Time-series analysis of MRI intensity patterns in multiple sclerosis

Dominik S. Meier,* and Charles R.G. Guttmann

Center for Neurological Imaging, Department of Radiology, Brigham and Women’s Hospital, Harvard Medical School, 221 Longwood Avenue, RFB 396, Boston, MA 02115

Received 13 February 2003; revised 2 May 2003; accepted 6 June 2003

Abstract

In progressive neurological disorders, such as multiple sclerosis (MS), magnetic resonance imaging (MRI) follow-up is used to monitor disease activity and progression and to understand the underlying pathogenic mechanisms. This article presents image postprocessing methods and validation for integrating multiple serial MRI scans into a spatiotemporal volume for direct quantitative evaluation of the temporal intensity profiles. This temporal intensity signal and its dynamics have thus far not been exploited in the study of MS pathogenesis and the search for MRI surrogates of disease activity and progression. The integration into a four-dimensional data set comprises stages of tissue classification, followed by spatial and intensity normalization and partial volume filtering. Spatial normalization corrects for variations in head positioning and distortion artifacts via fully automated intensity-based registration algorithms, both rigid and nonrigid. Intensity normalization includes separate stages of correcting intra- and interscan variations based on the prior tissue class segmentation. Different approaches to image registration, partial volume correction, and intensity normalization were validated and compared. Validation included a scan–rescan experiment as well as a natural-history study on MS patients, imaged in weekly to monthly intervals over a 1-year follow-up. Significant error reduction was observed by applying tissue-specific intensity normalization and partial volume filtering. Example temporal profiles within evolving multiple sclerosis lesions are presented. An overall residual signal variance of 1.4% ± 0.5% was observed across multiple subjects and time points, indicating an overall sensitivity of 3% (for axial dual echo images with 3-mm slice thickness) for longitudinal study of signal dynamics from serial brain MRI.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Serial MRI; Morphometric analysis; Brain; Multiple sclerosis; Neurodegenerative diseases

Introduction

In progressive neurological disorders, such as multiple sclerosis (MS), magnetic resonance imaging (MRI) follow-up is used to monitor disease activity and progression and to understand the underlying pathogenic mechanisms. A specific aim in MRI assessment of MS is to differentiate among inflammation, brain parenchymal degeneration, and reparatory processes, such as remyelination or fibrillary astrocytosis. Although MRI appears to be collectively sensitive to these changes, MRI-derived surrogate metrics thus far lack specificity in differentiating between them (Barkhof, 2002).

* Corresponding author. Center for Neurological Imaging Brigham and Women’s Hospital Fax: +1-617-264-5154.
E-mail address: meier@bwh.harvard.edu (D.S. Meier).

In this work we evaluate methods of time-series analysis (TSA) in an effort to characterize the dynamics of tissue damage and repair in MS. By the term time-series analysis we mean the direct quantitative analysis of signal intensity in the time domain, obtained from serial MRI. This TSA differs from longitudinal morphometric studies in that feature extraction (segmentation) and the associated data reduction at the level of a single exam are avoided; instead the integration into a four-dimensional (4D) spatiotemporal data set is sought. Typical serial quantitative MRI first obtains morphometric variables (surrogates) of interest from each exam and then studies their progression over time (Fisher et al., 2002; Ge et al., 2000; Kikinis et al., 1999; Pelletier et al., 2001; Rovaris et al., 1999).

The presented TSA pipeline includes segmentation and filtering steps applied at each time point, but in concept seeks the integration of the time-domain prior to surrogate...
definition and analysis. This can lead to a recursive scheme that iterates between the intertwined steps of segmentation and characterization, analogous to adaptive strategies that iterate between tissue segmentation and bias field correction (van Leemput et al., 1999; Wells et al., 1996b). This has the specific advantage that temporal continuity and differentiability constraints can be applied/exploited to improve feature extraction (segmentation) as well as spatial and intensity normalization, i.e. making use of the fact that the acquired images represent an ordered series. In principle this analysis task is not very different from the processing of functional MRI (fMRI) data. However, critical differences remain that caution against direct applicability of fMRI analysis tools: (1) in the follow-up of neurological diseases the time intervals between subsequent scans are often not constant and are long enough for significant scanner drift to occur; (2) hardware and software changes may prompt substantial differences in image transfer function; (3) in neurodegenerative diseases the dynamics of pathogenic processes have different rates and heterogeneous distributions; and (4) unlike fMRI, there is no reference paradigm against which signal variations can be matched. Finally, the metric of interest is not so much the identification or delineation of significant signal changes, as are their inherent dynamics. Because of these issues, along with limitations in follow-up exam frequency, the intensity signal and its dynamics have thus far received only little attention as direct metrics. It has been stressed, however, that the dynamics of MRI lesions may provide valuable specificity in terms of differentiating stages or clinical subtypes of MS (Arnold et al., 2002; Matthews et al., 2001).

In this article we examine the principal methods and propose a first set of tools to explore the time domain of signal intensity in serial MRI, focusing primarily on technical aspects of serial registration and intensity normalization. Different schemes for registration and intensity normalization are validated and compared as modules in a standard processing pipeline. Validation occurred with respect to reproducibility and sensitivity to change. Our testbed consists of frequent (up to weekly) MRI follow-up examinations of MS patients (Cotton et al., 2003; Guttmann et al., 1995; Weiner et al., 2000).

The spectrum of related work spans investigations of

Table 1
The main TSA processing steps and the implementations evaluated in this article

<table>
<thead>
<tr>
<th>Step</th>
<th>Implementation</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Tissue class segmentation</td>
<td>• Template-driven based on expectation-maximization principle, yielding ICC, WM, GM, CSF, WMSA</td>
<td>—</td>
</tr>
<tr>
<td>2 Intrascan intensity normalization</td>
<td>• Iteratively together with tissue class segmentation</td>
<td>R3m</td>
</tr>
<tr>
<td>3 Tissue masking</td>
<td>• Excluding image pixels outside the ICC and pixels classified as WMSA from contributing to the criterion function of registration (R3m) and interscan intensity normalization</td>
<td>—</td>
</tr>
<tr>
<td>4 Interascan intensity normalization</td>
<td>• Global histogram matching</td>
<td>N1</td>
</tr>
<tr>
<td>5 Spatial normalization</td>
<td>• Rigid registration</td>
<td>R1, R2, R3</td>
</tr>
<tr>
<td>6 Partial volume correction</td>
<td>• T2 maps</td>
<td>T2m</td>
</tr>
<tr>
<td>7 4D integration</td>
<td>• Nonrigid image warping</td>
<td>R5</td>
</tr>
<tr>
<td>8 ROI definition</td>
<td>• Region of interests for profile extraction are derived from either manual selection or automated from the tissue labels of the segmentation</td>
<td>—</td>
</tr>
<tr>
<td>9 Profile extraction and analysis</td>
<td></td>
<td>—</td>
</tr>
</tbody>
</table>

Note. Different versions and combinations of spatial and intensity normalization were compared and validated. Each of the serial scans is subject to an extensive and coupled processing pipeline, beginning with tissue class segmentation combined with a bias-field inhomogeneity correction. One study in the series (usually the baseline scan) is used as reference for both spatial and intensity normalization. The symbol abbreviations in the right column serve as shorthand for denoting module combinations into different pipelines. See also Tables 2 and 3 for further explanation of the abbreviations. Abbreviations: ICC, intracranial cavity; WM, white matter; GM, gray matter; CSF, cerebrospinal fluid; WMSA, white matter signal abnormalities, i.e. lesions; 4D, four-dimensional; ROI, region of interest.

Table 2
Combinations of tissue classes used for serial intensity normalization

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Derived from</th>
<th>Applied to</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1a</td>
<td>ICC − WMSA</td>
<td>ICC</td>
</tr>
<tr>
<td>N1b</td>
<td>CSF</td>
<td>ICC</td>
</tr>
<tr>
<td>N2</td>
<td>WM − WMSA</td>
<td>WM + WMSA</td>
</tr>
<tr>
<td></td>
<td>GM</td>
<td>GM</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>CSF</td>
</tr>
</tbody>
</table>

Note. Lesions (WMSA) are usually excluded from contributing to the match, but are treated as WM for adjustment. Tissue-class label lists are given to the normalization algorithm as input, hence combinations of normalization schemes are possible. The plus (+) and minus (-) signs indicate logical combinations (or, xor) of the binary label masks. N1b was tested as alternative to N1a, using the CSF for normalizing global image brightness and contrast.
registration error sources (Hajnal et al., 1995; Smith et al., 2001), difference image analysis (Tan et al., 2002b), and models for deformation analysis (Thirion et al., 1997). It also has been pointed out correctly that in serial analysis there is a need for a distinction between image resolution and accuracy of positional representation. In this context, difference images proved to be useful tools in identifying positional changes beyond image resolution levels (Hajnal et al., 1995) and in distinguishing between new, enlarging, and resolving lesion activity (Tan et al., 2002b).

Gerig et al., (2000) presented a set of pilot experiments on the discriminatory value of the time domain in serial MRI for purposes of lesion segmentation. Very promising results were obtained. It was shown that lesion segmentation is possible from the temporal information alone without use of absolute image intensities. Segmentation efforts later were also successfully applied to active lesions (Welti et al., 2001).

Rey et al., presented recent efforts of applying deformation analysis and even statistical fMRI tools for purposes of segmenting and characterizing evolving MS lesions (Rey et al., 2001, 2002; Thirion et al., 1997). No validation on clinical sets is yet available for those new techniques. In this article we provide some further methodological and validation results to help establish this important new field of quantitative serial MRI in neurological diseases. In this work we also shift the focus from lesion segmentation to lesion characterization and aim to establish a well-validated basis for application of TSA in large studies.

Methods

The processing strategy presented in this article focuses mainly on appropriate approaches for integration/fusion of serial 3D MRI scans into a 4D volume. The main steps for

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Method</th>
<th>Criterion</th>
<th>DOF</th>
<th>Interpolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Rigid registration</td>
<td>Mutual information</td>
<td>6</td>
<td>Trilinear</td>
</tr>
<tr>
<td>R2</td>
<td>Rigid + mean pose</td>
<td>Mutual information</td>
<td>6</td>
<td>Trilinear</td>
</tr>
<tr>
<td>R3</td>
<td>Rigid (SPM99)</td>
<td>Intensity differences</td>
<td>6</td>
<td>Sinc</td>
</tr>
<tr>
<td>R3m</td>
<td>Masked R3</td>
<td>Intensity differences</td>
<td>6</td>
<td>Sinc</td>
</tr>
<tr>
<td>R4</td>
<td>Affine</td>
<td>Intensity correlation</td>
<td>12</td>
<td>Trilinear</td>
</tr>
<tr>
<td>R5</td>
<td>Warping</td>
<td>Intensity correlation</td>
<td>&gt;104</td>
<td>Trilinear</td>
</tr>
</tbody>
</table>

Note. DOF, degrees of freedom; SPM, statistical parametric mapping (Friston et al., 1995; SPM, 1999).
such 4D integration are (1) spatial normalization: rigid and nonrigid registration of all serial sets to a common reference; (2) intensity normalization: adjustment of the intensity distributions; (3) artifact filtering: correction of partial volume effects and intensity inhomogeneities; (4) fusion into a 4D volume with creation of subtraction or ratio images (against a chosen baseline scan); and (5) time profile analysis and creation of parametric images and feature maps.

These steps are discussed in detail below. A summary of the proposed TSA procedure and the implemented and validated methods are summarized in Table 1.

**Image acquisition**

Development and validation occurred on MRI data sets from 46 MS patients (63% female), followed over the course...
of at least 1 year (Weiner et al., 2000). The mean age of the patient cohort was 40 ± 7 years (26–53 years). Patients were scanned, with written informed consent, in weekly to monthly intervals, with an axial dual echo protocol (PDw/T2w, TE=30/80 ms, TR=3000 ms, 192 phase-encoding steps, 256×256×54 voxels with 0.93 × 0.93 × 3 mm³ nominal voxel size). All scans were performed on a 1.5-Tesla machine (GE Signa, General Electrics, Milwaukee, WI).
Tissue class segmentation and intrascan intensity normalization

To account for field inhomogeneities, a bias-field correction algorithm was applied to each time point in conjunction with the tissue segmentation (Warfield et al., 1999; Wells et al., 1996b). The algorithm corrects for intrascan intensity variations by iterating between tissue class segmentation and bias field estimation (Fig. 4). An iterative combination of nearest neighbor multichannel clustering [expectation-maximization (EM) concept] and template-driven segmentation (Warfield et al., 1996) was applied to segment white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and white matter signal abnormalities (WMSA). The addition of a template-driven strategy to the EM criterion showed excellent accuracy and robustness (Wei et al., 2002). The tissue class segmentation formed the knowledge base for the tissue-specific intensity normalization described below.

Interscan intensity normalization

The normalization algorithm works with the segmentation data as an a priori knowledge base, using tissue class labels as input. It adjusts intensity distributions of each follow-up scan to match those of the chosen baseline. The tissue, from which the normalization is derived, and the regions to which it is applied, are distinguished in this application. This is necessary to obtain a hierarchical model and also to prevent the normalization from removing the intensity variations of interest. For example, the first-level normalization may be derived from the histogram of the entire ICC, the WM, or the CSF only but applied to the entire image. In a second iteration, normalization may be derived from normal-appearing WM (NAWM) only, but applied to the entire WM, including WMSA. The tissue class combinations tested in the experiments are summarized in Table 2.

Based on the results of the large-scale intensity relation analysis (see results), a two-parameter model assuming Gaussian intensity distributions was applied. Given a baseline (subscript 1 in equations) and a follow-up scan (subscript 2), the intensities of each tissue class in the follow-up scan are stretched by a factor \( \lambda = \sigma_1/\sigma_2 \) and shifted by an offset \( \kappa = \mu_1 - \lambda \mu_2 \), so that distribution mean \( \mu \) and standard deviation \( \sigma \) match the baseline scan. The corrected distribution then becomes \( \hat{I}_2 = \lambda I_2 + \kappa \) (where \( I_2 \) and \( \hat{I}_2 \) denote original and corrected intensities, respectively). This corresponds to individual adjustment of brightness and contrast for each image and each tissue class.

An alternative to a tissue-class-based intensity normalization is the construction of T2 maps from a dual echo scan. Modeling the PDw and T2w signals as monoexponentials with different TE yields the following:

\[
PDw = k \cdot e^{-TE_{PDw}/T2}
\]

and

\[
T2w = k \cdot e^{-TE_{T2w}/T2}
\]

and we obtain \( T2 \approx \ln(PDw/T2w)^{-1} \).

This T2 map in theory represents an absolute tissue property and would render further normalization unnecessary. The normalization scheme above was compared to this alternative in the validation experiments.

Note that the separate treatment of intra- and interscan variations arises from the processing pipeline, where intrascan intensity variations are addressed during the tissue class segmentation step (see above) and consequently not included in interscan normalization.

Table 4
Overview of data used for the different experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Patients</th>
<th>Time points</th>
<th>Scans</th>
<th>Total images</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interscan intensity</td>
<td>46</td>
<td>6–26 (22 ± 4)</td>
<td>1</td>
<td>733</td>
<td>1 year</td>
</tr>
<tr>
<td>relations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Scan–rescan</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>2 h</td>
</tr>
<tr>
<td>3. Method comparison</td>
<td>1</td>
<td>20</td>
<td>1</td>
<td>20</td>
<td>3 years</td>
</tr>
<tr>
<td>4. Robustness</td>
<td>16</td>
<td>21–26 (24 ± 1)</td>
<td>1</td>
<td>360</td>
<td>1 year</td>
</tr>
<tr>
<td>5. Example profiles</td>
<td>4</td>
<td>21–24 (23 ± 2)</td>
<td>1</td>
<td>93</td>
<td>1 year</td>
</tr>
</tbody>
</table>

Note. The entire data set included weekly to monthly-follow up of 46 patients, from which subsets were chosen for individual validation experiments.

Table 5
Comparison of tissue intensity slopes shown in Figure 2

<table>
<thead>
<tr>
<th></th>
<th>PDw</th>
<th>T2w</th>
<th>T2m</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC (ref)*</td>
<td>a = 1.000</td>
<td>a = 1.000</td>
<td>—</td>
</tr>
<tr>
<td>WM*</td>
<td>a = 0.916</td>
<td>a = 0.738</td>
<td>a &lt; 0.001</td>
</tr>
<tr>
<td>GM*</td>
<td>a = 1.063</td>
<td>a = 0.979</td>
<td>a &lt; 0.001</td>
</tr>
<tr>
<td>CSF*</td>
<td>a = 1.067</td>
<td>a = 1.432</td>
<td>a &lt; 0.001</td>
</tr>
<tr>
<td>WMSA*</td>
<td>a = 1.295</td>
<td>a = 1.350</td>
<td>a &lt; 0.001</td>
</tr>
<tr>
<td>ICC–WM*</td>
<td>( P &lt; 0.00001 )</td>
<td>( P &lt; 0.00001 )</td>
<td>—</td>
</tr>
<tr>
<td>WM–GM*</td>
<td>( P &lt; 0.00001 )</td>
<td>( P &lt; 0.00001 )</td>
<td>—</td>
</tr>
<tr>
<td>GM–CSF*</td>
<td>( P = 0.324 )</td>
<td>( P &lt; 0.00001 )</td>
<td>—</td>
</tr>
<tr>
<td>CSF–WMSA*</td>
<td>( P &lt; 0.00001 )</td>
<td>( P = 0.003 )</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. All slopes are significantly different from each other, except for the GM-CSF pair in PDw images. The slopes of the T2-map (Fig. 2c) are close to zero, as predicted by the monoexponential model.

* Slopes as shown in Fig. 2.

b \( P \) values of slope comparison.
Spatial normalization

Several algorithms and approaches to image registration, both rigid and nonrigid, were evaluated and compared. All methods were fully automated and used intensity-driven criteria based on mutual information (Wells et al., 1996a), intensity differences (Friston et al., 1995, SPM, 1999), or correlation coefficients (Guimond et al., 2001). The match criteria and degrees of freedom (DOF) used for each method are summarized in Table 2. Fine registration included 12 DOF (3 DOF each for translation, rotations, scaling, and shearing); the warping (R5) employed a simulation of an elastic deformation with thousands of DOF, where each pixel moves independently but constrained against overlap or separation (Guimond et al., 2001). The purpose of a second registration iteration involving nonrigid alignment was to account for eventual tissue shifts associated with atrophy (Guimond et al., 2002a), as well as voxel size differences and geometric distortions that have been observed in longitudinal studies (Smith et al., 2001).

Different strategies of coregistering serial MRI were also compared (Table 3): aligning all images to a mean reference (R2), as suggested by Smith et al., (2001), rather than aligning to an individual reference image (see Appendix); registration with lesion masking (R3m) (i.e., excluding signal anomalies from the criterion function; Brett et al., 2001; Guimond et al., 2002b); affine registration (R4); and image warping (R5) using a deformation model based on “demons” and an intensity-correlation criterion (Guimond et al., 2001).

Partial volume artifact correction

Ideally, all identifiable artifacts that affect intensity characteristics are to be removed prior to intensity normalization to prevent them from biasing the correction. One key artifact stems from different amounts of partial volume (PV) effect due to variations in head position relative to the slice plane. These variations in strength and direction of PV artifacts create an anisotropic blurring that cannot be compensated by spatial registration alone. This subvoxel effect becomes apparent in the difference image as coherent patterns of edge artifacts, similar to those of misregistration or tissue deformation.

One straightforward approach is to apply a deliberate additional “coloring” of the spatial autocorrelation, similar

<table>
<thead>
<tr>
<th>ROI/method</th>
<th>1: R1 (%)</th>
<th>2: R1 + N1a (%)</th>
<th>3: R1 + N1a + PV (%)</th>
<th>4: R1 + N1b + PV (%)</th>
<th>5: R1 + N1a + N2 + PV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>corp. call.</td>
<td>3.60</td>
<td>2.56</td>
<td>1.83</td>
<td>2.14</td>
<td>1.70</td>
</tr>
<tr>
<td>brainstem</td>
<td>4.43</td>
<td>2.06</td>
<td>1.22</td>
<td>1.60</td>
<td>1.01</td>
</tr>
<tr>
<td>confluent lesio</td>
<td>3.55</td>
<td>1.14</td>
<td>0.93</td>
<td>1.24</td>
<td>1.25</td>
</tr>
<tr>
<td>focal lesion</td>
<td>3.71</td>
<td>1.63</td>
<td>0.84</td>
<td>0.48</td>
<td>1.00</td>
</tr>
<tr>
<td>caudate</td>
<td>2.69</td>
<td>2.04</td>
<td>1.47</td>
<td>1.68</td>
<td>1.45</td>
</tr>
<tr>
<td>CSF</td>
<td>4.68</td>
<td>1.98</td>
<td>1.31</td>
<td>1.27</td>
<td>1.21</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>3.77% ± 0.71</strong></td>
<td><strong>1.90% ± 0.48</strong></td>
<td><strong>1.27% ± 0.36</strong></td>
<td><strong>1.40% ± 0.56</strong></td>
<td><strong>1.27% ± 0.27</strong></td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>—</td>
<td>0.001 (1)</td>
<td>0.001 (2)</td>
<td>0.016 (2)</td>
<td>0.006 (2)</td>
</tr>
</tbody>
</table>

Note. 1: rigid registration (R1); 2: additional overall (ICC-based) intensity normalization (N1a); 3: additional partial volume correction (PV); 4: global correction based on CSF signal (N1b); and 5: additional tissue class intensity normalization (N2). The six ROI are shown in Fig. 5. Significant improvements were observed for all methods compared to R1 alone. Further improvements were achieved by the addition of PV and N2 relative to N1. The bottom row gives the P values of method comparison (method in parentheses). Note, however, that this scan–rescan experiment does not contain scanner drift, as opposed to the results shown in Table 7.

Fig. 5. Regions of interest (ROI) chosen to evaluate signal variations in a scan-rescan experiment. Each ROI comprises five pixels (center + four-neighbors), indicated by the cross-mark (not to scale). The six ROIs were (1) right corpus callosum, (2) brainstem, (3) confluent periventricular WM lesion, (4) small focal lesion in cortical WM, (5) caudate nucleus (GM), (6) CSF, at the boundary of the posterior horn of the lateral ventricle. Locations were chosen near tissue boundaries to assess precision in areas most critical to registration and partial volume artifacts. The variations for these ROI are reported in Table 6.
to the temporal filtering applied in fMRI (Friston et al., 2000; Locascio et al., 1997; Woolrich et al., 2001). This filtering trades spatial resolution for sensitivity and spatio-temporal reliability of the signal. We therefore tested the efficacy of applying an additional one-dimensional Gaussian filter step in slice direction only (i.e., the predominant anisotropy direction, with a width corresponding to half the slice thickness), applied to all scans after normalization and prior to fusion into a 4D volume. The idea is to equalize the levels of blurring in all scans before creating ratio or subtraction images.

Figure 1 shows an example of the qualitative reduction of artifacts and enhanced sensitivity to change in a difference image with and without use of PV filtering. Recent
studies also showed that lesion conspicuity is improved in difference images (Tan et al., 2001).

Validation experiments

Five sets of validation experiments were performed, aimed at estimating precision and sensitivity of TSA-extracted intensity profiles (Table 4) and the efficacy of individual processing steps. Each experiment is described in detail below.

Intensity relations

To determine the appropriate intensity correction model, we evaluated the specific tissue appearance, relative to variations in global image brightness, on a large set of dual-echo scans (N=733 scans, covering the entire brain of 46 patients, 256×256×52, 0.93×0.93×3 mm³ voxels). Each image was segmented into the intracranial cavity (ICC) and its subcomponents of white matter (WM), gray matter (GM), and white matter signal anomalies (WMSA), and their mean intensities were plotted against the mean intensity of all tissue classes combined, i.e., the intracranial cavity (ICC). Based on regressions on these results, we developed a hierarchical tissue-specific intensity normalization, assuming normally distributed intensities for each class, as described above.

Scan–rescan validation

The scan–rescan exam involved a set of six repeated dual-echo scans, obtained from a single MS patient (patient 1). The patient left the scanner room between subsequent acquisitions and was positioned alternatingly by two radiology technicians (Guttmann et al., 1999). The amount of actual positional variability was assessed in terms of translation and rotation parameters obtained from the registration. The six scans thus represent the same anatomy, and changes across the series reflect acquisition-related artifacts, such as partial volume effects or residuals in registration and intensity normalization, in the absence of new pathology. The observed residual signal variance serves as estimate of the method’s precision and its expected sensitivity. Although this experiment does not include scanner drift, it includes all other sources of variation to be expected in real-life follow-up trials.

Six regions of interest (ROI, five-pixels; Fig. 5) were selected in different slices, representing different anatomical locations and tissue classes. To obtain a conservative estimate, ROIs at or close to tissue boundaries were selected, where artifacts such as partial volume effects and misregistration are likely to be expressed most strongly. Those regions are also physiologically relevant, because they are likely locations of developing pathologies and longitudinal change.

To estimate the contribution of misregistration to overall precision, registration error was evaluated both as residual intensity variance [coefficient of variation (CV)=SD/M]}
and as root-mean-square (RMS) of fiducial distances (in millimeters). The latter was obtained by permuting the baseline scan and applying both forward and backward transforms in series, i.e., by direct comparison of the registration parameters for permuted pairs (e.g., register scan 1→2 vs 2→1). Residual differences were expressed as fiducial point distances after passing through these two symmetrical registrations ($R_{ji}$ followed by $R_{ij}$): 

$$e = \sqrt{\frac{1}{2} \sum (p - (R_{ij} \cdot R_{ji}) \cdot p)}$$

where $p = [x, y, z]^T$ are the coordinates of the ROI fiducial points, and $R_{ij}$ denotes the registration matrix of time point $j$ to time point $i$.

### Table 7

<table>
<thead>
<tr>
<th>ROI</th>
<th>1: R3</th>
<th>2: R3</th>
<th>3: R3</th>
<th>4: R3 + N1</th>
<th>5: R3 + N1</th>
<th>6: N1 + R3</th>
<th>7: N1 + N2</th>
<th>8: N1 + N2</th>
<th>9: N1 + N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF</td>
<td>10.2%</td>
<td>7.7%</td>
<td>2.9%</td>
<td>2.5%</td>
<td>2.5%</td>
<td>1.2%</td>
<td>1.1%</td>
<td>1.1%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Anterior commissure (WM)</td>
<td>11.1%</td>
<td>47.5%</td>
<td>2.6%</td>
<td>1.9%</td>
<td>1.9%</td>
<td>3.5%</td>
<td>1.3%</td>
<td>1.3%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Periventricular Lesion (WM)</td>
<td>12.0%</td>
<td>0.6%</td>
<td>2.7%</td>
<td>1.4%</td>
<td>2.3%</td>
<td>1.4%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>0.8%</td>
</tr>
<tr>
<td>Caudate (GM)</td>
<td>10.6%</td>
<td>46.3%</td>
<td>2.1%</td>
<td>1.5%</td>
<td>1.2%</td>
<td>2.5%</td>
<td>1.5%</td>
<td>1.5%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Cortical WM</td>
<td>12.8%</td>
<td>2.0%</td>
<td>3.3%</td>
<td>2.5%</td>
<td>2.9%</td>
<td>2.3%</td>
<td>1.2%</td>
<td>1.3%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Cortical GM</td>
<td>12.1%</td>
<td>42.5%</td>
<td>3.1%</td>
<td>2.5%</td>
<td>2.8%</td>
<td>1.6%</td>
<td>1.6%</td>
<td>1.6%</td>
<td>0.8%</td>
</tr>
<tr>
<td>Mean</td>
<td>11% ± 1%</td>
<td>24% ± 23%</td>
<td>3% ± 0%</td>
<td>2.0% ± 0.5%</td>
<td>2.3% ± 0.6%</td>
<td>2.1% ± 0.9%</td>
<td>1.3% ± 0.2%</td>
<td>1.3% ± 0.2%</td>
<td>0.9% ± 0.1%</td>
</tr>
<tr>
<td>$P$</td>
<td>0.118 (1)</td>
<td>0.000 (1)</td>
<td>0.001 (3)</td>
<td>0.001 (3)</td>
<td>0.084 (3)</td>
<td>0.036 (3)</td>
<td>0.044 (7)</td>
<td>0.003 (8)</td>
<td>0.015 (7)</td>
</tr>
</tbody>
</table>

**Note.** Different strategies (1–9) for registration and intensity normalization are compared. The $P$ value (bottom row) gives the significance of the comparison with the most closely related method (in parentheses). Significant improvements in CV are achieved for all methods compared to registration alone (1). Normalization and lesion masking prior to registration (7) yielded further improvement. Extension to nonrigid spatial normalization (9) brought yet a further significant CV reduction. Each ROI contains five voxels (center + four-neighbors); the anatomic locations of the ROI are shown in Fig. 6. For a legend of the method components refer to Table 3 and Table 2.

### TSA pipeline comparison

An extended test of nine different TSA processing pipelines was performed on the same patient (patient 1) as used in the scan–rescan study (Guttmann et al., 1999; Kikinis et al., 1999). Validation metric was again residual signal variance (CV=SD/M), evaluated in small ROI (five pixels) of areas selected by visual inspection of an experienced observer as “silent” regions where no qualitative change occurs over the follow-up period. This second validation was necessary because the scan–rescan experiment cannot simulate long-term scanner drift or pathological processes, such as inflammation, neurodegeneration, or repair. The purpose of this test was to assure that the TSA integration methods (spatial and intensity normalization, bias, and PV corre-
across time (CV) was again evaluated in seven
applied to all of these cases and residual signal variance
period (Cotton et al., 2003). The same TSA pipeline was
of contrast-enhancing Gd-DTPA lesions) during the exam
active disease
images (Fig. 2). Regression results
examined how the mean intensity of each tissue class scales
fi
that the slopes of each tissue class are signiﬁcantly different
P < 10−5), with the exception of the GM-CSF pair in PDw
images (P = 0.32). These observed dissociations in tissue
response to overall image brightness are not merely arising
from varying contributions of individual tissue classes to the
global image brightness metric. The same results were ob-
erved with mean WM or GM intensity as metric for image
brightness. These results therefore suggested the use of a
tissue-speciﬁc linear intensity correction model, which was
applied in this work. The intensity relations of the T2 map
showed zero slopes (Fig. 2c). This indicates that such T2
maps represent absolute tissue properties and as such could
be employed in TSA without separate intensity normaliza-
tion. The T2 maps were therefore included in the validation
experiments.

An example of original and corrected tissue class histo-
grams is shown in Fig. 3. As expected from Fig. 2, diffe-
rential responses of tissue intensities are strongest in WM
and least in GM. A tissue-speciﬁc normalization is therefore
of particular importance for studies focusing on WM pa-
thology.

An example of the intrascan intensity normalization, re-
ﬂective of spatial variation, is shown in Fig. 4.

Scan–Rescan Validation

Four registration schemes were compared with respect to
spatial precision (in millimeters) and residual signal var-
ance, using the coefﬁcient of variation (CV) of PDw signal
intensities within each of six individual anatomical loca-
tions (Fig. 5). N1a intensity normalization was added to all
registration schemes to allow comparison. The standard-
deviation of the CV measure was obtained from ﬁve re-
peated registrations with permuted reference scans; it in-
dicates registration robustness. The average signal intensity
CV and average standard deviation across all permutations
were 1.6 and 0.6%, respectively. Registration error in mill-
imeters was assessed from residual location differences
after sending the ROI coordinates through all combinations
of pairwise registrations (N = 180), as described above.
An average registration error of 0.35 ± 0.20 mm was
observed. No signiﬁcant differences were found among
the four registration schemes (R1–R4; see Table 3). In
the subsequent natural history validation, we therefore
used only R3 and R3m as rigid registration method of
choice.

The effect of individual TSA processing steps on error
reduction is summarized in terms of signal CV in Table 6.
The images and ROI used were the same as above (Fig. 6).
The overall CV was reduced from 3.8 ± 0.7% (registration
only) to 1.3 ± 0.3% by adding intensity normalization and
partial volume ﬁltering.

To assess validity of the validation results in the scan–
rescan study, initial misalignment between scans for patient
1 was compared to the 3-year follow-up experiment re-
ported below. Translational misalignment for the scan–res-
can and 3-year follow-up was 3.7 ± 1.8 and 11.0 ± 2.0 mm,
respectively, and rotational misalignment was 4.8° ± 1.5°
and 4.8° ± 1.6°, respectively. Neither translational nor
rotational variance of misalignment were signiﬁcantly dif-
ferent between the two (F test for variance, P > 0.1); how-
ever, translational mean shifts were lower in the scan–
rescan study (t test for means, P < 10−6). The mean value
depends on the choice of reference scan and is therefore less
conclusive than the variance. Also, for translations above a
single voxel dimension no effects/differences in interpola-
tion effects are expected. Therefore, to test if this diffe-
rence in translation mean may affect the ﬁnal registration pose,
the registration (R3) was repeated another 49 times with (nor-
mally distributed) random translational displacements
added to the scan–rescan set to match the range observed in
the natural history study. In all 50 cases (N = 250 scans) the
registration converged to the same result. We therefore
conclude that the lower translational misalignment did not
signiﬁcantly affect the sensitivity measures reported in the
scan–rescan study and that they are representative for
achievable registration precision in a typical follow-up
exam.
TSA Pipeline Comparison

In the natural history study spanning a follow-up of 1 year (Table 4), an average CV of 11 ± 1% was observed with R3 alone, which was reduced by TSA processing to 1.3 ± 0.2% (Table 7). Intensity normalization and lesion masking prior to registration showed significant improvements (P = 0.036) over the inverse sequence. Extension to non-rigid spatial normalization (R5) brought yet a further significant reduction. The T2 mapping does yield a valid series (see below), but carries the problem of numerical instability, the closer the PDw and T2w of the mapped tissue match in intensity. To obtain a roughly comparable CV metric, the intensities of the T2 map were scaled to match those of the PDw, based on the histogram of the ICC (equivalent to a N1a normalization). The reduced robustness of the T2 map is apparent in the standard deviation of the CV in Table 7.

To assess potential effects of atrophy on the observed sensitivity measures in Table 7, we plot tissue volumetric data and the brain parenchymal fraction over the entire 3-year period in Fig. 7. Linear regression on the BPF showed a very small but significant trend of 0.5% reduction in BPF per year (P = 0.0015). No change in the warp field was observed. Patient 1 suffered one exacerbation around week 113, which was treated for 8 days with dexamethasone (Decadron, Merck, West Point, MA) (Guttmann et al., 1999).

TSA robustness validation

To test TSA feasibility on a wider spectrum of subjects, method 7 (N1a + N2 + R3m + PV; Table 7) was applied to N = 16 patients with 21–26 follow-up scans each (N = 360 scans total; Table 4). Seven ROI (equivalent to the ones shown in Fig. 6) with no apparent activity were selected by an experienced observer and the CV was calculated (Fig. 8).
A CV of 1.4% ± 0.5% was observed across all cases and all ROI; \((N = 112, \text{arising from seven ROI from 16 patients})\), which was a significant reduction from 2.4% ± 0.7% without the PV filter \((P < 10^{-9})\).

The amount of applied intra- and interscan intensity corrections are reported as histograms obtained from all processed images \((N = 360 \text{ scans, 16 patients})\) in Fig. 9 and Fig. 10. Intrascan correction was on average 7.6% ± 1.3%, interscan correction factors were 0.971 ± 0.183 \((N1a)\), 0.998 ± 0.054 \((N2:WM)\), and 0.989 ± 0.051 \((N2:GM)\), respectively. Example images of the intrascan correction are also shown in Fig. 4.

**Time profiles**

The following example profiles were taken from four different patients of the natural history study \((Table 4)\). They show signal dynamics of active and chronic MS lesions as well as NAWM and illustrate the effect of TSA processing on extracting the putative information of these dynamics.

Example profiles from a single lesion and the qualitative effect of TSA processing are illustrated in Fig. 11. The most dominant changes are recognizable in both the uncorrected and corrected profile, yet quantitative analysis of the dynamics is clearly not feasible without TSA correction. Example time profiles of another patient \((processed \text{ with the R3 + N1a + N2 + PV pipeline})\) over the same period \((a 56\text{-week follow-up study, weekly to monthly scans})\) are shown in Fig. 12.

A series of qualitative lesion characteristics are apparent from these profiles: The lesions in Fig. 12a and Fig. 12b both appear at the same time but have different rates and levels of recovery, indicating different processes due to lesion location and/or lesion type. Figure 12c shows a lesion cluster for the same patient in a different location. Chronic \((ROI 6)\) and new inflammatory lesions are visible, the latter again with different rates of evolution, but overall similar spatiotemporal shape characteristics. The intensity profiles also depict a heterogeneity in lesion dynamics, both within and between lesions. Slow and incomplete recovery at the lesion center compared to marginal regions is apparent. Note also that within individual lesions, level and rate of recovery seem to relate to the intensity of the initial insult. These results clearly suggest the maximal insult at the lesion center and support the notion of concentric lesion development \((Guttmann \text{ et al., }1995)\). Inflammation peaks during initial lesion formation usually last no longer than 1 month \((Fig. 12)\) and may therefore go undetected on follow-up scans more than 1 month apart.

The validation results in Table 7 and example profiles \((Figure 11)\) demonstrate that mere coregistration without intensity normalization is insufficient for TSA analysis and meaningful interpretation of the signal dynamics. Figure 13 shows the N1a intensity normalization for this series at each time point. A T2 map obtained from the same series is shown in Fig. 14; yielding similar qualitative profiles of inflammation and recovery.

Example profiles of three additional patients \((6, 12, \text{and }16)\) are shown below \((Figs. 15–17. )\)

**Discussion**

**Registration**

No significant differences with respect to signal sensitivity \((CV \text{ measure})\) were observed between the different rigid registration methods tested. Extension to nonrigid spatial normalization \((R5)\), however, yielded an additional significant reduction in signal CV. Consistently highest CV was observed in regions of the brainstem, \(R5 \text{ does, however, not mask WMSA from the cost function and therefore might remove variations of interest; it was therefore not applied to patients with apparent MRI activity.}\)

The elastic warping is applied with the intention to remove local distortion artifacts as well as biologically related deformation such as atrophy-related distal tissue shifts, so that the spatial acuity of local sampling across time is maintained. The contribution of diffuse atrophy to the warp is testable to a limited extent by monitoring longitudinal correlations between volumetric data \((e.g., \text{BPF})\) and the deformation field. No such correlations were observed with the patient data used for validation, and the deformation field remained around one pixel at all times \((Fig. 7)\). Atrophy rates, assessed by BPF, of \(-0.3/\text{year} \text{ and } -0.5/\text{year}\) have been reported in treatment and placebo cohorts, respectively \((Fisher \text{ et al., } 2002)\). The latter rate matches the one reported in Fig. 7. The basic follow-up period of the main patient cohort \((N = 46)\) in the natural history study was 1 year. We therefore consider it unlikely that atrophy contributed significantly to the results shown herein.

The sequencing of spatial and intensity normalization appears to be significant \((Table 7)\). Applying the normalization and lesion masking prior to registration yielded significantly higher precision than did the inverted order. This is most likely due to the optimization in the automated registration converging closer to the global optimum, especially if using a criterion function based on intensity differences.

**Partial volume filtering**

Longitudinal image analysis contains an inevitable inherent reduction in spatial resolution, occurring upon integrating the data into a 4D volume. Structures falling in between slice planes are smoothed in acquisition to an irrecoverable degree. Because direction and extent of this PV smoothing is different in each follow-up scan, it can no longer be dismissed as systematic bias. The principal purpose of the applied “coloring” PV filter was to make this de
facto reduction in resolution explicit. In consequence we would expect the PV filter to yield a signal improvement in a domain different from the filter dimension, which is apparent the validation results (Table 7 and Fig. 8), i.e., we filter in the spatial direction where the artifact occurred but find improvement in the time direction where we wish to achieve maximal sensitivity. Sensitivity to change is enhanced at the loss of spatial resolution. For example, it is not possible to accurately track a 1-mm lesion with a serial MRI protocol of 3-mm slice thickness. Even though the lesion may appear in some or all of the scans, the time-series signal will exhibit random fluctuations due to the dominant PV effect caused by pseudorandom position of the lesion relative to the slice plane. For 3D acquisition methods with isotropic image data (Mugler et al., 2000; Tan et al., 2002a) we expect this artifact to be less of an issue.

*Intensity normalization*

An average CV of 1.4 ± 0.5% was observed consistently across patients (N = 16) and different ROI. We therefore expect “true” signal changes above 3% (95% confidence interval as twice the CV) to be detectable reliably. This residual 3% sensitivity also indicates that small shoulders and other changes in profile curvature might reflect actual tissue changes versus TSA artifacts. Such small changes might be obfuscated by applying N1 only. For characterization of intraleSION heterogeneity the sensitivity achieved by most of the tested normalization methods is probably sufficient. However, for studying lesion dynamics and profile curvature, it may be necessary to choose more advanced and computationally more expensive spatiotemporal normalization methods (e.g., method 9 in Table 7) to reliably dissociate artifacts from signal variations representing actual tissue changes.

The reason for a hierarchical model (i.e., N1 followed by N2) arose from assumptions about the physical origin of signal variations as well as concerns on spatial continuity and differentiability of the final intensity signal. Brightness differences arising from technical sources, such as different receiver gain, are more effectively addressed by a method that removes differences in global image brightness before correcting residual differences between tissue classes. With this approach the residual tissue-specific corrections are small and artifactual edge enhancement at tissue boundaries is avoided.

It should be pointed out that tissue segmentation is an integral part of the TSA concept and that the steps of static morphometric assessment and integrated TSA complement each other. Several longitudinal and cross-sectional morphometric studies on the significance of T1 lesions used relative hypointensity as reference measure (Bitsch et al., 2001; van Walderveen et al., 1998), i.e., lesion intensity relative to NAWM. This effectively corrects for both scanner variability and normal interindividual tissue variability and in essence corresponds to a first-order model of tissue-specific normalization (N1) as used here. Note that the WMSA are coupled with the NAWM in the normalization scheme, i.e., they do not participate in calculating the correction but are combined with NAWM when the correction is applied. Absolute measurement of MR properties, such as the T2 mapping shown in Fig. 14, bypass a normalization step by obtaining an absolute tissue property. However, the logaritmic mapping involved can lead to numerical instabilities in cases where the intensities of both channels are similar (Table 7).

*Relations to pathophysiology*

This article focuses on assessing the feasibility and precision of measuring MRI intensity profiles from long-term serial MRI follow-up. The structure inherent in the temporal intensity profile is apparent, but it remains to be shown if disease-specific surrogates can be derived from this metric. We intend to address this issue in future work. The following section discusses some aspects of relevance for such assessment.

Several investigators have suggested that the branching point on MS lesion development into a remyelinating or inactive lesion appears to occur early during the active state of the lesion (Bitsch et al., 2001; Doussset et al., 1998). A consistent characteristic of the obtained lesion profiles is that levels of residual hyperintensity (incomplete recovery) are found consistently in regions where the initial signal peak was strongest (Fig. 12). It remains to be tested to which extent the observed episodes of conspicuous hyper-intensities actually reflect inflammation and/or edema. We intend to apply the TSA method in long-term follow-up to evaluate the predictive power of these early dynamics toward residual lesion size and within-lesion changes. The high sensitivity of the TSA method also invites the study of NAWM in the investigation of subclinical activity, which is estimated to be of more continuous nature than suggested by clinical surrogates (Barkhof, 1999; Rudick, 2001).

Our motivation to study the dynamics of pathological tissue evolution, using surrogate MRI signal changes, was reinforced by observations where MRI metrics related to differential measures provided closer physiological ties and better specificity. Contrast-enhancing lesions (Gd-DTPA) have repeatedly been reported to have strong predictive power and clinical relevance (Simon et al., 2000; Weiner et al., 2000). Changes in enhancing lesion number and T2 lesion volume also correlated with clinical activity (attacks) (Weiner et al., 2000), clinical progression (Lee et al., 1998), as well as markers of immunological activity (Khoury et al., 2000). Similarly, a longitudinal study on T1 hypointensities (Bitsch et al., 2001) observed a relative difference between remyelinating and inactive demyelinating lesions, not only in the level of hypointensity but also in their rate of temporal change. Furthermore, significant differences in the rate of brain atrophy were observed in treated MS patients compared to a placebo group (Rudick et al., 1999).
Recent studies have begun to investigate the potential improvement of MRI specificity by looking at difference images (Lee et al., 1998; Tan et al., 2002b), as shown in Fig. 1. It remains to be shown if TSA-derived, lesion-specific time signatures can facilitate the discrimination of different lesion types and pathological processes that are indistinguishable on static images, such as edema, gliosis, astrocytosis, de- and remyelination, and axonal loss. Such longitudinal connections are advocated by the significant correlations found in large studies between baseline inflammatory activity and later T1 hypointense lesion development and also between baseline T1 lesions and follow-up T2 lesions (Simon et al., 2000) and permanent tissue destruction (Bitsch et al., 2001; Filippi et al., 2001).

We have previously pointed out the qualitatively distinct lesion evolution pattern apparent in serial MRI, indicating that specific information about disease activity and the histopathologic evolution of a lesion may be extracted not only from contrast-enhanced but also from nonenhanced serial MR (Guttmann et al., 1995). In a collaborative effort we have also demonstrated that the time-intensity profiles contain sufficient discriminative power to directly segment lesions (Gerig et al., 2000). Together with the findings presented here we conclude that TSA of serial MRI provides a feasible complement to morphometry in longitudinal studies. It may also contribute to the currently unanswered role of MRI in routine follow-up of therapeutic interventions and assessment of treatment efficacy (Arnold et al., 2002).

Limitations and future work

Signal variations, as studied herein, represent changes of tissue quality, location, or both. Similarly, observed distortions (in both geometric and intensity domains) may be of technical nature (e.g., susceptibility) or reflective of biological change (e.g., tissue shifts or atrophy). We sought a self-normalization scheme that dissociates between those sources of variation based on a priori models about their spatiotemporal properties. The adopted intensity normalization scheme does not use an external calibration phantom and therefore does not impact current clinical scanning protocols. In a first-order model, we therefore assumed the global MRI properties of a given tissue class to remain constant over the entire brain. The specificity of TSA to distinguish more global and slowly varying changes caused by diffuse disease from acquisition-related variations within the same model should therefore be studied more closely.

The presented TSA pipeline includes steps of lesion segmentation at each time point. This is beneficial, especially if available in an automated fashion, but with the premise that if 4D data are available, this information may then be used to revise and update the segmentation, leading to a large-scale iterative/intertwined strategy. The presented TSA pipeline does not yet reiterate the segmentation steps, i.e., the segmentation steps guiding the interscan intensity normalization are not constrained for temporal continuity, and segmentation errors propagate to the intensity normalization step.

Applicability of feature maps of temporal characteristics (Gerig et al., 2000; Rey et al., 2001) and direct profile curve parameterization to distinguish lesion types and lesion development will be explored. Recent pilot experiments have attempted to detect such small changes in an fMRI analysis framework (Rey et al., 2001), but no results have been reported yet.

The TSA methods introduced in this article are applicable to image series other than dual-echo scans. Multicontrast TSA, involving simultaneous analysis of multiple MRI channels may also be of interest for future study (Barkhof, 2002).

Conclusions

Methods for automated analysis of brain tissue dynamics, captured in internally normalized MRI intensity profiles, have been presented. The focus of this article is on strategies for correct 4D integration to minimize artifacts due to misregistration, partial-volume effects, and scanner drift. The following main conclusions arise from this work:

1. Fully automated intensity-based registration with sub-voxel precision is feasible in the context of long-term follow-up of white matter disorders. Normalization and pathology masking prior to registration reduced the final signal variance.
2. The use of a partial volume filter addressed the reduction in spatial resolution inherent to longitudinal signal analysis and enhanced robustness and sensitivity to change at the cost of spatial resolution.
3. Individual tissue classes appear to scale differently with changes in overall image brightness, suggesting a tissue-specific interscan intensity normalization scheme.
4. Self-normalization of intensity variations based on tissue-specific intensity histograms suggest that intensity variations above 3% can be reliably detected.
5. Different patterns of lesion dynamics have been observed qualitatively, especially in rate and level of recovery to the initial insult. Connections between intensity and location of this initial activity and subsequent recovery should be investigated.

Acknowledgments

This work was supported in part by funding provided by the National Institutes of Health (R01 NS35142 and P41 RR13218-01). We thank Drs. Howard Weiner and Ferenc Jolesz for providing the longitudinal MRI data. We also thank Drs. Simon Warfield and William Wells III for their segmentation and registration algorithms, respectively.
Appendix

Average rigid transform

To register an image series to a mean pose, the average of multiple spatial transforms was determined by spherical interpolation of quaternion representations. To find a mean pose for spatial normalization, all scans are first registered to a reference scan (e.g., baseline scan \( t_j \)). The quaternion pose to align scan \( j \) with the baseline scan \( k \) is given as \( R_{jk} = [r_{jk}, \alpha_{jk}, t_{jk}] \), given in the form of a rotation axis \( r_{jk} = [x, y, z]_{jk} \) for its spherical CS angles \( \{\theta, \phi\}_{jk} \) and a rotation angle \( \alpha_{jk} \) plus a translation vector \( t_{jk} = [dx, dy, dz]_{jk} \).

Defining the baseline scan as \( k = 1 \), the average of the inverse transforms from \( k = 1 \) to each scan \( j \) is found and scaled to half size. We denote this mean pose as \( R_{10} \) as follows:

\[
R_{10} = \left\{ \frac{1}{N} \sum_j \theta_j, \frac{1}{N} \sum_j \phi_j, \sum_j w_j \alpha_j, \sum_j w_j t_j \right\},
\]

\[
w_j = \frac{\pi}{2} \sum_j \beta_j \frac{\beta_j}{\pi}, \beta_j = \cos^{-1} (r_{1j} \cdot r_{10}).
\]

The interpolation weight for the angles and translations is given by \( w_j \), which scales the rotation angles proportionally to the angle between its axis and the axis of the mean rotation \( r_{10} \). In other words, translations and rotation angles are weighted proportionally to how close they are to the mean transform.

The baseline scan is placed in this new mean pose and the remaining registration matrices are all adjusted by \( \tilde{R}_{1j} = R_{10} \cdot R_{1j} \).

References


