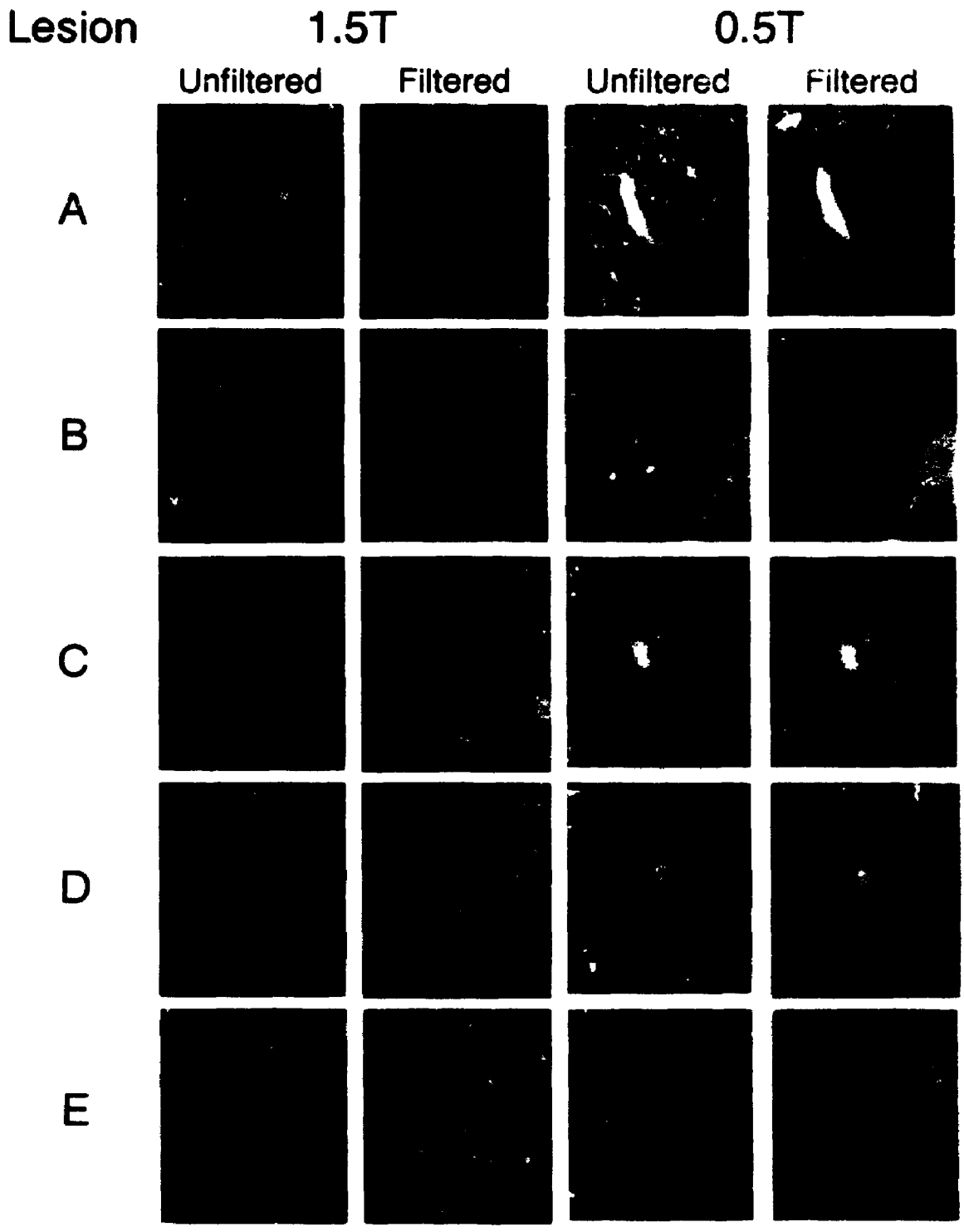


Figure 4.3 Detailed views from the 1.5T and 0.5T unfiltered and filtered exams of lesions A through E.

These images are close-ups of each lesion taken in the second echo at each field strength. The lesions range from large (A) to small (B) and from focal (D) to diffuse (E). Lesion C is a ventricular cap, while lesions B and D are deep white-matter lesions. Lesions were selected to cover the range of lesions which may be visible in MR exams of MS patients. All these lesions are visible in at least two exam slices.

	1.5T		0.5T	
	Unfiltered	Filtered	Unfiltered	Filtered
$\hat{\sigma}_T$ (cm^3)	1.16	1.13	1.00	1.05
$\hat{\sigma}_L$ (cm^3)	1.11	1.10	0.97	1.02
$\hat{\sigma}_O$ (cm^3)	0.14	0.09	0.13	0.13
$\hat{\sigma}_{LO}$ (cm^3)	0.21	0.17	0.12	0.10
$\hat{\sigma}_e$ (cm^3)	0.23	0.19	0.19	0.18
SEM_{inter} (cm^3)	0.34	0.27	0.26	0.24
$\hat{\rho}_{inter}$	0.91	0.94	0.93	0.95
SEM_{intra} (cm^3)	0.23	0.19	0.19	0.18
$\hat{\rho}_{intra}$	0.96	0.97	0.96	0.97

Table 4.4 The reliability and variability of lesion quantification in 1.5T and 0.5T filtered and unfiltered exams.

Estimates of total standard deviation ($\hat{\sigma}_T$), reliability ($\hat{\rho}$), and standard error of measurement (SEM) for quantification in unfiltered and filtered exams acquired at 1.5T and 0.5T. Also shown are the components of total standard deviation due to lesions ($\hat{\sigma}_L$), operators ($\hat{\sigma}_O$), the interaction between lesions and operators ($\hat{\sigma}_{LO}$), and the random variability within operators ($\hat{\sigma}_e$). Units are cm^3 except for the reliability coefficients (ρ), which are dimensionless

4.2.5 Spatial error distribution.

The impact of field strength and anisotropic filtering on the variability of lesion boundary determination was examined by analyzing the spatial distribution of the operator variabilities. My system was used to mark and record voxels labelled as "lesion" in each experiment described above, after which the selection frequency, F(x,y,z), for a single voxel with coordinates (x,y,z) was determined using the following formula:

$$F(x, y, z) = \frac{N(x, y, z)}{N_T}, \quad 1 \le N(x, y, z) \le N_T \tag{4.8}$$

where N(x,y,z) was the number of times the voxel with coordinates (x,y,z) was determined to be "lesion", and N_T was the total number of measurements in each experiment. The spatial distribution of voxel selection frequencies was coded by a color-scale where white represents 100% selection and black represents 3% selection. The spatial distribution is plotted in Figure 4.4 for quantifications of all five lesions in unfiltered and filtered 1.5T and 0.5T exams.

4.3 Results

4.3.1 Anisotropic Filtering.

Tables 4.2 and 4.3 list the mean signals and standard deviations measured from homogenous regions of WM, GM, CSF, lesion and air in unfiltered and filtered 1.5T and 0.5T exams, respectively. These tables reveal that anisotropic filtering has little impact on the mean measured signals of the tissues, or air, as expected. However, filtering does reduce the standard deviation of both tissue and air. In particular, filtering reduced the standard deviation in air by approximately 65% in both 0.5T and 1.5T exams.

4.3.2 Lesion Volume Measurement.

Figure 4.5 shows histograms of the average lesion volumes measured by each operator in the 4 experiments quantifying the effect of field-strength and filter method on operator variability and reliability. Comparison of volume measurements between operators reveals that volume measurements ranged from a maximum of 3.21 cm^3 (1.5T-Unfiltered, Lesion A, Operator 5) to a minimum of 0.15 cm^3 (1.5T-Filtered, Lesion B, Operator 6) with a mean of 1.30 cm^3. The individual standard errors of measurement ranged from a maximum of 0.27 cm^3 (1.5T-Unfiltered, Lesion E, Operator 2) to a minimum of 0 cm^3 (0.5T-Filtered, Lesion D, Operator 6). Measurements of small focal lesions B and D were highly consistent across operators, while measurements of the large diffuse lesion E were highly

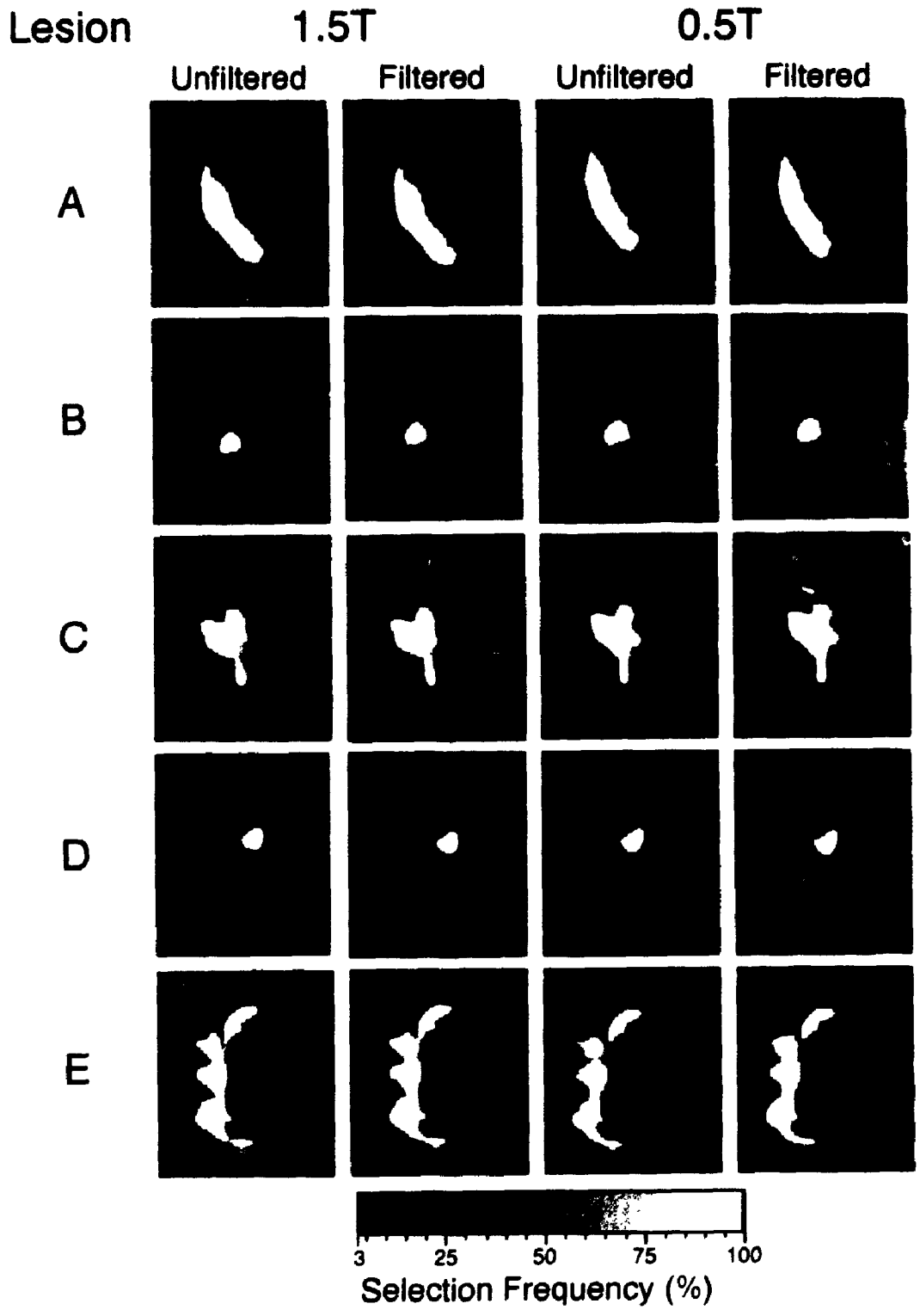


Figure 4.4 Detailed views showing the frequency of voxel selection for quantification of lesions A to E in 1.5T and 0.5T unfiltered and filtered exams.

Color values indicate the number of times each voxel was labelled lesion by 6 operators performing 5 trials each. Dark regions correspond to voxels labelled as lesion in 3% of the trials, while white regions correspond to regions always labelled lesion.

variable both between and within operators. Overall, quantification in 0.5T exams provided slightly smaller lesion volume estimates than quantification in 1.5T exams, with the average difference being 0.06 cm^3. Even though this difference is small, representing less than 5% of the mean lesion volume, it has a p-value of 0.0002. This difference may be due to slight differences in patient registration between exams. However, there was no significant difference between the mean lesion volumes measured in unfiltered and filtered exams.

Table 4.4 lists the estimated total standard deviation, ($\hat{\sigma}_T^2$) along with its contributing components of standard deviation, for each experiment. The total standard deviation ranged from 1.00 cm^3 (0.5T-Unfiltered) to 1.16 cm^3 (1.5T-Unfiltered). In each experiment the lesion component ($\hat{\sigma}_L^2$) was the largest contributor to total standard deviation. This component ranged from 0.97 cm^3 (0.5T-Unfiltered) to 1.11 cm^3 (1.5T-Unfiltered). The large lesion component reflects the differences between lesions selected for analysis. The operator ($\hat{\sigma}_O^2$) and operator-lesion ($\hat{\sigma}_{LO}^2$) components are small in comparison, having a maximum value of 0.21 cm^3 (operator-lesion component, 1.5T-Unfiltered). The intra-operator standard deviation ($\hat{\sigma}_\varepsilon^2$) ranged from 0.18 cm^3 (0.5T-Filtered) to 0.23 cm^3 (1.5T-Unfiltered).

The standard deviation components listed in Table 4.4 were used to calculate both the inter-operator standard error of measurement (SEM_{inter}) and the intra-operator standard error of measurement (SEM_{intra}), using equations (4.6) and (4.7). Both SEM_{inter} and SEM_{intra} are listed in Table 4.4 for each experiment. SEM_{inter} is a measure of the variability between operators in each experiment, while SEM_{intra} is a measure of the variability within the typical operator in each experiment. In general, SEM_{inter} is larger than SEM_{intra} since operators are more likely to agree with themselves than with each other. In my experiments, SEM_{inter} ranged from 0.24 cm^3 (0.5T-Filtered) to 0.34 cm^3 (1.5T-Unfiltered), and SEM_{intra} ranged from 0.18 cm^3 (0.5T-Filtered) to 0.23 cm^3 (1.5T-Unfiltered).

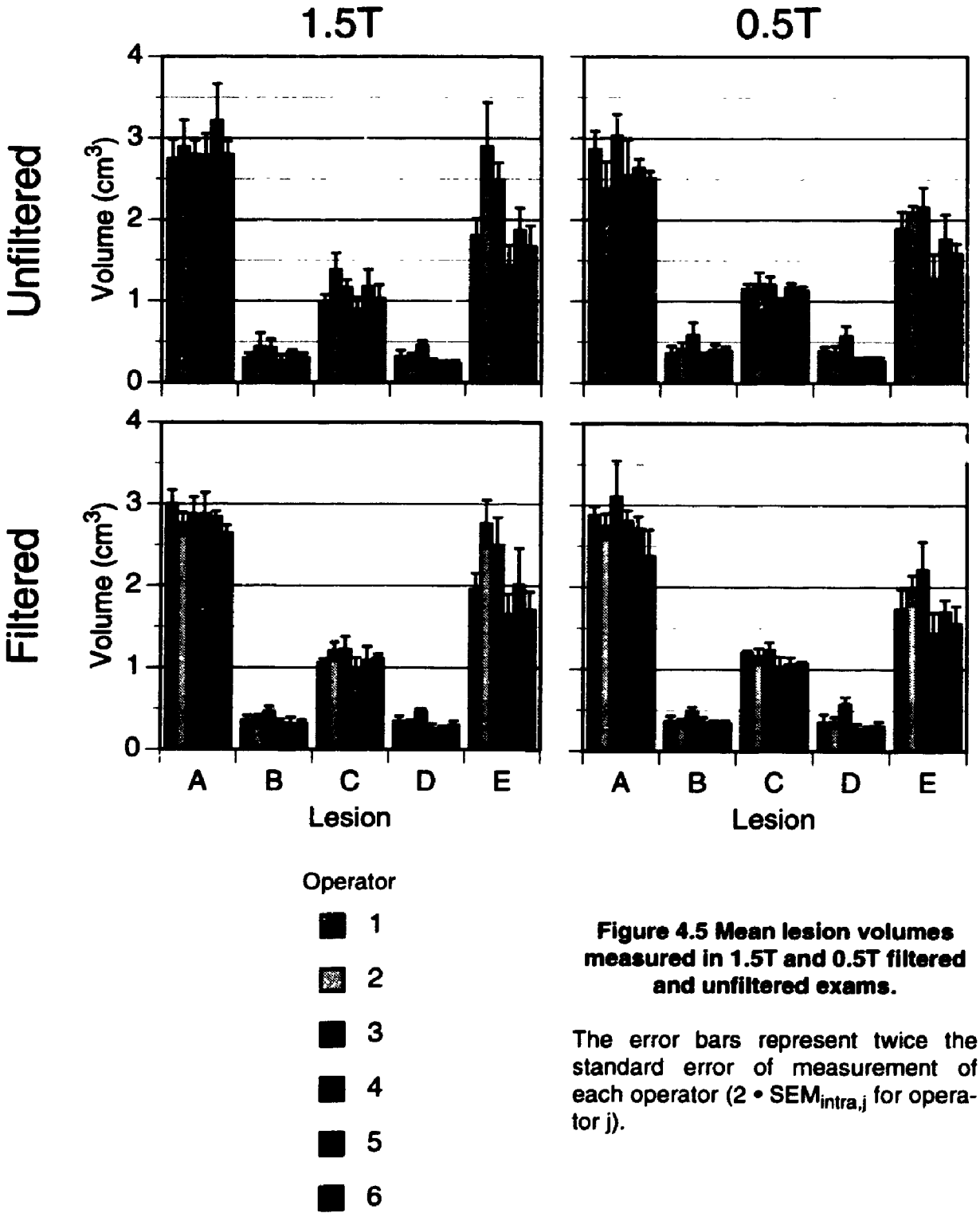


Figure 4.5 Mean lesion volumes measured in 1.5T and 0.5T filtered and unfiltered exams.

The error bars represent twice the standard error of measurement of each operator ($2 \cdot SEM_{intra,j}$ for operator j).

Comparison of SEMs in Table 4.4 between field strengths and between unfiltered and filtered exams revealed that SEM_{inter} in 1.5T unfiltered exams (0.34 cm^3) was significantly higher than SEM_{inter} in any other experiment (maximum p-

value of 0.016). In addition, there was no significant difference between SEM_{inter} in either 1.5T filtered (0.27 cm^3), 0.5T unfiltered (0.26 cm^3), or 0.5T filtered exams (0.24 cm^3). Similarly, SEM_{intra} in 1.5T unfiltered exams (0.23 cm^3) was significantly higher than in any other experiment (maximum p-value of 0.037), and there was no difference between SEM_{intra} in either 1.5T filtered (0.19 cm^3), 0.5T unfiltered (0.19 cm^3), or 0.5T filtered exams (0.18 cm^3).

The standard deviation components listed in Table 4.4 were also used to determine the reliability coefficients, ρ_{inter} and ρ_{intra} using equations (4.4) and (4.5). These coefficients may be regarded as correlation coefficients, and can range from 0 to 1.0, with higher values corresponding to increased reliability.[23] In my experiments ρ_{inter} ranged from 0.91 (1.5T-Unfiltered) to 0.95 (0.5T-Filtered), while ρ_{intra} was 0.96 in unfiltered exams, and 0.97 in filtered exams. The high reliabilities in these experiments are largely due to the diverse sizes and types of lesions selected for quantification. The similarity of the inter-operator reliabilities indicates that the operators were largely interchangeable when measuring the lesions in these experiments. The similarity of the intra-operator reliabilities indicates that the variability within the typical operator was small compared to the diversity of the lesions.

4.3.3 Spatial error distribution.

Figure 4.4 shows the frequency of voxel quantification for lesions A through E in the four experiments. In these visualizations, a sharper transition between pure white (region always quantified) and the gray-scale MR image (region never quantified) indicates better operator performance. In each experiment the lesions are surrounded by regions of low quantification frequency. The large lesions C and E have a "smooth" transition between frequently and infrequently quantified regions, while the large lesions A has sharp transitions, except in the immediate periventricular regions. The small lesions B and D also have sharp transitions. Lesion C contains two focal regions with high quantification frequency, and a third

diffuse region quantified in less than half of the operator trials. Unlike lesion A, this region is further away from the ventricle surface. The large, diffuse, multi-focal lesion E contains a number of regions with high quantification frequency. Each of these regions has a smooth transition to low quantification frequency, indicating that in some trials operators considered the focal regions of lesion E to be connected and in others to be disjoint. In some experiments lesion E also has large posterior and anterior "lobes" with low quantification frequency (< 50%), indicating operator variability is high in these regions.

Comparison of the frequency distributions in 1.5T and 0.5T exams from Figure 4.4 reveals that quantification is quite similar at both field strengths. However, quantification in 0.5T exams reduced the large anterior "lobes" of low quantification frequency in lesions A and E, and reduced the large posterior "lobes" in lesions C and E. This reduction results in lower volume estimates for these large lesions in 0.5T exams. There is very little difference in the frequency distributions of small lesions B and D between field strengths. Comparison of the frequency distributions in unfiltered and anisotropically filtered exams from Figure 4.4 reveals little effect on the quantification of small lesions B and D. However, filtering increased the quantification frequency of anterior "lobes" in lesions A, C and E.

Although useful for visualizing the spatial distribution of operator variability, the images of Figure 4.4 are not quantitative. Quantitative analysis of quantification frequencies for lesion E is shown in Figure 4.6. These graphs show the volume (number of voxels) at each quantification frequency for the four experiments. Under ideal conditions, every lesion voxel would be identified by all operators all of the time. Therefore, under ideal conditions these graphs would consist of a single point on the ordinate axis. In reality, variability within and between operators results in some voxels being selected with lower frequency, which in turn causes spreading of the volume distribution over lower frequencies. Thus, in these graphs high ordinate intersection and reduced volumes at frequencies below 100% indicate better operator performance. Comparison of the graphs in Figure

4.6 between field strengths reveals that quantification in 0.5T exams reduced volumes selected at frequencies below 30% (Figure 4.6b and 4.6d). In addition, quantification in filtered exams increased the ordinate intersection at both field strengths (Figure 4.6a and 4.6c). These effects are likely responsible for the reduced operator variabilities in filtered and 0.5T exams.

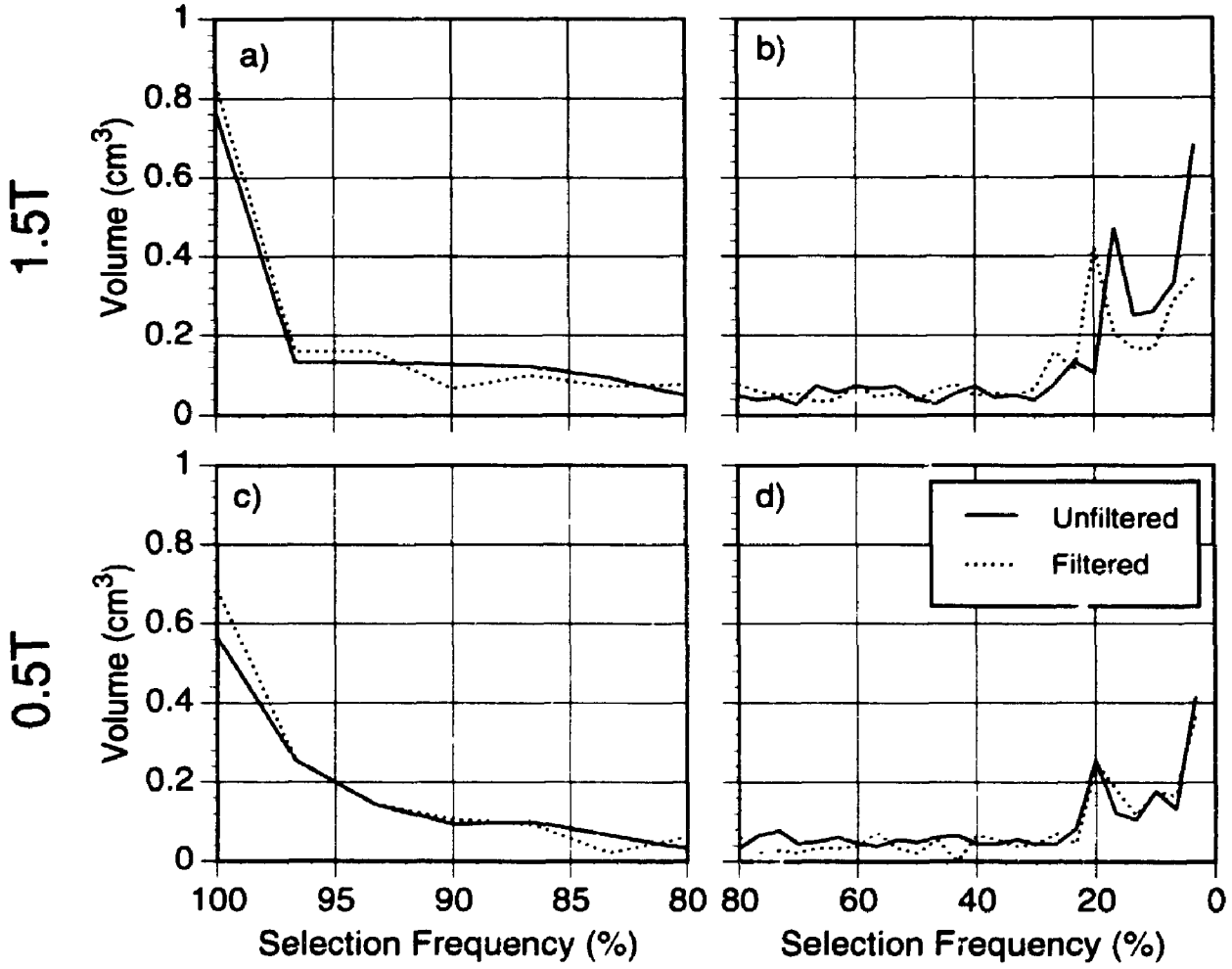


Figure 4.6 The volume selected at each quantification frequency in 1.5T and 0.5T exams.

Histograms showing the volume (number of voxels) of lesion E selected in 1.5T and 0.5T exams. a) and c) are detailed plots of volumes selected in more than 80 percent of the trials in 1.5T and 0.5T exams, respectively. b) and d) are plots of the volumes selected in less than 80% of the trials in 1.5T and 0.5T exams, respectively. In each graph the solid line indicates the volume at each frequency for unfiltered exams. and the dashed line indicates the volume at each frequency for anisotropically filtered exams.

4.4 Discussion

The patient study presented in this chapter has examined the impact of field strength, and anisotropic filtering on the reliability and variability of MS lesion quantification. Results indicate that inter- and intra-operator variabilities in 1.5T unfiltered exams were significantly larger than those in 1.5T filtered, 0.5T unfiltered, and 0.5T filtered exams. However, the statistical and experimental methods presented have not addressed the accuracy of quantification. The impact on accuracy of these variables is difficult to establish since true lesion volumes are difficult to determine. Nevertheless, knowledge of the precision of lesion quantification is essential for clinical trials concerned with changes in lesion volumes in response to therapy. The minimum change in lesion volume, Δv, which can be confidently detected depends on the desired confidence level and operator variability.[22] Table 4.5 lists Δv calculated at the 95% confidence level for each experiment based on the data in Table 4.4, when either the same operator or two different operators make successive volume measurements. When the same operator makes both measurements (Δv_{intra}), the minimum detectable change ranges from 0.64 cm^3 (1.5T-Unfiltered) to 0.50 cm^3 (0.5T-Filtered). When different operators make successive measurements (Δv_{inter}), the minimum detectable change ranges from 0.94 cm^3 (1.5T-Unfiltered) to 0.67 cm^3 (0.5T-Filtered). These values are largely independent of the lesions selected in my experiments, and therefore can be used to estimate the quantification variability of lesions with volumes between 0.3 cm^3 and 3.0 cm^3, which are measured in exams acquired using the protocol listed in Table 4.1. If Δv is too large for application in a clinical trial, it can be reduced by $\sqrt{n}$ by comparing the means of 'n' "before" to 'n' "after" measurements.

My results can also be used to estimate the number of lesions which must be monitored to detect a desired average change in lesion volume, at a desired power and significance level, using:[24, p. 182]

	1.5T		0.5T	
	Unfiltered	Filtered	Unfiltered	Filtered
Δv_{inter} (cm^3)	0.94	0.75	0.72	0.67
Δv_{intra} (cm^3)	0.64	0.53	0.53	0.50
n_{inter}	8	5	5	4
n_{intra}	4	3	3	3

Table 4.5 Minimum lesion volume changes (Δv), and the number of lesions which must be monitored needed to detect a volume change of 0.5 cm^3 (n).

Δv_{inter} is the minimum detectable change in lesion volume from two successive quantifications by different operators which are significantly different at the 95% confidence level. Δv_{intra} is the minimum detectable change in lesion volume from two successive quantifications by the same operator which are significantly different at the 95% confidence level. n_{inter} is the number of lesions which must be monitored to detect an average change in volume of 0.5 cm^3 with a power of 80% and a two-sided confidence of 95% when different operators make lesion volume measurements. n_{intra} is number of lesions to be monitored when the same operator makes all volume measurements.

$$n \geq 2\left(\frac{(z_\alpha + z_\beta) \cdot SEM}{d}\right)^2 \qquad (4.9)$$

where n is the number of lesions to be monitored; d is the desired mean difference in lesion volume; z_α is the standard normal deviate exceeded in either direction with probability α, where α is the desired significance level (ie. for $\alpha = 0.05$, $z_\alpha = 1.96$); and, z_β is the standard normal deviate exceeded in one direction with probability β where $1-\beta$ is the desired power (ie. for a power of 0.80, $\beta = 0.20$, and $z_\beta = 0.842$). If the same operator makes both measurements, then SEM_{intra} is used in equation (4.9); however, if different operators make the measurements then SEM_{inter} is used. For d = 0.5 cm^3, Table 4.5 lists values for n using $\alpha = 0.05$ and $\beta = 0.20$. It can be observed that fewer lesions must be monitored when measuring a 0.5 cm^3 average volume change with the same significance and power in 0.5T exams, compared to 1.5T exams. In addition, anisotropic filtering reduces the number of lesions which must be monitored, in 1.5T exams and when different operators make measurements in 0.5T exams.

4.4.1 Anisotropic Filtering.

Examination of Figure 4.4 reveals a number of effects of anisotropic filtering on lesion quantification. For example, in 1.5T exams filtering increased the voxel selection frequency of the "low-contrast lobes" on the anterior of Lesion E, increasing operator variability in this region. However, filtering in these exams also caused a slight increase in the frequency of selection within the interior of lesions A,C and E, and a slightly sharper transition along the perimeter of all 5 lesions, lowering operator variability. In 0.5T exams, filtering increased the selection frequency of the "low-contrast lobes" on the anterior of Lesion E, and resulted in a few voxels selected in lesions A, and C where none occurred in unfiltered exams. My tests to determine the significance of the difference between SEM's showed that filtering had no impact on SEM's in 0.5T exams, but significantly reduced both SEM_{inter} and SEM_{intra} in 1.5T exams. Thus, the overall effect of filtering in 1.5T exams was to reduce operator variability, while the effect in 0.5T exams was less clear.

These results may be due to three important characteristics of the anisotropic diffusion filter, which are to: 1) increase the detail signal-to-noise ratio (see equation (4.10)); 2) sharpen edges; and, 3) preserve edge locations.[18] The first two characteristics should aid differentiation of lesion from non-lesion regions, and reduce operator variability. However, it appears that in some situations, filtering caused operators to select lesion regions they otherwise would not, and also increased their variability in regions of low selection frequency. The third filter characteristic maintains object boundaries, and therefore should not change the mean measured lesion volume. This was observed in these experiments where no significant difference in lesion volume resulted between filtered and unfiltered exams.

4.4.2 Field Strength.

Although signal-to-noise ratio is improved at higher field strengths, Table 4.4 reveals that for quantification in unfiltered exams, both inter- and intra-operator variability are less at 0.5T than at 1.5T. This result may be due to differences in the detail signal-to-noise ratios (dSNRs) between various brain tissues at each field strength. The dSNR between two tissues, A and B, is defined as:[25,26]

$$dSNR = \frac{|S_A - S_B|}{\sigma} \tag{4.10}$$

where S_A and S_B are the mean image intensities measured from homogenous regions inside tissues A and B respectively, and σ is an estimate of image noise. In this experiment, I used the standard deviation of signal in air as an estimate for σ.[27] Table 4.6 lists the dSNRs between WM, GM, CSF and lesion calculated using the values listed in Tables 4.2 and 4.3 for unfiltered 1.5T and 0.5T exams. Since filtering did not affect the mean signal but did reduce similarly the standard deviation in air (i.e. σ) in all exams, the dSNRs were increased similarly in filtered exams and therefore are not tabulated here.

	Early echo		Late echo	
	1.5T	0.5T	1.5T	0.5T
WM - GM	18.5	15.7	16.9	18.5
WM - CSF	16.4	7.7	53.7	42.8
WM - Lesion	38.5	35.4	56.5	59.5
GM - CSF	2.1	8.0	36.8	24.4
GM - Lesion	20.0	19.7	39.6	41.1
CSF - Lesion	22.1	27.7	2.8	16.7

Table 4.6 The detail signal-to-noise ratios (dSNRs) between various brain tissues in 1.5T and 0.5T unfiltered exams.

dSNRs were calculated using values listed in Tables 4.2 and 4.3. An estimate of image noise was obtained by using the standard deviation of signal in air.

Examination of Table 4.6 reveals that dSNRs can vary between the early and late echoes in each exam. For some tissues, dSNR is considerably higher in one echo than the other. For example, at both field strengths, CSF-Lesion dSNR is higher in the early echo, while GM-CSF dSNR is higher in the late echo. This is important since operators may use information in either or both echoes while performing tissue quantification, and higher dSNRs should reduce operator variability when measuring tissue volumes.

Examination of Table 4.6 also reveals that dSNRs vary between the two field strengths. For example, in the late echo, the CSF-Lesion dSNR is much higher at 0.5T than at 1.5T, while in the early echo the WM-CSF dSNR is higher at 1.5T. If all other factors remain constant, then signal-to-noise ratio (SNR), and thus dSNR, should be proportional to field strength.[28] Therefore, imaging at 1.5T rather than 0.5T should result in a 3 fold increase in dSNR. However, Table 4.1 reveals that the imaging parameters were not constant between the 1.5T and 0.5T exams. In particular, the bandwidth was reduced for the 0.5T exams. Since SNR is inversely proportional to the square-root of the bandwidth,[28] the reduction in bandwidth will increase SNR, and therefore increase dSNR. The dependence of dSNR on field strength is further complicated because T1 increases with increasing field strength for most tissues. In the imaging protocols listed in Table 4.1, TR was fixed between field strengths. The longer tissue T1's in 1.5T exams result in less longitudinal recovery between imaging experiments at the higher field strength, and reduced dSNR. This effect may be particularly acute for dSNRs between CSF and other tissues, since the T1 of CSF is very long compared to other tissues.

In these experiments, the WM-Lesion, GM-Lesion and CSF-Lesion dSNRs are of particular interest since they indicate how difficult it is to differentiate MS lesion from normal tissue. Table 4.6 reveals that there is very little difference in the dSNRs of WM-Lesion and GM-Lesion between field strengths, in either echo. Thus, the image information available to operators when differentiating lesion

from WM and GM should be similar at both field strengths. However, the CSF-Lesion dSNR is higher in the 0.5T exams in both echoes. This is particularly important since most lesions occur in the periventricular white matter. Indeed, in these experiments both Lesions A and C border ventricles (see Figure 4.3). The improved dSNR between these lesions and ventricular CSF in 0.5T exams may be responsible in part, for the lower operator variability at the lower field strength.

4.5 References

[1] C. Isaac, D.K. Li, M. Genton, C. Jardine, E. Grochowski, M. Palmer, L.F. Oger, D.W. Paty, "Multiple sclerosis: A serial study using MRI in relapsing patients". Neurology, **38(10)**, 1511-1515, (1988).

[2] E.W. Willowghby, E. Grochowski, J. Oger, L.F. Kastrukoff, D.W. Paty, "Serial magnetic resonance scanning in multiple sclerosis: A second prospective study in relapsing patients". Annals of Neurology, **25**, 43-49, (1989).

[3] R.A. Koopmans, D.B.K. Li, J.J. Oger, J. Mayo, D.W. Paty, "The lesions of multiple sclerosis: Imaging of acute and chronic stages". Neurology, **39**, 959-963, (1989).

[4] J.O. Harris, J.A. Frank, N. Pattronas, D.E. McFarlin, H.F. McFarland, "Serial Gadolinium-enhanced Magnetic Resonance Imaging Scans in Patients with Early, Relapsing-Remitting Multiple Sclerosis: Implications for Clinical Trials and Natural History". Annals of Neurology **29**, 548-444, (1991).

[5] M.D. Weinshenker, R.N. Bass, S. Karlik, G.C. Ebers, G.P.A. Rice, "An open trial of OKT3 in patients with multiple sclerosis." Neurology 41, 1047-1052, (1991).

[6] D.W. Paty, D.K.B. Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group "Interferon beta-1b is effective in relapsing-remitting multiple sclerosis". Neurology, **43**, 662-667, (1993).

[7] D.E. Goodkin, J.S. Ross, S. Vanderbrug-Medendorp, J. Konecsni, R.A. Rudick, "Magnetic resonance imaging lesion enlargement in multiple sclerosis". Archives of Neurology, **49**, 261-263, (1992).

[8] Paty, D. W., Bergstrom, M., Palmer, M., MacFadyen, J. , Li, D. "A quantitative magnetic resonance image of the multiple sclerosis brain". Abstract. Neurology **35(1)**, 137, (1985).

[9] D.W. Paty, C.D. Isaac, E. Grochowski, *et. al.* "Magnetic resonance imaging (MRI) in multiple sclerosis (MS): a serial study in relapsing and remitting patients with quantitative measurements of lesion size". Abstract. Neurology **36(suppl 1)**, 177, (1986).

[10] I.E.C. Ormerod, D.H. Miller, W.I. McDonald, *et. al.* "The role of NMR imaging in the assessment of multiple sclerosis and isolated neurological lesions; a quantitative study". Brain, **110**, 1579-1616, (1987).

[11] L. Kappos, D. Stadt, M. Ratzka, *et. al.* "Magnetic resonance imaging in the evaluation of treatment in multiple sclerosis". Neuroradiology, **30**, 299-302, (1988).

[12] J. Kesselring, I.E.C. Ormerod, D.H. Miller, E.P.G.H. du Boulay, W.I. McDonald *Magnetic Resonance Imaging in Multiple Sclerosis*. New York, Thieme Medical Publishers, 1-93, (1989).

[13] R.A. Koopmans, D.K.B. Li, J.J.F. Oger, L.F. Kastrukoff, C. Jardine, L. Costley, S. Hall, E.W. Growchowski, D.W. Paty, "Cronic Progressive Multiple Sclerosis: Serial Magnetic Resonance Brain Imaging Over Six Months". Annals of Neurology, **26**, 248-256, (1989).

[14] A.J. Thompson, A.G. Kermode, D.G. MacManus, *et. al.* "Patterns of disease activity in multiple sclerosis: clinical and magnetic resonance imaging study". British Medical Journal, **300**, 631-634, (1990).

[15] F. Pannizzo, M.J.B. Stallmeyer, J. Friedman, *et al.* "Quantitative MRI studies for assessment of multiple sclerosis". Magnetic Resonance in Medicine, **24**, 90-99, (1992).

[16] D.H. Miller, F. Barkhof, I. Berry, L. Kappos, G. Scotti, A.J. Thompson A.J. "Magnetic Resonance Imaging in Monitoring the Treatment of Multiple Sclerosis: Concerted Action Guidelines." Journal of Neurology Neurosurgery and Psychiatry, **54**, 683-688, (1991).

[17] P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion." IEEE Transactions on Pattern Analysis and Machine Intelligence, **12**, 629-639, (1990).

[18] G. Gerig, O. Kübler, R. Kikinis F.A. Jolesz. "Nonlinear Anisotropic Filtering of MRI Data." IEEE Transactions on Medical Imaging, **11 (2)**, 221-232, (1992).

[19] D.H. Lee, A.D. Vellet, M. Eliasziw, L. Vidito, G.C. Ebers, G.P. Rice, L. Hewett, S. Dunlavy, "Magnetic Resonance Imaging Field Strength: Prospective Evaluation of the Diagnostic Accuracy of MR for the diagnosis of Multiple Sclerosis at 1.5T and 0.5T". Radiology,194(1), 256-262, (1995).

[20] J.R. Mitchell, S.J. Karlik, D.H. Lee, A. Fenster, "Computer Assisted Identification and Quantification of Multiple Sclerosis Lesions in MR Volumes of the Brain". Journal of Magnetic Resonance Imaging, 4, 197-208, (1994).

[21] J.R. Mitchell, S.J. Karlik, D.H. Lee, M. Eliasziw, G.P. Rice, A. Fenster. "The Variability of Manual and Computer Assisted Quantification of Multiple Sclerosis Lesion Volumes." submitted to Medical Physics, (1995).

[22] M. Eliasziw, S.L. Young, M.G. Woodbury, K. Fryday-Field. "Statistical Methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example". Physical Therapy, 74, 777-788, (1994).

[23] G. Bartzokis, J. Mintz, P. Marx, *et. al.* "Reliability of in vivo volume measures of hippocampus and other brain structures using MRI". Magnetic Resonance Imaging, **11**, 993-1006, (1993).

[24] P. Armitage, G. Berry. *Statistical methods in medical research.* 2nd edition. Oxford: Blackwell Scientific Publications, 1-559, (1990).

[25] J.M. Boone, G.S. Shaber, M. Tecotzky. "Dual-energy mammography: A detector analysis." Medical Physics, **17 (4)**, 665-675, (1990).

[26] H.N. Cardinal, D.W. Holdsworth, M. Drangova, B.B. Hobbs, A. Fenster. "Experimental and theoretical x-ray imaging performance comparison of iodine and lanthanide contrast agents." Medical Physics, **20 (1)**, 15-31, (1993).

[27] R.M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images". Medical Physics, 12(2), 232 (1985).

[28] W.A. Edelstein, G.H. Glover, C.J. Hardy, R.W. Redington. "The Intrinsic Signal-to-Noise Ration in NMR Imaging." Magnetic Resonance in Medicine, **3**, 604-618, . (1986).

CHAPTER 5 MAGNETIC RESONANCE MULTISPECTRAL ANALYSIS OF MULTIPLE SCLEROSIS LESIONS

5.1 Introduction

Magnetic Resonance Imaging (MRI) provides very sensitive detection of the lesions of Multiple Sclerosis (MS).[1,2] Indeed, quantification of changes in total lesion volume in MR exams has been used as an indicator of disease activity in clinical trials evaluating new therapies.[3-13] However, the quantification of total lesion volume alone may not indicate completely the spatial and temporal pathological changes of active lesions.[14,15] Furthermore, the ability to discriminate between potentially reversible and irreversible tissue damage may be essential for the development of effective therapies.[7] MR spectroscopy and spectroscopic imaging show great potential for assessing reversibility in MS lesions.[16-18]Arnold *et al*[16] have demonstrated that the ability to study changes in metabolites in voxels by MR spectroscopy provides important information about the functional variability of lesions. However, at the present time, MR spectroscopy has limited resolution, which increases partial volume effects and limits its ability to determine compositional variability within lesions.

In addition to changes in metabolites, there are complex changes to the T2 spectrum of white matter (WM) occurring with the onset and progression of MS. Differences between the T2 spectral distributions of normal appearing WM and MS lesions have been reported in the long T2 range,[19] in the intermediate (~100msec) T2 range,[20,21] and more recently, in the short (10-50msec) T2 range.[22,20] Analysis of these T2 spectral distributions, therefore, may lead to a more specific functional measure of the disease than the mostly morphological information provided by current volume measurement techniques.

In this chapter I present a technique to analyze standard spin-echo MR exams of MS patients, and provide information about the underlying T2 spectrum of

lesions and other tissues. This technique is based upon multispectral analysis of these exams,[23,24] and has the advantage of utilizing the inherent T2 information without sacrificing high spatial resolution. I constructed a test phantom containing materials with known MR characteristics, and used it to test the theoretical basis of my technique. I then used my technique to analyze and visualize white matter, cerebrospinal fluid, and MS lesions in serial MR exams of a chronic-progressive, and two relapsing/remitting MS patients.

5.2 Materials and Methods

5.2.1 Multispectral Analysis: Theory

Many MR imaging protocols are inherently multispectral because they can provide multiple images with different contrast characteristics of the same anatomy. The multiple images may be used to construct a 2-dimensional (2-D) histogram, or feature space,[23,24] describing the frequency of intensity combinations in the MR images. MR exams of MS patients commonly involve acquisition of dual echo, spin-echo exams consisting of an early "proton-density" (PD) weighted echo, and a later T2 weighted echo.[25] The intensity at each location with coordinates (x,y) in the feature space constructed from these exams is proportional to the number of image locations with the PD intensity, x, on the abscissa, and the T2 intensity, y, on the ordinate axis.[26] Distinct tissues in the exam images will correspond, ideally, to distinct clusters in the feature space. However, radio-frequency nonlinearities and partial volume effects can cause blurring of tissue clusters.[27]

If we assume a perfect slice profile, and ignore motion and diffusion effects, then for a "pure" tissue with cluster mean (x,y) imaged such that the repetition time is much greater than the echo time, the early echo intensity, x, can be approximated using:

$$x = \rho \cdot \left(1 - e^{\frac{-TR}{T1}} \right) \cdot e^{\frac{-TE_1}{T2}} \tag{5.1}$$

where ρ is the relative tissue proton density, T1 and T2 respectively are the mean spin-lattice and spin-spin relaxation times, TR is the repetition time, and TE_1 is the echo time. Similarly, the later echo intensity, y, can be approximated using:

$$y = \rho \cdot \left(1 - e^{\frac{-TR}{T1}} \right) \cdot e^{\frac{-TE_2}{T2}} \tag{5.2}$$

where TE_2 is the second echo time. One can use these equations to derive an expression for the distance, d, between the cluster mean, (x,y),and the origin of feature space:

$$d = \sqrt{x^2 + y^2} \tag{5.3}$$

Using equations (5.1) and (5.2) to expand equation (5.3) gives:

$$d^2 = \rho^2 \cdot \left(1 - e^{\frac{-TR}{T1}} \right)^2 \cdot e^{\frac{-2TE_1}{T2}} + \rho^2 \cdot \left(1 - e^{\frac{-TR}{T1}} \right)^2 \cdot e^{\frac{-2TE_2}{T2}} \tag{5.4}$$

If one assumes that TE_2 is greater than TE_1, then one can define ΔTE such that:

$$TE_2 = TE_1 + \Delta TE \tag{5.5}$$

Substituting this expression into equation (5.4) and simplifying gives:

$$d^2 = \rho^2 \cdot \left(1 - e^{\frac{-TR}{T1}} \right)^2 \cdot e^{\frac{-2TE_1}{T2}} \cdot \left(1 + e^{\frac{-2\Delta TE}{T2}} \right) \tag{5.6}$$

and therefore:

$$d = x \cdot \sqrt{1 + e^{\frac{-2\Delta TE}{T2}}} \tag{5.7}$$

For most tissues imaged in MR exams of MS patients, TR is much greater than T1, TE_1 is much less than T2, and the early echo intensity, x, is heavily weighted by proton density (i.e. $x \cong \rho$) Thus, for tissues imaged with dual spin-echo exams in which TR is much greater than either TE, the distance between the cluster mean and the origin of feature space is primarily determined by the tissue's average T2, ρ, and the delay between the echoes, ΔTE.

One can also derive an expression for the angle, θ, formed between the abscissa and the line connecting the origin of feature space and the cluster mean:

$$\theta = \operatorname{atan}\left(\frac{y}{x}\right) \tag{5.8}$$

Substituting equations (5.1) and (5.2) into (5.8) and simplifying gives:

$$\theta = \operatorname{atan}\left(\frac{e^{\frac{-TE_2}{T2}}}{e^{\frac{-TE_1}{T2}}}\right) \tag{5.9}$$

which can be further simplified using equation (5.5) to yield:

$$\theta = \operatorname{atan}\left(e^{\frac{-\Delta TE}{T2}}\right) \tag{5.10}$$

Thus, the angle, θ, between the abscissa and the cluster mean is determined solely by ΔTE and the average tissue T2. The relationships between ρ, T2, and the location in feature space are shown graphically in Figure 5.1, where I have plotted iso-ρ and iso-T2 contours over feature space.

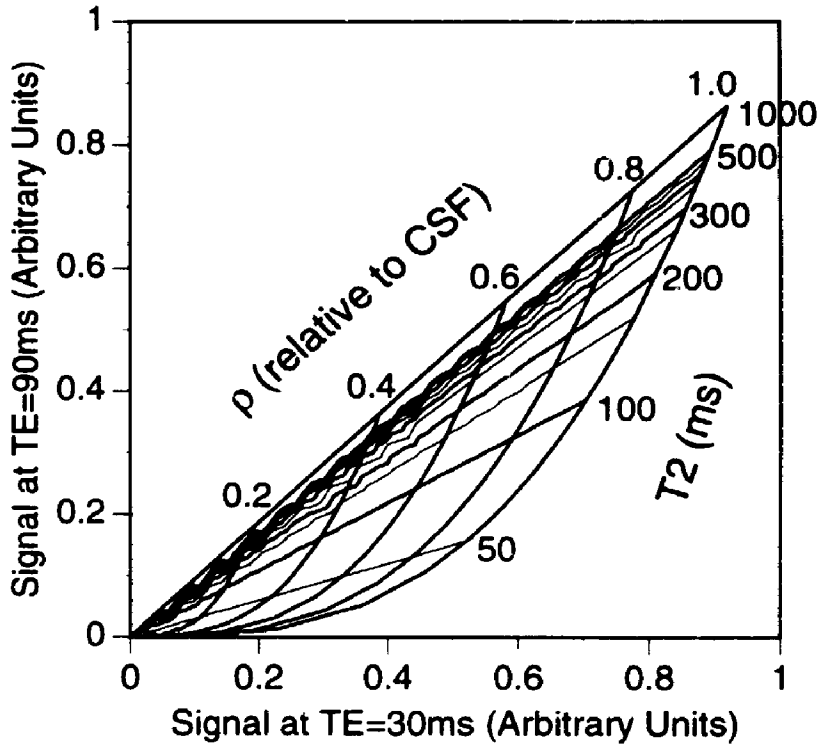


Figure 5.1 iso-ρ and iso-T2 contours in feature space.

These data were generated using equations (5.1) and (5.2) and assuming constant T1 << TR. The abscissa and ordinate represent the spin-echo intensities from echoes at TE_1=30ms, and TE_2=90ms, respectively. The contours indicate regions of constant T2 and constant proton density (relative to CSF).

5.2.2 Multispectral Analysis: Phantom Study

The effects of T1, T2 and ΔTE on cluster location in feature space were examined using a test phantom consisting of materials with known T1 and T2 values in the physiological range. Phantom construction was based on work reported by Kraft,[28] and Mitchell,[29] and consisted of nine 40ml glass tubes each filled with an agarose gel and sealed to prevent deterioration over time. The T2 and T1 in each tube were controlled, respectively, by varying the concentration of the agarose gelling agent in each material, and by doping with varying concentrations of nickel ion, Ni^{2+}. Phantoms constructed in this manner have T1 and T2 values that are very stable between 10 and 50°C, between 0.1 Tesla and 4 Tesla, and over periods of one year or more.[28,29]

The T1 and T2 times of each material were measured at 20°C with a 1.5T relaxometer (Nalorac, Martinez CA/SMIS, Surrey England). T2 times were measured using a nonnegative least squares algorithm to fit 600 points acquired using

a CPMG sequence.[30] T1 values were measured by fitting 35 points acquired using an inversion-recovery sequence with logarithmically spaced inversion times. Table 5.1. lists the T1 and T2 times measured for each phantom material,

Material	T2(ms)	T1(ms)	% Agarose	mM Ni^{2+}
1	29	281	4	5
2	30	1234	4	0.5
3	32	612	4	2
4	97	299	1	5
5	103	624	1	2
6	157	1459	1	0.5
7	191	301	0.2	5
8	347	647	0.2	2
9	568	1544	0.2	0.5

Table 5.1 T1 and T2 times, percent agarose, and concentration of Ni^{2+} for the 9 test materials.

These materials were used to study the effect of T1, T2 and ΔTE on mean signal location in feature space. The amount of agarose was varied to control T2 and the concentration of Ni^{2+} was varied to control T1.

along with the percentage of agarose and the concentration of Ni^{2+} in each material.

Images of the test phantom were acquired on a 0.5T GE Signa imaging system (GE Medical Systems, Milwaukee) using the standard MS patient exam protocol listed in Table 5.2. Four echoes were obtained with echo times of 15ms, 30ms, 60ms, and 90ms respectively. Software on the Signa console was used to determine the mean signal from circular, 125mm^2 regions-of-interest inside each material in each echo. The mean material signals from pairs of echoes were then plotted in feature spaces, and the distance from the origin, and angle with the abscissa were measured for each material in each feature space.

Image Type	T2 Weighted
Image Plane	Axial
Field-of-view	20 cm
TR	3100 ms
TE	15/30/60/90 ms
Number of slices	1
Slice thickness	2 cm
Matrix size	256 x 192
NEX	1
Bandwidth	12.8 kHz
Scan time	10:32

Table 5.2 Imaging protocol for the phantom experiment.

A quadrature bird cage head coil was used to acquire all images.

5.2.3 Multispectral Analysis: Patient Study

Multispectral analysis of MS lesions was performed on spin-echo exams of one chronic-progressive and two relapsing/remitting MS patients. These patients are part of an ongoing clinical trial to determine the efficacy of beta interferon in the treatment of MS. This trial is being conducted at my institution, and others, in collaboration with the Ares-Serono Group (Geneva, Switzerland). Analysis was performed by acquiring dual spin-echo exams of these patients, quantifying lesions in each exam, normalizing exam intensities, then examining the feature space distribution of lesion intensities over time.

MR Exam Protocol. Exams were acquired on the GE Signa imaging system described above using the dual-echo imaging protocol listed in Table 5.3. Two exams of each relapsing/remitting patient were acquired over a six month period. The chronic-progressive patient was examined approximately every 4 weeks over a seven month period for a total of 7 examinations. Great care was taken to reproduce the oblique axial slice position and orientation as closely as possible for each exam. This was accomplished by acquiring a series of sagittal images using

the "localizer" protocol listed in Table 5.3, for each exam. The inferior margins of the genu and splenium of the corpus callosum were identified in these images, and used to establish the anterior-posterior axis and the superior-inferior position for the oblique axial slices. Next, the tragus of the left and right ears were identified in the "localizer" images and used to establish the left-right axis of the oblique axial slices. Film copies of the slice prescription were created after the initial, or baseline, scan and referred to when represcribing slice position during later examinations. After landmark identification in the "localizer" images, the "multi-echo" protocol listed in Table 5.3 was used to acquire PD and T2 weighted images at each of the oblique axial slices in the patient's brain. All PD and T2 weighted images were then compared by a neuroradiologist to baseline images to ensure that the slice position did not vary by more than one half of a slice thickness over the entire exam. Exams not meeting these criteria were rejected, or retaken if possible.

	Localizer	Multi-echo
Image Type	T1 Weighted	T2 Weighted
Image Plane	Sagittal	Axial-oblique
Field-of-view (cm)	22	22
TR (ms)	400	3100
TE1/TE2 (ms)	20	33/90
Number of slices	9	36
Slice thickness (mm)	5	5
Slice gap (mm)	0	0.5
Matrix size	256 x 128	256 x 192
NEX	1	1
Bandwidth (kHz)	16	9.14/3.76
Scan time	0:56	9:21

Table 5.3 Imaging parameters used in the patient examinations.

A quadrature bird cage head coil was used to acquire all images.

All exams were transferred via ethernet to Sparc workstations (Sun Microsystems Computer Corporation, Mountainview Calif.) for further processing using a system I developed for lesion quantification and analysis.[31] To facilitate manipulation by my system, the dynamic range of each exam was reduced to 8 bits by dividing each image intensity by 4. PD and T2 weighted volumes were then created from each exam by stacking all early echo images and all later echo images, respectively.

Lesion Selection and Quantification. The following criteria were applied to lesions selected for analysis: 1) In at least one exam, each lesion was larger than 1 cm^3 and simultaneously visible in 3 or more slices; and, 2) each lesion was located in the most homogenous region of the head coil, that is in slices superior to the fourth ventricle. Criterion 1 ensured that there were a significant number of lesion voxels for analysis. It also allowed us to limit the impact of partial volume effects in the slice direction by omitting analysis in the most superior and most inferior slices covering each lesion. Criterion 2 helped limit major variations in tissue signals over time.

Three lesions were selected for analysis in Patient A (chronic progressive), three in Patient B (relapsing/remitting), and two in Patient C (relapsing/remitting). My system was used to quantify these lesions by manually outlining the lesion boundaries in all but the most superior and most inferior slices covering each lesion. Outlining occurred in either image, while the lesion boundary was displayed in both images. Detailed adjustment of lesion boundaries was possible using the editing procedures provided by my system. The volume of each lesion was calculated by counting the number of exam voxels entirely contained within the lesion boundary, then multiplying by the voxel volume.

Tissue Intensity Analysis. As described earlier, great care was taken to limit major variations in tissue signals over time by careful patient and slice repositioning and analysis in the most homogenous region of the head coil. Nevertheless,

minor signal variations within tissues still occurred between exams. Since each feature space shows the distribution of early and late echo exam intensities, small signal fluctuations can cause warping of the feature space. Exam intensities were calibrated to limit feature space distortions.

My calibration procedure is based on three assumptions: 1) one reference tissue or material will maintain constant PD and T2 characteristics over time, even with disease progression; 2) scaling factors can be obtained from changes in the PD and T2 intensities of the reference tissue or material; and, 3) small signal fluctuations can be corrected by linear scaling of each feature space axis. I selected cerebrospinal fluid (CSF) as the internal reference tissue for these experiments. Therefore, my system was used to determine the intensity distribution of CSF from multiple circular regions-of-interest placed inside the lateral ventricles in each exam. The total number of CSF samples varied between patients, depending upon the size of their lateral ventricles. Patients A, B and C had 936, 488 and 1456 CSF sample voxels selected, respectively. The feature space mean of each set of samples was then determined, and used to estimate the total CSF cluster mean. The abscissa scaling factor for the i^{th} exam, $S_{i,x}$ was calculated using:

$$S_{i,x} = \frac{\bar{x}_{baseline}}{\bar{x}_i} \tag{5.11}$$

where $\bar{x}_{baseline}$ and $\bar{x}_i$ are the mean CSF early echo intensities from the first exam, and the i^{th} exam, respectively. Similarly, the ordinate scaling factor for the i^{th} exam $S_{i,y}$ was calculated using:

$$S_{i,y} = \frac{\bar{y}_{baseline}}{\bar{y}_i} \tag{5.12}$$

where $\bar{y}_{baseline}$ and $\bar{y}_i$ are the mean CSF late echo intensities from the first exam, and the i^{th} exam, respectively. The scaling factors were applied to each axis of each exam to force alignment of the CSF cluster means. My system was also

used to determine the cluster mean of normal appearing white matter (WM) in each patient exam. The feature space location and distribution of WM intensities was used to estimate the disease-free intensity distribution for comparison to the diseased, or lesion, distributions. My system was used to select multiple circular regions-of-interest inside normal appearing WM. Regions were selected in slices superior to the fourth ventricle, and across each slice to sample any WM intensity fluctuations. No WM samples were taken from regions of diffuse hyperintensity or near visible lesions. Patients A, B and C had 840, 720 and 784 WM sample voxels selected, respectively. The feature space mean of each set of samples was determined, and used to estimate the total WM cluster mean.

Visualization and comparison of the lesion intensity distributions in normalized exams was facilitated by reducing the dimension of the distributions, while minimizing the corresponding loss of information. Dimension reduction was performed by projecting calibrated intensity distributions from 2-D to 1-D. Information content of the projected data was maximized by projecting onto the first principle component[32] of the WM-GM cluster in feature space, since each of the tissues of interest has a first principle component similar to that of the WM-GM cluster.[31] I used my system to identify the WM-GM cluster in feature space, and determine its principle axis. Projections onto this axis were summed into intensity "bins" to form an equivalent WM-GM (eW-G) spectrum. I selected empirically a bin width of 4 as a compromise between higher resolution (narrow bins) and better signal summation (wider bins). Finally, the origin of each eW-G spectrum was set to the projected mean of the internal reference tissue, CSF, to allow comparison of different spectra over time.

Lesion eW-G spectra. After feature space calibration, eW-G spectra were determined from regions-of-interest (ROI's) placed within and around the lesions selected in each patient. For each slice, my system was used to define elliptical ROI's which covered the maximum extent of each lesion observed in all exams.

The ROI's were extended into the WM bordering each lesion to ensure complete coverage of the lesion and its edge.

5.2.4 Multispectral Analysis: Visualization

My system for the quantification and analysis of multispectral MR exams was used to visualize high intensity regions inside MS lesions. The was accomplished by setting a threshold of +30 in the eW-G spectra of each lesion, then highlighting the lesion voxels in the MR exams which fell above the threshold. 3-D volume renderings of the manually quantified lesion regions in Patient A, and their high intensity "cores", at each date were also produced using a commercially available package (SunVision Visualization Package, Advanced Visual Systems, Waltham Mass.) 3-D visualizations were produced by ray tracing through the MR volumes, then translucently rendering the lesion regions in red, and opaquely rendering their high intensity "cores" in yellow. As an anatomical landmark, CSF voxels in the MR exams were also highlighted using my system.[31] These voxels were translucently rendered in blue.

5.3 Results

5.3.1 Phantom Study

Table 5.1 lists the T1 and T2 times of the phantom materials measured with the relaxometer. These materials can be grouped into 3 categories based on their T2 times: short T2's near 30 ms (materials 1,2 and 3); medium T2's near 100 ms (materials 4,5 and 6); and long T2's of 200 ms or more (materials 7,8 and 9). These materials also can be grouped into categories based on their T1 times: short T1's near 300ms (materials 1,4 and 7); medium T1's near 600 ms (materials 3, 5 and 8); and, long T1's of 1200 ms or more (materials 2,6 and 9). Figure 5.2 shows one plot of a feature space showing the mean signals measured in the 30ms and 90ms echoes. A comparison of the values in Table 5.1 to the plot in Fig-

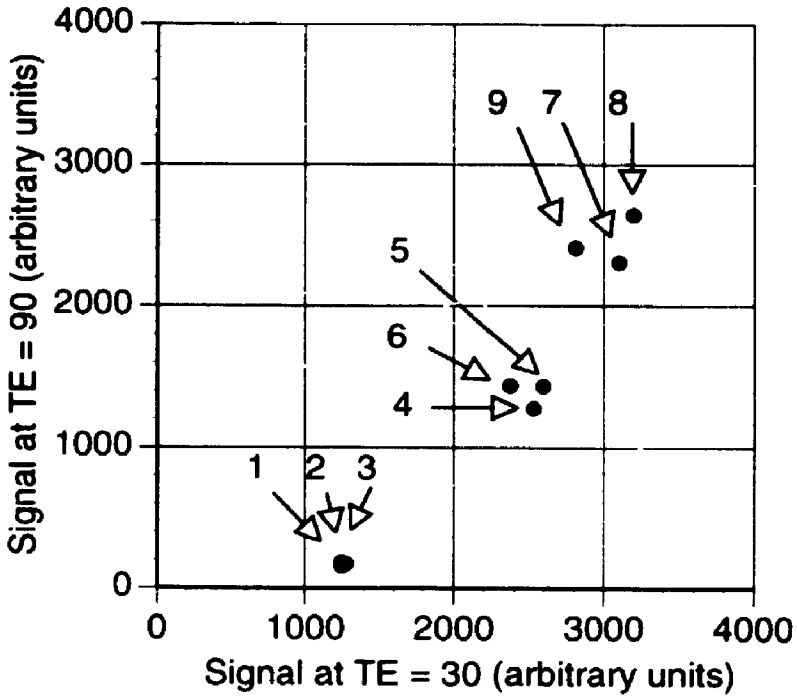


Figure 5.2 The feature space corresponding to test phantom images acquired at echo times of 30ms and 90ms.

The abscissa and ordinate represent the intensities present in the 30ms and 90ms echoes, respectively. The numbers indicate the locations of the mean signal measured for each of the materials listed in Table 5.1.

ure 5.2 reveals that the phantom materials group in feature space according to T2, as predicted by equations (5.7) and (5.10).

Figure 5.3 shows the feature space coordinate (d,θ) as a function of ΔTE for one material from each group in Figure 5.2 (i.e. materials 3,5 and 8). Figure 5.3a is a plot of the measured and expected (equation (5.7)) distance, d, between the origin and the mean material signal versus ΔTE. Figure 5.3b is a plot of the measured and expected (equation (5.10)) angle, θ, versus ΔTE. In general, there is very good agreement between the measured (points) and expected (lines) values, in both graphs. Figure 5.3a shows that d decreases with increasing ΔTE, for each material, as predicted by equation (5.7). Similarly, Figure 5.3b shows that there is a decrease in θ with increasing ΔTE, as predicted in equation (5.10). For materials with T2's near 30ms, there is a very rapid drop in θ with increasing ΔTE.

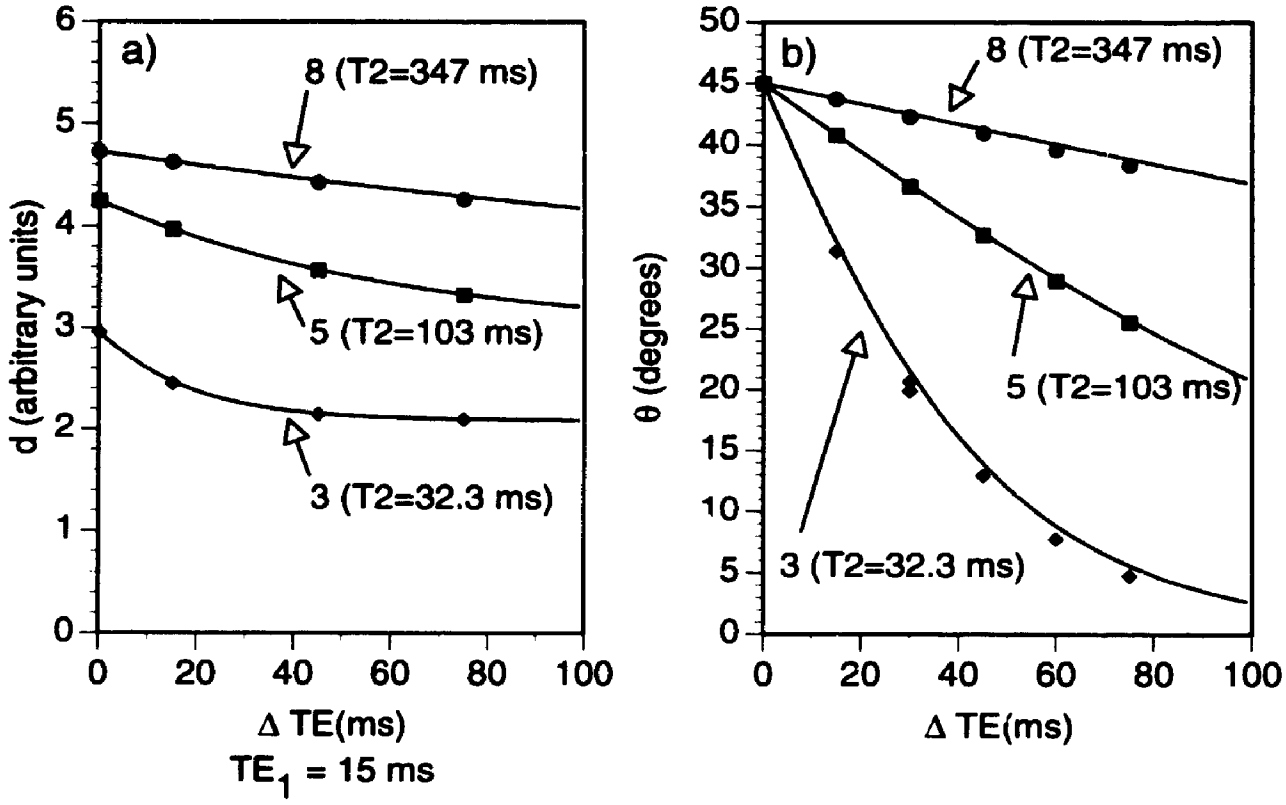


Figure 5.3 The effects of T2 and ΔTE on feature space position.

Plots of the mean distance (equation (5.7)) and angle (equation (5.10)) in feature space for phantom materials 4,5 and 6. a) shows the distance, d, between the material mean and the origin, as a function of ΔTE. b) shows the angle, θ, between the material mean and the abscissa as a function of ΔTE. The points represent the measured values and the solid lines are plots of equations (5.7) and (5.10).

5.3.2 Patient Study

Feature Space Calibration. Table 5.4 lists the mean early and late echo signals for CSF and WM in Patient A at each exam date. These values were used in equations (5.11) and (5.12) to determine the scaling factors, also listed in Table 5.4. The mean abscissa scaling factor for the non-baseline exams from the three patients was 0.98 with a maximum of 1.06 and a minimum of 0.93. The mean, maximum, and minimum ordinate scaling factors were 0.99, 1.06 and 0.94, respectively. Thus, the feature space calibration process changed exam intensities by less than 2% on average, and 7% at most.

	Echo	Day 0	Day 26	Day 54	Day 82	Day 110	Day 166	Day 196
CSF	Early	135.6	140.8	145.3	142.4	137.8	132.9	138.8
	Late	122.9	127.9	131.1	127.7	123.9	120.8	125.0
WM	Early	131.9	137.3	141.0	135.1	135.0	130.6	133.1
	Late	63.2	66.2	67.8	64.1	65.3	62.9	62.9
Scaling Factor	$S_{i,x}$	1.00	0.96	0.93	0.95	0.98	1.02	0.98
	$S_{i,y}$	1.00	0.96	0.94	0.96	0.99	1.02	0.98

Table 5.4 Mean signals and scaling factors for the 7 exams of Patient A.

$S_{i,x}$, and $S_{i,y}$ are the scaling factors derived using equations (5.11) and (5.12) for each exam of Patient A. Listed In the first four rows are the mean signal intensities in the early (33 ms) and late (90 ms) echoes for cerebrospinal fluid (CSF) and white matter (WM).

The effect of the calibration is shown in Figure 5.4. where I have plotted the intensity distributions of CSF and WM, before and after calibration, for Patient A. Even without calibration (Figure 5.4a and 5.4c) there is good agreement in the shapes of the CSF and WM distributions between exams in this patient. However, before calibration the means of the CSF distributions (Figure 5.4a) ranged from 180.1 (Day 0) to 192.7 (Day 54) with overall mean 184.5 and standard deviation 5.4, while the means of the WM distributions (Figure 5.4c) ranged from 143.7 (Day 166) to 155.7 (Day 54) with overall mean 148.7 and standard deviation 4.1. After calibration the CSF distributions had an overall mean of 0 and standard deviation 0, as expected by my calibration procedure (equations (5.11) and (5.12)). The overall mean of the WM distributions was shifted to -35 because the calibration centers the origin of the spectra on CSF. The important effect was to reduce the standard deviation of the WM means to 1.4.

Lesion Volume Measurement. My system was used to quantify lesion volumes in each exam of each patient, as described above. Figure 5.5 shows the measured volumes of lesions 1 to 3 in the seven exams of Patient A. The volume of lesion 1 was approximately 2 cm^3 during the first 3 exams. On Day 82 it nearly

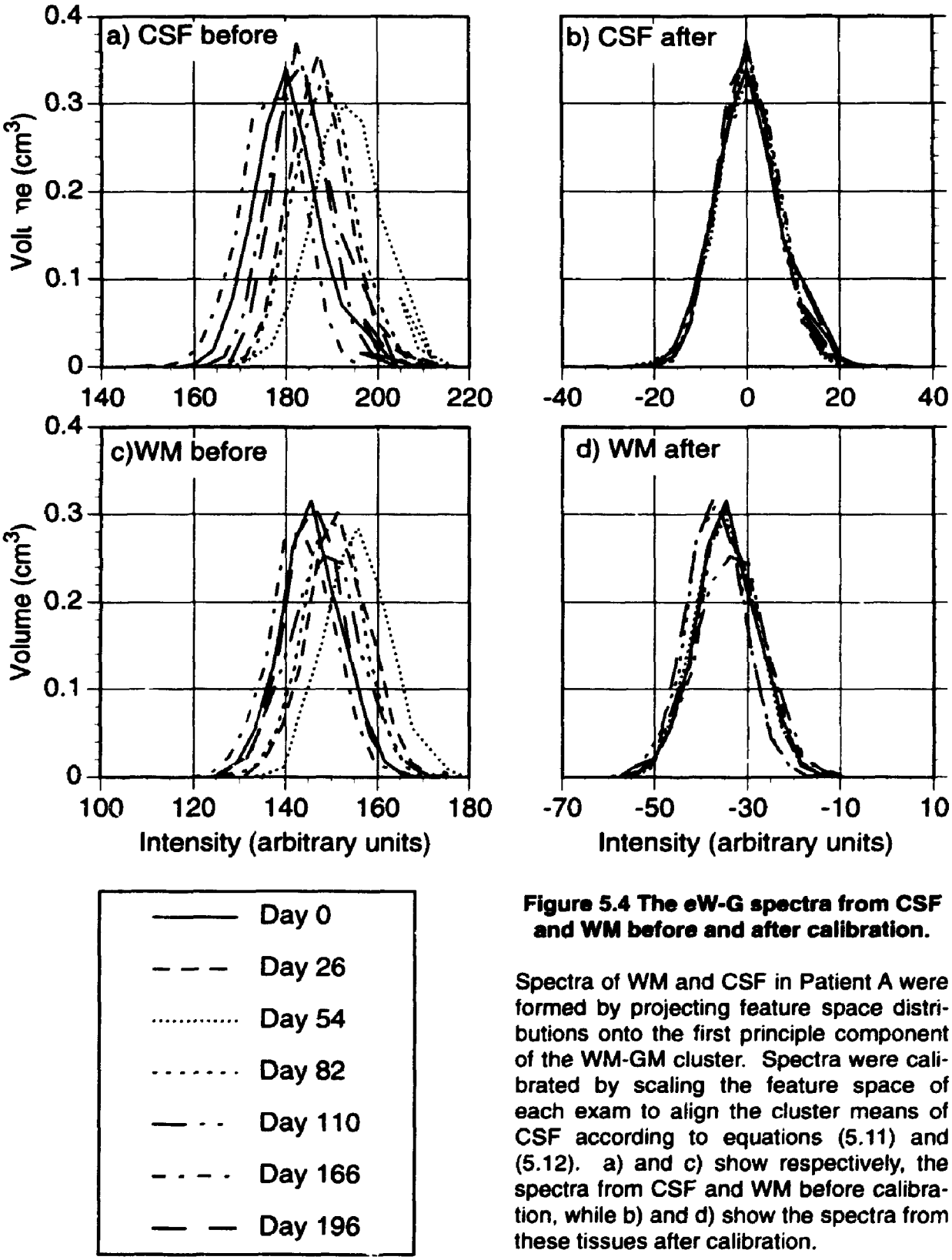


Figure 5.4 The eW-G spectra from CSF and WM before and after calibration.

Spectra of WM and CSF in Patient A were formed by projecting feature space distributions onto the first principle component of the WM-GM cluster. Spectra were calibrated by scaling the feature space of each exam to align the cluster means of CSF according to equations (5.11) and (5.12). a) and c) show respectively, the spectra from CSF and WM before calibration, while b) and d) show the spectra from these tissues after calibration.

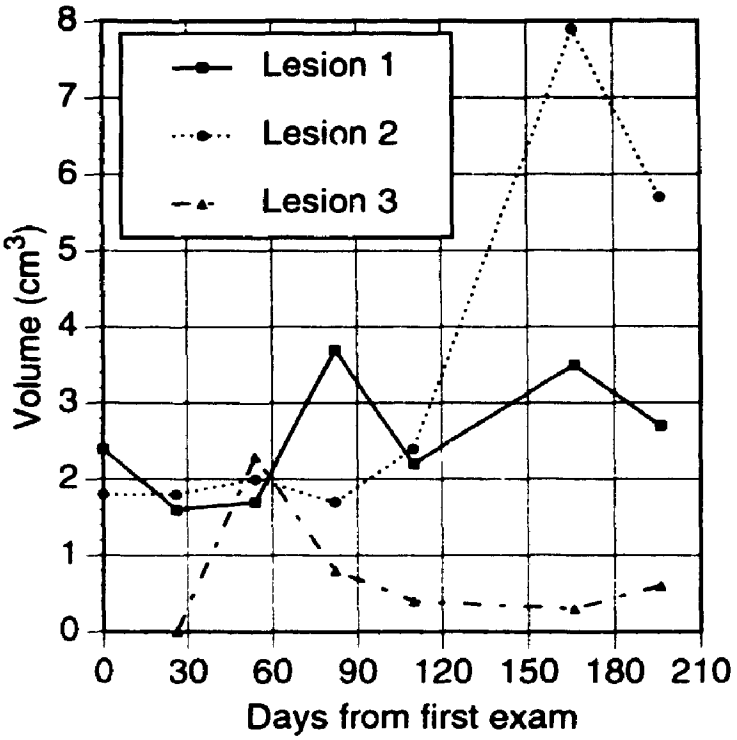


Figure 5.5 The volumes of the three lesions in Patient A over a seven month period.

This patient had examinations performed approximately every 4 weeks, for a total of 7 exams. Of the many lesions in this patient, three were selected for analysis based on their size, visibility in 3 or more slices, and location in the most homogenous region of the head coil. Lesion volumes were then quantified using my system.

doubled in volume to 4 cm^3. Further fluctuations in volume are seen during the last three exams with measured volumes of 2.2 cm^3, 3.5 cm^3 and 2.7 cm^3 respectively. During the first four exams the measured volume of Lesion 2 was also constant at approximately 2 cm^3. However, by Day 166 it had more than tripled in size to 7.9 cm^3, with a reduction to 5.8 cm^3 by the last exam. Lesion 3 was not visible during the first two exams, appearing suddenly at Day 54 with a volume of 2.4 cm^3. During the next 3 exams it shrank to 0.3 cm^3, and increased slightly by the last exam to 0.6 cm^3.

Tables 5.5 and 5.6 list the lesion volumes measured in the exams of Patients B and C, respectively. Some of the lesions in these patients underwent dramatic changes in volume between the baseline and 6 month follow up exam. In Patient B, lesion 1 almost tripled in volume, from 8.4 cm^3 to 22.0 cm^3. Lesion 2 in Patient B was barely visible in the baseline exam with a volume of 0.5 cm^3, yet in the follow up exam it had a volume of 3.3 cm^3. Lesion 3 in Patient B, however, had a constant volume near 6.8 cm^3 in the two exams. In Patient C, lesion 1 also under-

went a dramatic change in volume, decreasing by two thirds, from 3.3 cm^3 in the baseline exam to 1.1 cm^3 in the follow up exam. However, like lesion 3 in Patient B, lesion 2 in Patient C had a constant volume, near 1.0 cm^3, during this period.

	Day 0	Day 181
Lesion 1	8.4 cm^3	22.0 cm^3
Lesion 2	0.5 cm^3	3.3 cm^3
Lesion 3	6.9 cm^3	6.8 cm^3

Table 5.5 Lesion volumes measured in Patient B on Day 0 and Day 181.

	Day 0	Day 168
Lesion 1	3.3 cm^3	1.1 cm^3
Lesion 2	1.0 cm^3	0.8 cm^3

Table 5.6 Lesion volumes measured in Patient C on Day 0 and Day 168.

eW-G Spectra: Patient A. Figure 5.6 shows the eW-G spectra inside a 5.2 cm^3 ROI covering lesion 1 in the seven exams of Patient A. The spectrum on Day 0 has a prominent peak at -20, and a high intensity "tail" region decreasing between +10 and +50. The peak at -20 corresponds to a large, diffuse, slightly hyperintense volume inside the ROI. The intensity of this region is less than that of CSF but higher than that of normal WM. The "tail" corresponds to focal hyperintensities within the diffuse region. On Day 26 the prominent peak near -20 has increased in amplitude to 0.2 cm^3, while the "tail" has diminished, or disappeared. By Day 54 the spectrum is more symmetric with further diminishing of the focal hyperintensities. On Day 82 a dramatic change has occurred in the spectrum. The peak at -20 is reduced by half, while the "tail" has increased in amplitude. The small peak at +75 represents a very bright volume within the ROI. After Day 82 the changes are small, but with a gradual loss of volume at high intensities, and a shape closer to that on Day 0.

Figure 5.7 shows the eW-G spectra inside a 9.0 cm^3 ROI covering lesion 2 in the seven exams of Patient A. On Days 0 through 82 the spectrum is essentially constant, and similar to that of lesion 1 (Figure 5.6), with a peak volume of 0.3 cm^3 near intensity -30, and a "tail" region. However, on Day 110 a number of subtle changes have occurred. The volume of intensities above +20 has increased,

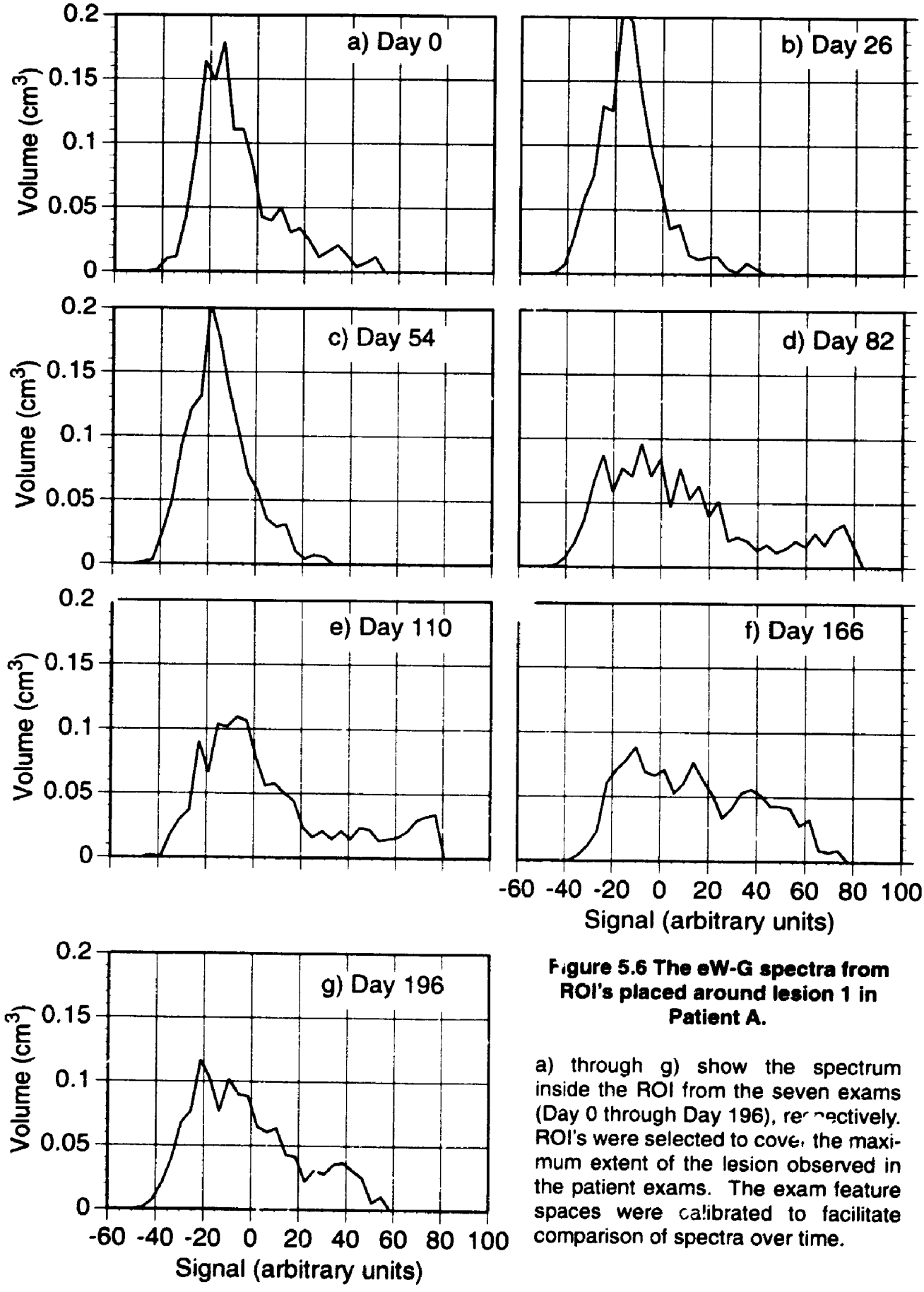


Figure 5.6 The eW-G spectra from ROI's placed around lesion 1 in Patient A.

a) through g) show the spectrum inside the ROI from the seven exams (Day 0 through Day 196), respectively. ROI's were selected to cover the maximum extent of the lesion observed in the patient exams. The exam feature spaces were calibrated to facilitate comparison of spectra over time.

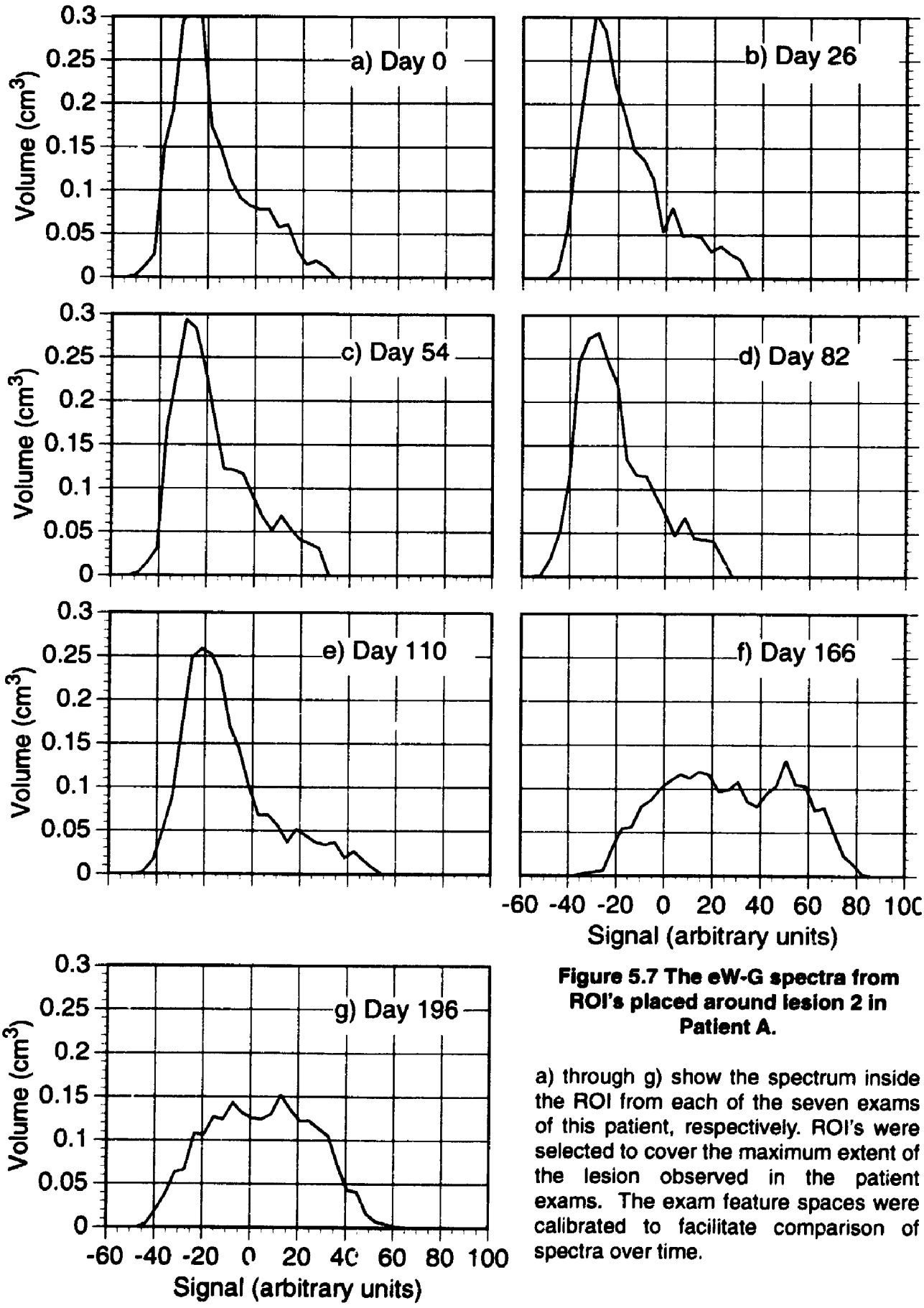


Figure 5.7 The eW-G spectra from ROI's placed around lesion 2 in Patient A.

a) through g) show the spectrum inside the ROI from each of the seven exams of this patient, respectively. ROI's were selected to cover the maximum extent of the lesion observed in the patient exams. The exam feature spaces were calibrated to facilitate comparison of spectra over time.

with intensities extending out to +55, and the prominent peak has dropped in amplitude to 0.25 cm^3 with a slight increase in intensity from -30 to -20. A dramatic change has occurred by Day 166. The peak at -20 has disappeared with a shift of the spectrum to volumes at higher intensities. These changes indicate that within the ROI, WM which appeared normal in the earlier exams has become abnormal. By the last exam, the entire spectrum has become less intense but remained similar in shape.

Figure 5.8 shows the eW-G spectra inside a 2.8 cm^3 ROI covering lesion 3 in the seven exams of Patient A. During the first two exams, lesion 3 was not visible, and the corresponding spectra are of normal appearing WM, with a single prominent peak near -30. On Day 54, however, lesion 3 was clearly visible in the MR exam, and the spectrum changed dramatically. The peak at -20 was greatly reduced, with volumes near 0.05 cm^3 between -20 and +20, and a high intensity "tail". Between Days 82 and 196 the volume at high intensity (> +20) reduced and the volume near intensity -20 (close to normal WM) reappeared. However, some high intensity volumes, above +20, remained. These high intensity volumes correspond to focal hyperintensities within the lesion.

eW-G Spectra: Patients B and C. Figures 5.9 and 5.10. show eW-G spectral changes in the baseline and follow-up exams of two relapsing/remitting patients. Both patients had substantial lesion burden, and the changes observed are both subtle (Figure 5.9c and 5.10b) and large (Figures 5.9a and 5.9b).

5.3.3 Visualization

Figure 5.11 shows 3-D visualizations of the quantified lesion volumes (shown in Figure 5.5) from the three lesions in Patient A, at each exam date. In these visualizations, the lesions are translucently rendered in red. Within each lesion, high intensity regions falling above +30 in the eW-G spectra (Figure 5.6) are rendered opaquely in yellow. Lesion 1, a periventricular cap, has a volume of 2.4 cm^3 on Day 0, and a small high intensity "core" (Figure 5.6a). During the next two exams,

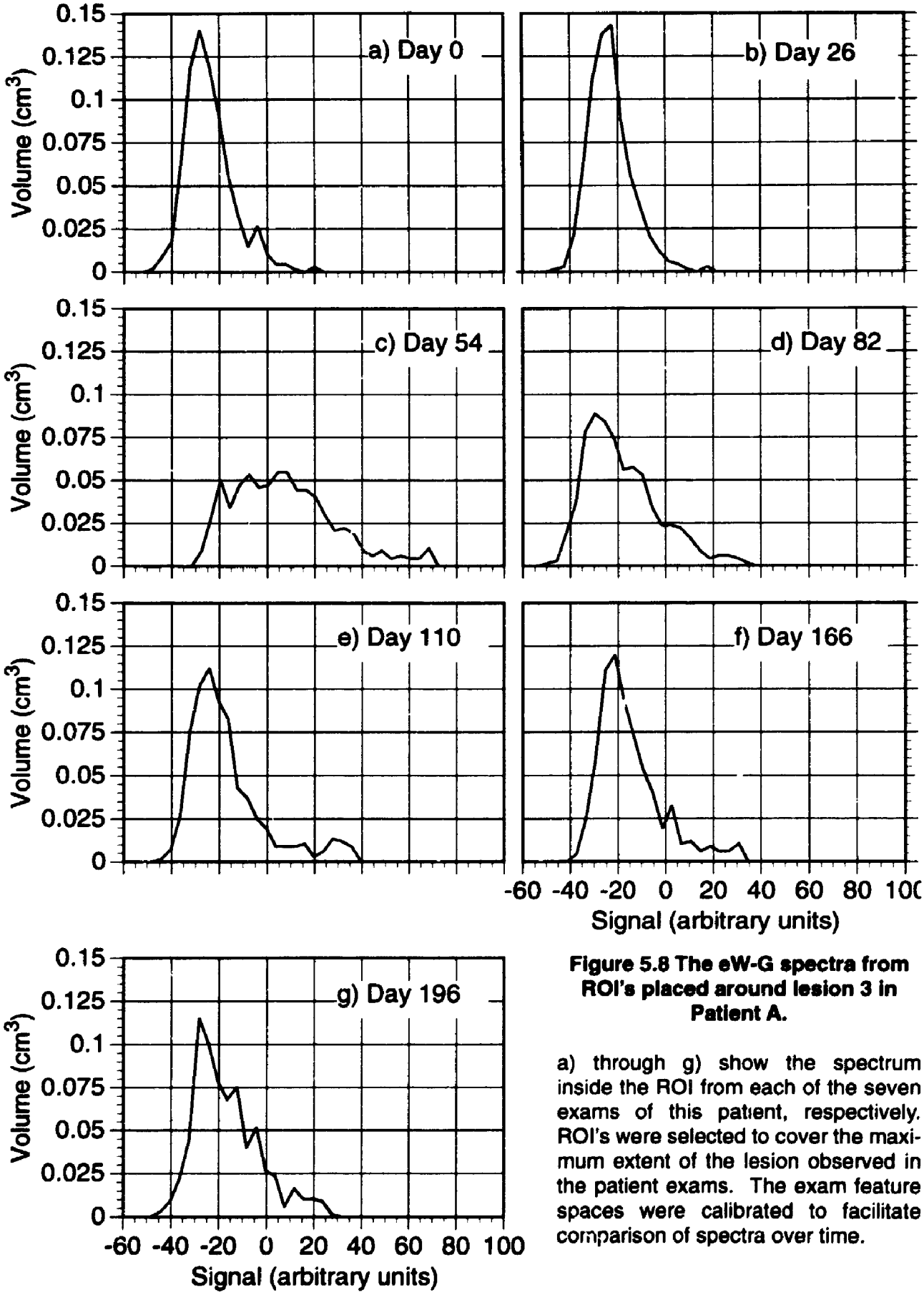


Figure 5.8 The eW-G spectra from ROI's placed around lesion 3 in Patient A.

a) through g) show the spectrum inside the ROI from each of the seven exams of this patient, respectively. ROI's were selected to cover the maximum extent of the lesion observed in the patient exams. The exam feature spaces were calibrated to facilitate comparison of spectra over time.

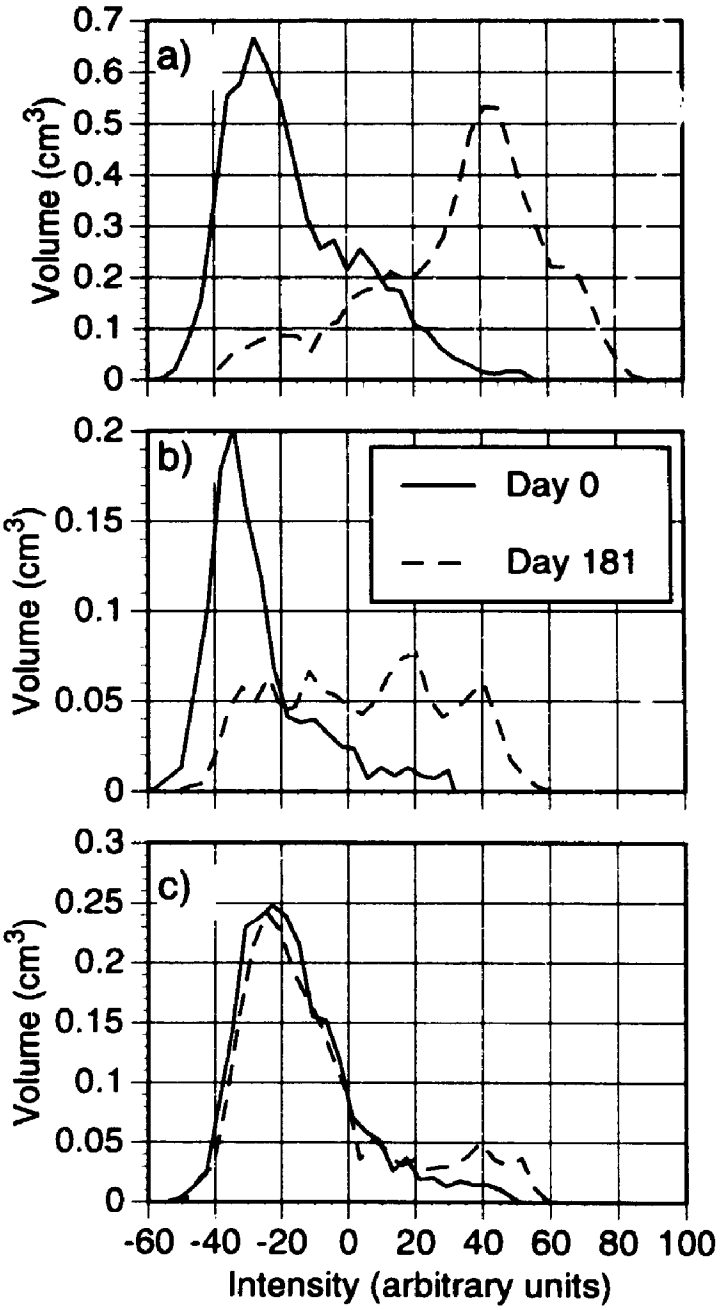


Figure 5.9 The eW-G spectra from ROI's placed around three lesions in Patient B.

a) through c) show the spectra from ROI's with volumes of 25.3 cm^3, 4.8 cm^3, and 9.0 cm^3, selected to cover the maximum extent of lesions 1 to 3, respectively. In each figure the solid line indicates the measured spectrum in the baseline exam (Day 0), and the dashed line indicates the spectrum measured in the follow-up exam (Day 181). The feature space of each exam was calibrated using the mean signal from CSF on Day 0.

the volume of this lesion decreases (to 1.6 cm^3), and the high intensity "core" disappears, as seen in the eW-G spectra of Figure 5.6b and 5.6c. However, on Day 82 the lesion volume increases (to 3.7 cm^3) and a large "core" appears superior to the lateral ventricle (Figure 5.6d). In the final three exam dates, the high intensity "core" fluctuates, as depicted in the eW-G spectra (Figures 5.6e-g).

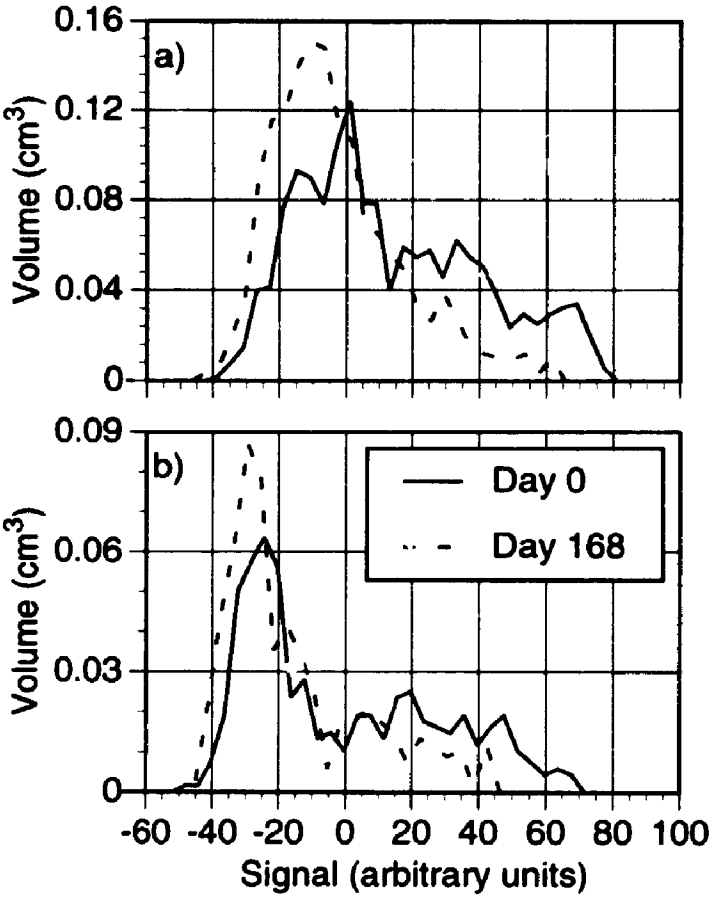


Figure 5.10 The eW-G spectra from ROI's placed around two lesions in Patient C.

a) and b) show spectra from ROI's with volumes of 6.0 cm^3 and 2.4 cm^3 selected to cover the maximum extent of lesions 1 and 2, respectively. In each figure the solid line indicates the measured spectrum in the baseline exam (Day 0), and the dashed line indicates the spectrum measured in the follow-up exam (Day 168). The feature space of each exam was calibrated using the mean signal from CSF on Day 0.

Lesion 2, also a periventricular cap, maintains constant volume near 2 cm^3 during the first four exams and does not have a high intensity "core" (Figure 5.7a-d). However, on Day 110 the volume increases slightly (to 2.4 cm^3) and a small "core", with a single "finger" extending in the superior direction, appears deep within this lesion (Figure 5.7e). By Day 136 the volume increased dramatically (to 7.9 cm^3) and a large, very bright cap developed on the superior side of the lesion. This large change is also seen in the eW-G spectrum in Figure 5.7f. The large "core" within this cap has a number of small "fingers" extending radially from the center. In the last exam the lesion volume decreased but the cap was still present. The "core" within the cap was also reduced, covering the area where the small "finger" extended on Day 110.

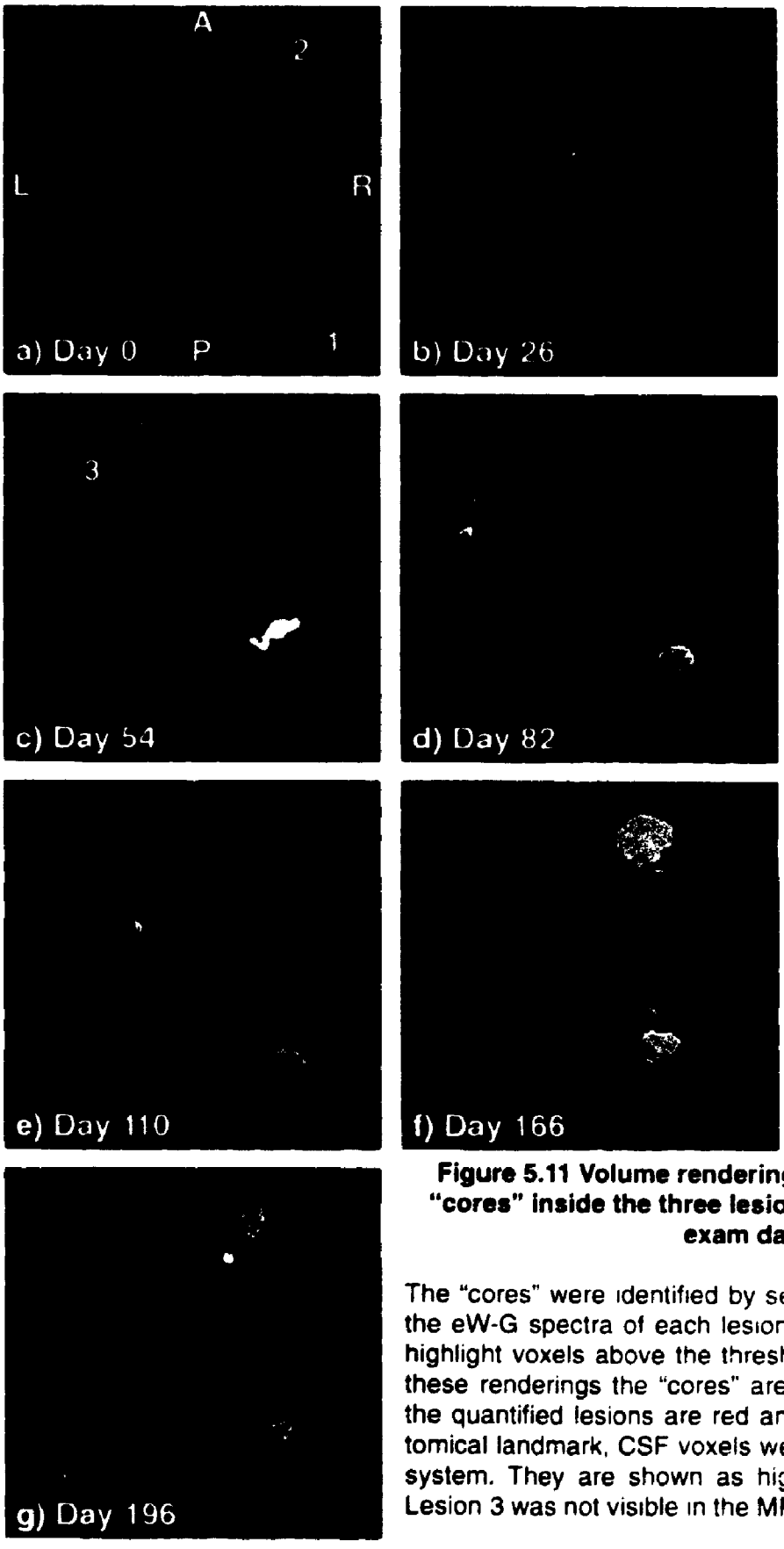


Figure 5.11 Volume renderings of the high intensity "cores" inside the three lesions of Patient A, at each exam date.

The "cores" were identified by setting a threshold of +30 in the eW-G spectra of each lesion, then using my system to highlight voxels above the threshold in the MR images. In these renderings the "cores" are opaque and yellow, while the quantified lesions are red and translucent. As an anatomical landmark, CSF voxels were also identified using my system. They are shown as highly translucent, and blue. Lesion 3 was not visible in the MR exams until Day 54.

Lesion 3 did not appear until Day 54 (Figure 5.11c), when it had a volume of 2.3 cm^3, and a large, hyperintense "core" (see the spectrum in Figure 5.8c) surrounded by a thin low intensity "border". Between Day 54 (Figure 5.11c) and Day 82 (Figure 5.11d) the lesion shrank (to 0.8 cm^3), and the "core" contracted significantly. Over the next three exams the lesion changed little, remaining a small focal hyperintensity with volume near 0.5 cm^3.

5.4 Discussion

Figure 5.3 shows the feature space coordinate, (d,θ), as a function of ΔTE for three of the phantom materials. This figure shows that there is very good agreement between the measured and expected values for d and θ. Equation (5.7) reveals that coordinate d is determined by ρ, T2 and ΔTE, while equation (5.10) reveals that the coordinate ϑ is determined solely by T2 and ΔTE. Thus, when ΔTE is fixed, ρ and T2 determine the tissue location and distribution in feature space. In particular, equations (5.7) and (5.10) indicate that large values of d and θ correspond to tissue regions with high ρ and long T2, while small values for d and θ indicate lower ρ and shorter T2. In the eW-G spectrum, volumes at high intensities correspond to tissue regions with high ρ and/or long T2, while volumes at low intensity correspond to tissue regions with low ρ and/or short T2. Thus, the eW-G spectrum not only provides a means of visualizing a tissue's feature space distribution, it also reflects the tissue's underlying ρ and T2 distribution.

In repeated examinations of the same patient, when ΔTE is fixed, changes in feature space position will be due to changes in ρ and T2, and also to signal variability between exams. My calibration procedure (equations (5.11) and (5.12)) reduces variability between exams so that underlying changes in ρ and T2 have a more prominent effect on feature space position. As described earlier, I selected ventricular CSF as the internal reference tissue for feature space calibration. Although a selective increase in immunoglobulins (oligoclonal banding) occurs in

the CSF of most MS patients,[33] these changes are macroscopically subtle[34] and likely to have little effect on the signal characteristics measured in MR exams of MS patients.

However, the pathological changes which occur in WM and lesions, and their impact on MR signals and the corresponding eW-G spectra, are complex and difficult to describe. Among the reasons for these difficulties are: 1) the physiology and evolution of plaques are variable, and not understood completely; 2) the impact of pathological changes within WM and lesions on proton density, T1, and T2 are not understood completely; and, 3) changes in the underlying tissue parameters result in complex changes in the images, and the feature spaces constructed from those images. Each of these issues will be addressed in turn.

There is some controversy regarding the earliest pathological event leading to demyelination and plaque formation.[35] Indeed, in MS patients there may be diffuse abnormalities in normal appearing WM, which may be primary or secondary in nature.[33] Nevertheless, many authors believe, and gadolinium DTPA enhanced MR imaging suggests, that blood-brain-barrier (BBB) breakdown is the initial pathological event.[36] This is followed by local inflammation, which may be severe,[35] increased cellularity in some cases, and the presence of lymphocytes.[33] Subsequently, infiltration by myelin phagocytes may lead to demyelination, or alternatively, demyelination may lead to macrophage activation, and microglial hyperplasia.[35] In any case, active demyelination is a characteristic of the acute, early lesion. Generally, oligodendrocytes are present in the early lesion, and remyelination may occur leading to complete remission, or more commonly, to the formation of partially remyelinated lesions, or shadow plaques. This entire cycle may recur, leading to further demyelination and/or remission. Lesions may remain in this active state for many weeks or months and are characterized by increased gliosis and a hypercellular edge which may extend 3 cm into surrounding tissue. In contrast, older inactive lesions are characterized by an absence of myelin breakdown products, hypocellular cores, complete demyelination, oligo-

dendrocyte loss, and axonal degeneration. However, older chronic lesions may still be edematous, even when gadolinium enhanced MRI suggests that the BBB is intact.

Many investigations have been performed to study the impact of these pathological events on MR relaxation times.[19,20,37-41] A review of this research is presented by Matthews.[33] Generally, investigators have found that lesions with similar pathology have similar MR characteristics. In addition, T1 and T2 tend to increase with increased inflammation and demyelination, though not exclusively. In particular, excess water was found to increase both T1 and T2, but higher levels of protein in the water tend to lower T2 but not T1. Thus a focal severe inflammation with relative hypocellularity may lead to very long T2's, while inflammation accompanied by an immune cellular response may lead to moderately increased T2's, compared to normal WM. Remyelination reduces T1 and T2, with complete remyelination causing the lesion site to be indistinguishable from normal WM in MRI. However, the formation of shadow plaques is likely to result in relaxation times which are slightly longer than normal WM, depending upon the degree of remyelination. Older chronic gliosed plaques tend to have longer relaxation times than acute active plaques, though the distinction is not clear. Finally, recent evidence suggests that even normal appearing WM in MS patients may have increased T1 and T2, compared to normal controls.[41,19]

The observations described above may allow physiological interpretation of the lesion eW-G spectra measured in these experiments. Perhaps the most prominent feature of many of these spectra is the presence of a peak at intensities near -30 seen in many of the spectra in all three patients (Figures 5.6 to 5.10). In some of these, no lesion is visible within the ROI in the MR exams, and the prominent peak in each case corresponds to normal appearing WM. Other spectra, however, have a peak with similar shape but which is slightly more intense than normal WM (see Figure 5.6a, Figure 5.7e and Figure 5.10a). These peaks

correspond to diffuse, slightly hyperintense regions and may represent regions of active, post inflammatory disease, or partially remyelinated shadow plaques.

Figure 5.8 (lesion 3 in Patient A) provides an interesting example of the evolution of an eW-G spectrum, in light of the pathology discussed above. During the first two exams, no lesion was present within the ROI covering the site of lesion 3. By Day 54 (Figure 5.8c) an acute event had occurred, and a lesion was visible inside the ROI. Over the next 142 days, and 4 exams, the spectrum "regrew" a prominent peak near -30. Indeed, the spectra from the last two exams are very similar to that on Day 0. The "regrowth" may represent reduced inflammation or remyelination within and around the lesion site.

Lesion 1 in Patient A (Figure 5.6) also undergoes an acute change between Days 54 (Figure 5.6c) and 82 (Figure 5.6d). The resulting eW-G spectrum (Figure 5.6d) is very similar to that in Figure 5.8c. Both spectra are characterized by peaks near -20, flat distributions between -10 and +20 and a "tail" region above +20 with peaks at very high intensity. The very high intensities, above +60 within each spectrum, suggest long T2's and may correspond to focal severe hypocellular inflammation. However, after the acute event, the spectra in Figures 5.6e-g do not "regrow" a prominent peak near WM. Instead, these spectra maintain constant volume at high intensities. In Figure 5.6f the spectrum of lesion 1 is characterized by three small peaks near -10, +10 and +40. It is interesting to note the similarities between this spectrum, and that in Figure 5.9b, Day 181. The similarity in these two spectra may indicate a similar pathological state in these two lesions, in patients A and B; however I have no evidence to support this speculation and further studies with MR spectroscopy[16,17] and animal models of MS[39,40] would be required.

The small spectral peak at +40 in Figure 5.6f may correspond to inflammation accompanied by an immune cellular response, and thus a lower T2 and lower intensity, than hypocellular inflammation (acellular edema). Spectral peaks near

+40 are also present in the distributions in Figure 5.6g, Figures 5.9a-c and Figure 5.10a, and may represent similar pathological states in these regions. Finally, the large, nearly unimodal distribution around +40 in Figure 5.9a suggests a nearly uniform pathological state throughout a large region. This lesion changed from a large periventricular cap surrounded by "normal" WM on Day 0, to a very large, very intense plaque extending through 7 exam slices on Day 181.

In addition to the potential to provide information about the pathological state of MS lesions, multispectral analysis may be able to provide information about lesion activity. Currently, activity in MR images is measured by the appearance and reappearance of lesions, and the change in volume of existing lesions.[36] There are a number of problems associated with lesion volume measurement alone as a gauge of disease activity: first, quantification is subject to inter- and intra-operator variability which may hinder measurement of disease activity;[42] and second, there is much pathological evidence to suggest that activity in existing lesions may occur within the lesions, and may not necessarily involve changes in volume.[33] For example, Figure 5.5 reveals that the measured volume of lesion 1 in Patient A was nearly identical on Day 0 and Day 110. However, differences in the eW-G spectra at each date (Figures 5.6a and 5.6e) may indicate a more abnormal pathological state in the later exam. Similarly, the volumes of this lesion on Days 82 and 166 are nearly identical, yet the eW-G spectra on these dates (Figures 5.6d and 5.6f) are not, though neither is normal. Finally, lesion 3 in Patient B has nearly identical volume in both the baseline and follow-up exams. However, the eW-G spectrum at the later date (Figure 5.9c) reveals a small decrease in the peak at -20 and a corresponding increase at +40, which may suggest an increase in inflammation.

Application of the eW-G spectrum to measurement of disease activity may also reduce the effects of inter- and intra-operator variability. Currently, measures of disease activity based on lesion volume changes require exact specification of lesion borders in repeated exams, and may be subject to considerable inter- and

intra-operator variability.[42] However, the eW-G spectra may be determined from ROI's placed to cover the lesion and its border, eliminating the need to specify exact lesion borders. Small changes to the size and shape of the ROI between exams, which add or remove regions of normal appearing WM to the ROI, will only change the amplitude of low intensity peaks within the eW-G spectra. Therefore, measures of disease activity based on changes in "high" eW-G intensities should be less sensitive to small changes in the ROI, and eliminate the need to specify the lesion border exactly. This approach may be useful when measuring activity in large regions of diffuse disease where prominent lesion edges are difficult to define or subject to large inter- and intra-operator variability.

5.5 References

[1] I.R. Young, A.S. Hall, C.A. Pallis, G.M. Bydder, N.J. Legg, R.E. Steiner, "Nuclear Magnetic Resonance Imaging of the Brain in Multiple Sclerosis". Lancet, 2,1063-1066 (1981).

[2] J. Kesselring, I.E.C. Ormerod, D.H. Miller, E.P.G.H. du Boulay, W.I. McDonald *Magnetic Resonance Imaging in Multiple Sclerosis.* New York, Thieme Medical Publishers, 1-93 (1989).

[3] C. Isaac, D.K. Li, M. Genton, C. Jardine, E. Grochowski, M. Palmer, L.F. Oger, D.W. Paty, "Multiple sclerosis: A serial study using MRI in relapsing patients". Neurology, **38(10)**, 1511-1515 (1988).

[4] E.W. Willowghby, E. Grochowski, J. Oger, L.F. Kastrukoff, D.W. Paty, "Serial magnetic resonance scanning in multiple sclerosis: A second prospective study in relapsing patients". Annals of Neurology, **25**, 43-49 (1989).

[5] R.A. Koopmans, D.K.B. Li, J.J.F. Oger, L.F. Kastrukoff, C. Jardine, L. Costley, S. Hall, E.W. Growchowski, D.W. Paty, "Chronic Progressive Multiple Sclerosis: Serial Magnetic Resonance Brain Imaging Over Six Months". Annals of Neurology, **26**, 248-256 (1989).

[6] A.J. Thompson, A.G. Kermode, D.G. MacManus, et al. "Patterns of disease activity in multiple sclerosis: clinical and magnetic resonance imaging study". British Medical Journal, **300**, 631-634 (1990).

[7] J.O. Harris, J.A. Frank, N. Pattronas, D.E. McFarlin, H.F. McFarland, "Serial Gadolinium-enhanced Magnetic Resonance Imaging Scans in Patients with Early, Relapsing-Remitting Multiple Sclerosis: Implications for Clinical Trials and Natural History". Annals of Neurology **29**, 548-444 (1991).

[8] M.D. Weinshenker, R.N. Bass, S.J. Karlik, G.C. Ebers, G.P.A. Rice, "An open trial of OKT3 in patients with multiple sclerosis." Neurology 41, 1047 (1991).

[9] D.W. Paty, D.K.B. Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group "Interferon beta-1b is effective in relapsing-remitting multiple sclerosis". Neurology, **43**, 662-667 (1993).

[10] L. Kappos, D. Stadt, M. Ratzka, et al. "Magnetic resonance imaging in the evaluation of treatment in multiple sclerosis". Neuroradiology, **30**, 299-302 (1988).

[11] I.E.C. Ormerod, D.H. Miller, W.I. McDonald, et al. "The role of NMR imaging in the assessment of multiple sclerosis and isolated neurological lesions: a quantitative study". Brain, **110**, 1579-1616 (1987).

[12] S.J. Karlik, M.K. Vandervoort, M. Hopkins, J.H. Noseworthy "Serial quantitative brain MRI in multiple sclerosis using visual and volumetric techniques". Abstract Neurology, 40, 142 (1990).

[13] F. Pannizzo, M.J.B. Stallmeyer, J. Friedman, *et al.* "Quantitative MRI studies for assessment of multiple sclerosis". Magnetic Resonance in Medicine, 24, 90-99 (1992).

[14] D.W. Paty, M. Bergstrom, M. Palmer, J. MacFadyen, D. Li, "A quantitative magnetic resonance image of the multiple sclerosis brain". Abstract. Neurology, 35(1), 137 (1985).

[15] S.S. Gebarski, T.O. Gabrielsen, S. Gilman, J.E. Knake, J.T. Latack, A.M. Aisen, "The initial diagnosis of multiple sclerosis: clinical impact of magnetic resonance imaging". Annals of Neurology, 17, 469-474 (1985).

[16] D.L. Arnold, P.M. Matthews, G. Francis, J. O'Connor, J. Antel, "Proton magnetic resonance spectroscopic imaging for metabolic characterization of demyelinating plaques". Annals of Neurology, 31, 235-241, (1992).

[17] J.S. Wolinsky, P.A. Narayana, M.J. Fenstermacher, "Proton magnetic resonance spectroscopy in multiple sclerosis". Neurology, 40(11), 1764-1769, (1990).

[18] S.S. Gill, R.K. Small, D.G. Thomas, P. Patel, R. Porteous, N.V. Bruggen, D.G. Gadian, R.A. Kauppinen, S.R. Williams, "Brain metabolites as 1h NMR markers of neuronal and glial disorders". NMR in Biomedicine, 2, 196-200, (1989).

[19] J. Armspach, D. Gounot, L. Rumbach, J. Chambron, "In Vivo Determination of Multiexponential T_2 Relaxation in the Brain of Patients with Multiple Sclerosis", Magnetic Resonance Imaging, 9, 107-113 (1991).

[20] A. MacKay, K. Whittall, J. Adler, D. Li, D. Paty, D. Graeb, "In Vivo Visualization of Myelin Water in Brain by Magnetic Resonance", Magnetic Resonance in Medicine, 31, 673-677 (1994).

[21] L. Rumbach, J.P. Armspach, D. Gounot, I.J. Namer, J. Chambron, J M. Warter, M. Collard, "Nuclear Magnetic Resonance T2 Relaxation Times in Multiple Sclerosis". Journal of Neurological Sciences, 104, 176-181, (1991).

[22] B.K. Rutt, J.D. Mitchell, "Myelin-selective Imaging with 3-D Fast Spin-Echo MRI". Book of Abstracts, Proceedings of the SMRM, 137, (1993).

[23] M.W. Vannier, R.L. Butterfield, D. Jordan, W.A. Murphy, R.G. Levitt, M. Gado, "Multispectral analysis of magnetic resonance images". Radiology, 154, 221-224 (1985).

[24] M.W. Vannier, R.L. Butterfield, D.L. Rickman, D.M. Jordan, W.A. Murphy, P.R. Biondetti, "Multispectral magnetic resonce image analysis". CRC Critical Reviews in Biomedical Engineering, 15(2), 117-144 (1987).

[25] D.H. Miller, F. Barkhof, I. Berry, L. Kappos, G. Scotti, A.J. Thompson "Magnetic Resonance Imaging in Monitoring the Treatment of Multiple Sclerosis: Concerted Action Guidelines." Journal of Neurology Neurosurgery and Psychiatry, 54, 683-688, (1991).

[26] O. Kubler, G. Gerig, "Segmentation and Analysis of Multidimensional Data-Sets in Medicine". NATO ADI, Series F60, 63, (1990).

[27] G. Gerig, J. Martin, R. Kikinis, O. Kubler, M. Shenton, F.A. Jolesz, "Automating segmentation of dual-echo MR head data". Information Processing in Medical Imaging - 12th International Conference, Colchester, A. C. F., and Hawkes, D. J., editors. 175-187, IPMI, Springer-Verlag, London, UK, July (1991).

[28] K. A. Kraft, P. P. Fatouros, G. D. Clarke and P. R. S. Kishore, "An MRI Phantom Material for Quantitative Relaxometry," Magnetic Resonanc in Medicine, 5, 555-562 (1987).

[29] M.D. Mitchell, H.L. Kundel, L.Axel, P.M. Joseph, "Agarose as a Tissue Equivalent Phantom Material For NMR Imaging," Magnetic Resonance Imaging, 4, 263-266 (1986).

[30] K.P. Whittall, M.J. Bronskill, R.M. Henkelman, "Investigation of analysis techniques for complicated NMR relaxation data". Journal of Magnetic Resonance, 95, 221-234 (1991).

[31] J.R. Mitchell, S.J. Karlik, D.H. Lee, A. Fenster, "Computer Assisted Identification and Quantification of Multiple Sclerosis Lesions in MR Volumes of the Brain". Journal of Magnetic Resonance Imaging, 4, 197-208 (1994).

[32] P. Armitage, G. Berry, *Statistical methods in medical research.* 2nd edition. Oxford: Blackwell Scientific Publications, 1-559 (1990).

[33] W.B. Matthews, A. Compston, I.V. Allen, C.N. Martyn, *McAlpine's Multiple Sclerosis.* Edinburgh, Scotland: Churchill Livingstone, 2nd edition, 1991; 330-332.

[34] J.N. Whitaker, E.N. Benveniste, S. Zhou, "Cerebrospinal Fluid" in *The Handbook of Multiple Sclerosis*, S.D. Cook editor. New York: Marcel Dekker Inc., 251-270 (1990).

[35] J.W. Prineas, "Multiple Sclerosis: Pathology of the Early Lesion." In: R.M. Herndon, F.J. Seil, *Multiple Sclerosis: Current Status of Research and Treatment*, New York: Demos Publications, 1994.

[36] D.W. Paty, J.J.F Oger, R.A. Koopmans, E.W. Willoughby, W.A. Stewart, D.K.B. Li, "Magnetic Resonance Studies in Multiple Sclerosis: Diagnosis, Quantification of Disease Burden and Assessment of Disease Activity." In: R.M. Herndon, F.J. Seil, *Multiple Sclerosis: Current Status of Research and Treatment*, New York: Demos Publications, 1994.

[37] W.A. Stewart, E.C. Alvord Jr., S. Hruby, L.D. Hall, D.W. Paty, "Magnetic Resonance Imaging of Experimental Allergic Encephalomyelitis in Primates", Brain, 114, 1069-1096 (1991).

[38] Y. Fu, K. Tanaka, S. Nishimura, "Evaluation of Brain Edema using Magnetic Resonance Proton Relaxation Times", Advances in Neurology, 52, 165-176 (1990).

[39] S.J. Karlik, J.J. Gilbert, C. Wong, M.K. Vandervoort, J.H. Noseworthy, "NMR Studies in Experimental Allergic Encephalomyelitis (EAE): Factors which contribute to T1 and T2 values." Magnetic Resonance in Medicine, 14, 1-11, (1990).

[40] S.J. Karlik, E.A. Grant, D.H. Lee, J.H. Noseworthy, "Gadolinium Enhancement in Acute and Chronic Progressive Experimental Allergic Encephalomyelitis in the Guinea Pig." Magnetic Resonance in Medicine, 30: 326-331, (1993).

[41] D. Lacomis, M. Osbakken, G. Cross, "Spin-Lattice Relaxation (T1) Times of Cerebral White Matter in Multiple Sclerosis", Magnetic Resonance in Medicine, 3, 194-202 (1986).

[42] J.R. Mitchell, S.J. Karlik, D.H. Lee, M. Eliasziw, G.P.A. Rice, A. Fenster, "The Variability of Manual and Computer Assisted Quantification of Multiple Sclerosis Lesion Volumes." submitted to Medical Physics, (1995).

CHAPTER 6 CONCLUSIONS

This chapter summarizes the major results presented in this thesis. Each Chapter 2 through 5 is addressed in turn. At the end of this summary I discuss some of the potential applications of my work.

6.1 Summary of Results

6.1.1 A computerised system for MR exam analysis

MRI is the principle imaging technique for the diagnosis of Multiple Sclerosis (MS). However, manually quantifying the number and extent of lesions in MR images is arduous. I have developed a computerised 3-D quantitative system to assist in the identification and analysis of MS lesions in MR proton density weighted (PD), and T2 weighted volumes of the head. The system provides intuitive, interactive operations which allow flexible extraction of information from the data. Use of the system to analyze MR exams of a phantom containing regular "lesions" revealed that accurate (average error < 0.21 cm^3) and precise (10% or better for "lesions" > 1 cm^3) measurements of objects less than 7 cm^3 is possible, and that an estimate of the quantization error predicted the uncertainty in the volume.

Analysis of four MR exams of a chronic-progressive MS patient taken over an 18 month period was performed. A 2-D histogram showing the frequency of voxels with particular PD and T2 weighted intensities revealed a distinct cluster only in histograms of slices with lesions. Measurement and 3-D volume rendering of lesions clearly showed changes in lesion shape, position and size.

6.1.2 The Variability of Lesion Quantification

Quantification of changes in the appearance of lesions in MR exams is a common technique to assess disease activity. However, accurate quantification tech-

niques are subject to inter- and intra-operator variability, which may hinder monitoring of disease progression. I have developed a new algorithm to assist an experienced operator in the quantification of MS lesions in standard spin-echo MR exams. The accuracy of assisted and manual quantification under known conditions was studied using exams of a test phantom, while inter-and intra-operator reliability and variability were studied using exams of an MS patient. Results from the phantom study show that accuracy is improved by assisted quantification. The patient exam results indicate that assisted quantification reduced inter-operator variability from 0.34 cm^3 to 0.17 cm^3, and reduced intra-operator variability from 0.23 cm^3 to 0.15 cm^3. In addition, the minimum significant change between two successive measurements of lesion volume by the same operator, was 0.64 cm^3 for manual quantification, and 0.42 cm^3 for assisted quantification. For two different operators making successive measurements, the minimum significant change was 0.94 cm^3 for manual quantification, but only 0.47 cm^3 for assisted quantification. Finally, the number of lesions to be monitored for an average change in volume at a given power and significance level was reduced by a factor of 2 to 4 by assisted quantification. These results suggest that assisted quantification may have practical applications in clinical trials, especially those that are large, multi-centre, or extended over time, and therefore require lesion measurements by one or more operators.

6.1.3 The Effects of Field Strength and a Noise Reduction Filter on Quantification

Recently, guidelines for the use of MRI in the monitoring of MS have recommend the use of imaging systems with mid field (0.5T to 1.0T) or high field (greater than 1.0T) strengths. Higher field strengths provide many advantages, including increased signal-to-noise ratios (SNR). SNR also may be increased by post-processing algorithms which reduce noise. I evaluated the impact on operator variability of: a) lesion quantification in high-field (1.5T) versus mid-field (0.5T) exams; and, b) an anisotropic diffusion filter algorithm which reduces image noise

without blurring or moving object boundaries. Inter-and intra-operator reliability and variability were studied using repeated quantification of lesions in 1.5T and 0.5T filtered and unfiltered MR exams of an MS patient. Results indicate that inter-operator variability in 1.5T unfiltered exams was 0.34 cm^3, and was significantly larger than that in 1.5T filtered (0.27 cm^3), 0.5T unfiltered (0.26 cm^3), and 0.5T filtered (0.24 cm^3) exams. Similarly, intra-operator variability in 1.5T unfiltered exams was 0.23 cm^3, and was significantly larger than that in 1.5T filtered (0.19 cm^3), 0.5T unfiltered (0.19 cm^3), and 0.5T filtered (0.18 cm^3) exams. In addition, the minimum significant change between two successive measurements of lesion volume by the same operator, was 0.64 cm^3 in 1.5T unfiltered exams, and 0.5 cm^3 in other exams. For two different operators making successive measurements, the minimum significant change was 0.94 cm^3 in 1.5T unfiltered exams, but only 0.7 cm^3 in other exams. Finally, the number of lesions to be monitored for an average change in volume at a given power and significance level was greater by 30% to 60% for quantification in 1.5T unfiltered exams. These results suggest that inter- and intra-operator variability are reduced by anisotropic filtering, and by quantification in 0.5T exams. Reduced operator variabilities may result from higher detail signal-to-noise ratios (dSNR's) in 0.5T and filtered exams.

6.1.4 Analysis of Lesion Intensity Composition

Although quantification of the lesion burden from serial MR exams of MS patients is a common technique to asses disease activity in clinical trials, pathological change may occur within a lesion without a corresponding change in volume. Therefore, measures of lesion volume *and* composition may improve the sensitivity of detecting disease activity.

I have developed a new technique which provides information about MS lesion composition from standard spin-echo MR exams. The new technique is based on the multispectral "feature space" intensity distributions of the lesions and normal

tissues. I analyzed MR exams of materials with known T1 and T2 times and showed that feature space position from spin-echo exams is largely determined from ρ (proton density), T2, and the inter-echo delay. Information about tissue composition was obtained by reducing the multidimensional intensity distribution to 1-D, while minimizing the loss of information. This technique was used to analyze lesions in standard spin-echo MR exams of three MS patients. Lesion distributions were compared between exams by first calibrating the exams based upon the intensity distributions of CSF, an internal reference tissue. Results reveal that the calibration procedure eliminated variability in the distributions of CSF, as expected. Calibration also greatly reduced variability in the intensity distributions of normal appearing white matter. Many of the lesion distributions had a distinctive peak at low intensity corresponding to normal appearing WM. Within the lesion distributions, increases in high intensity peaks were accompanied generally by reductions in the WM peak. Increases at high intensities may correspond to increased inflammation, while regrowth of the WM peak may represent reduced inflammation, or remyelination at lesion sites. Serial analysis of the lesion distributions revealed some dramatic fluctuations, even when lesion volume remained constant.

6.2 Future Work

This section describes some of the potential applications of the system and techniques I have developed. These applications fall generally into three categories: 1) improvements to the existing system; 2) further studies of the sources and effects of variability; and, 3) clinical applications. Each will be discussed in turn.

6.2.1 System Improvements

Chapter 2 includes a description of the system I developed for the quantification of MS lesions from MR images of the head. This system was originally devel-

oped as a research tool, and was not intended to be applied clinically. However, clinical applications do exist, yet a number of improvements should be made to facilitate clinical application. In particular, the system should be simplified to provide only a few basic functions, including manual and assisted quantification. This should not only make it easier to use but improve its performance. Improved performance is particularly important to allow analysis of larger, higher resolution images. Performance may also be improved by using more sophisticated data structures internally, to represent images and classified regions. New data structures may be essential to allow prompt manipulation of large 3-D images which can be obtained with current MR protocols.

The system in its current configuration allows analysis of single or dual-echo images. The latter are often two different views of the same patient anatomy. However, current MR technology is capable of providing many additional views. These may be additional echoes, or they may be images whose contrast is weighted by other tissue specific properties such as magnetization transfer, or diffusion. The system should be expanded to allow more than two views of the same anatomy. If there are 'n' views of the patient, this will involve manipulation of an 'n'-dimensional feature space, since each feature space axis corresponds to one view. Multidimensional feature spaces may allow better separation of tissue clusters. In turn this should improve quantification techniques, such as the assisted quantification described in Chapter 3, which rely on 2-D tissue cluster identification.

However, multidimensional feature spaces will make visualization of tissue intensity distributions very difficult. Techniques like that described in Chapter 5, which reduce feature space to 1 or 2 dimensions while minimizing the loss of information, may be critical in an expanded system. The additional information provided by multidimensional feature spaces also may allow more specific physiological measures in the spectra resulting from dimension reduction.

6.2.2 Additional Studies of Accuracy and Variability

Much of the work reported in this thesis concerned the variability of lesion quantification. The issue of accuracy was not addressed generally, since actual lesion volumes are difficult to determine. Nevertheless, information about quantification accuracy may be of benefit when designing new algorithms, or assessing results in clinical trials.

In Chapters 2 and 3, accuracy studies were performed using a test phantom which contained "lesions" doped with graphite to provide contrast with the surrounding agarose "brain". These phantoms were idealistic and failed to mimic actual conditions in the following ways: the lesions were very regular with consistent slice-to-slice profiles; the "lesions" and "brain" had different intensities than actual lesions, WM, GM or CSF; the contrast between "lesions" and "brain" was very high; and, there was no attempt during phantom construction to match the ρ, T1 and T2 of actual brain tissues. Even in these idealistic conditions, there was variability and inaccuracy in lesion quantification. However, more clinically relevant measurements of accuracy require an even more realistic lesion model.

Better models may be obtained by constructing better phantoms, or by using animal models of MS[1,2]. Better phantoms may be constructed using techniques described in Chapter 5. It may be possible to control the T1 and T2 of agarose "lesions" and "brain" by varying the concentration of Ni^{2+} and the percentage of agarose in these tissues, respectively. In addition, irregular "lesions" with known volumes may be constructed by cutting the agarose, or allowing it to gel, in arbitrary shapes, then measuring lesion volumes using a volume displacement technique. Ultimately however, an animal model of MS may be required to measure accuracy. This would involve imaging animals, then sacrificing them to determine lesion extent. One benefit of this approach would be the ability to correlate MRI measurements, such as the eW-G spectra reported in Chapter 5, with lesion pathology.

The variability study reported in Chapters 3 and 4 assumed that the operators were drawn at random from a larger population of operators. This has the effect of increasing the uncertainty in operator performance, and increasing the estimates of operator variability.[3] In some laboratories, and in some clinical trials, the operators are known in advance, fixed, and highly experienced. In this situation, the results of Chapters 3 and 4 overestimate operator variability, and thus the standard error of measurement (SEM). It might be useful to repeat the experiment described in Chapter 3 with a group of highly experienced lesion quantifiers, and use a slightly different statistical analysis which assumes that they are the only operators of interest, and not randomly drawn. The results will be specific to the group of operators, but might indicate lower operator variability, or less distinction between manual and assisted quantification.

The experiment described in Chapter 3 might also provide a useful framework for comparing the performance of new quantification techniques. Other investigators can use a similar design, and the same statistical methods, to compare their algorithms to my algorithm, or manual quantification. All that is required is that enough lesions, operators, and repeated measures are utilized to give significant results. Techniques for estimating these numbers are described elsewhere, and a pilot study may be required.[4] Nevertheless, the results should be largely independent of the lesions and operators selected, and thus useful for comparing the effects of new algorithms. Finally, this design may be used to determine the effect on operator variability of new imaging sequences. For example, fluid-attenuated inversion recovery (FLAIR) reduces the signal from CSF, heavily weights images by T2, and increases the visibility of WM lesions.[5,6] Operators could repeat measurements of lesions in standard spin-echo images, and in FLAIR images to determine the SEM and reliability in each case.

6.2.3 Clinical Applications

In Chapter 5 I described a technique to visualize the intensity distributions of lesions in multidimensional feature space by projecting the distributions down to 1 dimensional equivalent WM-GM spectra (eW-G). These spectra reflect the underlying ρ and T2 distribution of lesions, and may provide information about lesion pathology. Additional insight may be provided by further studies involving Experimental Allergic Encephalomyelitis (EAE), an animal models of MS.[1,2] Both acute and relapsing EAE models can be induced in guinea pigs, and their clinical status monitored. By imaging then sacrificing animals with different degrees of disability, and with different EAE models, it may be possible to correlate eW-G spectral characteristics with specific pathological findings and/or clinical status in these guinea pigs. Also, it may be possible to determine the effect, if any, of increased blood-brain barrier (BBB) permeability on eW-G spectra. Gadolinium DTPA enhanced MRI of the previously mentioned guinea pigs could be performed to indicate regions of increased permeability. eW-G spectra could then be measured in these areas, and the spectral shapes compared to controls.

These studies may also help define a quantitative measure of eW-G spectra which is related to pathology Changes in such a measure over time may be used then to indicate disease activity. Since the eW-G spectra is determined from standard spin-echo MS patient exams, its sensitivity at detecting disease activity could be examined by retrospective analysis of MR exams from completed or ongoing clinical trials. Changes in the spectral measure over time could be calculated and compared to clinical and lesion volume based measures of activity.

The data reduction techniques employed to construct the eW-G spectra may be useful to help visualize additional information provided by new imaging sequences. For example, magnetization transfer imaging (MTI) shows great promise for determining lesion age, and indicating regions of demyelination.[7,8] The spectra determined from multidimensional feature spaces which include data

from MTI, or other new sequences, may provide a more specific measure of pathology. Similarly, new functional imaging techniques may be utilized to incorporate functional information into feature spaces. My system may be used then to analyze function in specific anatomical regions, or anatomy in specific functional areas.

An experiment could be performed also to determine the sensitivity of assisted quantification at determining disease activity. Retrospective measurement of lesion burden could be made using assisted quantification in MR exams from completed or ongoing clinical trials. The reduced variability of assisted quantification should allow detection of smaller changes in lesion volume, and thus more sensitive indication of disease activity, than manual quantification techniques.

Finally, the system I have developed should have practical application for the visualization and analysis of MR images of other neurological disorders. Virtually any pathological state which can be observed in an MR exam of any protocol, may be analyzed using this system. This will include, for example, neurological lesions due to vascular dementia, stroke, cancer or Alzheimer's Dementia.

6.3 References

1 S.J. Karlik, J.J. Gilbert, C. Wong, M.K. Vandervoort, J.H. Noseworthy, "NMR Studies in Experimental Allergic Encephalomyelitis (EAE): Factors which contribute to T1 and T2 values." Magnetic Resonance in Medicine, 14, 1-11, (1990).

2 S.J. Karlik, E.A. Grant, D.H. Lee, J.H. Noseworthy, "Gadolinium Enhancement in Acute and Chronic Progressive Experimental Allergic Encephalomyelitis in the Guinea Pig." Magnetic Resonance in Medicine, 30: 326-331, (1993).

3 M. Eliasziw, S.L. Young, M.G. Woodbury, K. Fryday-Field, "Statistical Methodology for the Concurrent Assessment of Interrater and Intrarater Reliability: Using Goniometric Measurements as an Example". Physical Therapy, 74(8), 777-788 (1994).

4 P. Armitage, G. Berry, *Statistical methods in medical research.* 2nd edition. Oxford: Blackwell Scientific Publications, 1-559 (1990).

5 D.J. Thomas, J.M. Pennock, J.V. Hajnal, I.R. Young, G.M. Bydder, R.E. Steiner, "Magnetic Resonance imaging of spinal cord in multiple sclerosis by fluid-attenuated inversion recovery". Lancet, 341(8845), 593-594, (1993).

6 A. Oatridge, J.V. Hajnal, F.M. Cowan, C.J. Baudouin, I.R. Young, G.M. Bydder, "MRI diffusion-weighted imaging of the brain: contributions to image contrast from CSF signal reduction, use of a long echo time and diffusion effects". Clinical Radiology, 47(2), 82-90, (1993).

7 J.F. Hiehle, R.I. Grossman, K.N. Ramer, F. Gonzalez-Scarano, J.A. Cohen, "Magnetization transfer effects in MR-detected multiple sclerosis lesions: comparison with gadolinium-enhanced spin-echo images and nonenhanced T1-weighted images". American Journal of Neuroradiology, 16(1), 68-77, (1995).

8 M. Filippi, A. Campi, V. Dousset, C. Baratti, V. Martinelli, N. Canal, G. Scotti, G. Comi, "A magnetization transfer imaging study of normal-appearing white matter in multiple sclerosis". Neurology, 45, 478-482, (1995).