

MRI in Neurosciences

Guest Editors: Alayar Kangarlu, Ramin V. Parsey, and Eric C. Bourekas





MRI in Neurosciences

International Journal of Biomedical Imaging

MRI in Neurosciences

Guest Editors: Alayar Kangarlu, Ramin V. Parsey,
and Eric C. Bourekas



Copyright © 2012 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "International Journal of Biomedical Imaging." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Haim Azhari, Israel

K. Ty Bae, USA

Richard H. Bayford, UK

F. J. Beekman, The Netherlands

J. C. Chen, Taiwan

Anne Clough, USA

Carl Crawford, USA

Daniel Day, Australia

Eric Hoffman, USA

Jiang Hsieh, USA

M. Jiang, China

Marc Kachelriess, Germany

Cornelia Laule, Canada

Seung W. Lee, Republic of Korea

A. K. Louis, Germany

Jayanta Mukherjee, India

Vasilis Ntziachristos, Germany

Scott Pohlman, USA

Erik L. Ritman, USA

Jay Rubinstein, USA

Peter Santago, USA

Lizhi Sun, USA

Kenji Suzuki, USA

Jie Tian, China

Michael W. Vannier, USA

Yue Wang, USA

Ge Wang, USA

Guowei Wei, USA

D. L. Wilson, USA

Sun K. Yoo, Republic of Korea

Habib Zaidi, Switzerland

Yantian Zhang, USA

Jun Zhao, China

Yibin Zheng, USA

Tiange Zhuang, China

Yu Zou, USA

Contents

MRI in Neurosciences, Alayar Kangarlu, Ramin V. Parsey, and Eric C. Bourekas
Volume 2012, Article ID 579192, 2 pages

Tissue-Based MRI Intensity Standardization: Application to Multicentric Datasets, Nicolas Robitaille, Abderazzak Mouiha, Burt Crépeault, Fernando Valdivia, Simon Duchesne, and The Alzheimer's Disease Neuroimaging Initiative
Volume 2012, Article ID 347120, 11 pages

Arterial Spin Labeling (ASL) fMRI: Advantages, Theoretical Constrains and Experimental Challenges in Neurosciences, Ajna Borogovac and Iris Asllani
Volume 2012, Article ID 818456, 13 pages

Metabolite Mapping with Extended Brain Coverage Using a Fast Multisection MRSI Pulse Sequence and a Multichannel Coil, Zhengchao Dong, Feng Liu, Alayar Kangarlu, and Bradley S. Peterson
Volume 2012, Article ID 247161, 8 pages

Label Fusion Strategy Selection, Nicolas Robitaille and Simon Duchesne
Volume 2012, Article ID 431095, 13 pages

Modulating the Focus of Attention for Spoken Words at Encoding Affects Frontoparietal Activation for Incidental Verbal Memory, Thomas A. Christensen, Kyle R. Almryde, Lesley J. Fidler, Julie L. Lockwood, Sharon M. Antonucci, and Elena Plante
Volume 2012, Article ID 579786, 11 pages

MRI Texture Analysis in Multiple Sclerosis, Yunyan Zhang
Volume 2012, Article ID 762804, 7 pages

Landmark Optimization Using Local Curvature for Point-Based Nonlinear Rodent Brain Image Registration, Yutong Liu, Balasrinivasa R. Sajja, Mariano G. Uberti, Howard E. Gendelman, Tammy Kielian, and Michael D. Boska
Volume 2012, Article ID 635207, 8 pages

Editorial

MRI in Neurosciences

Alayar Kangarlu,¹ Ramin V. Parsey,¹ and Eric C. Bourekas²

¹ *Department of Psychiatry, Columbia University, New York, NY 10032, USA*

² *Department of Radiology, The Ohio State University, Columbus, OH 43210, USA*

Correspondence should be addressed to Alayar Kangarlu, ak2334@columbia.edu

Received 24 January 2012; Accepted 24 January 2012

Copyright © 2012 Alayar Kangarlu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Magnetic Resonance (MR) has empowered neuroscience with a tool to investigate the inner structure and workings of the central nervous system (CNS). Unlike many imaging techniques, MR offers multiple modalities in one package, enabling scientists access to a wide range of unknowns from microstructure to physiology of the brain. In this issue, developments in structural and functional MR imaging, spectroscopy, image processing, and applications are presented that could expand MRI's ability in unraveling some of the mysteries of the brain. These works demonstrate that in spite of its rapid growth MRI still has much room for further development and penetration in neuroscience research.

Dong et al. address two of the major challenges in spectroscopy, that are, whole brain coverage and sensitivity. They present a technique for taking advantage of multichannel technology, a powerful signal detection scheme, to increase sensitivity of MR spectroscopy of the brain. Furthermore, Dong et al. offer a multiplanar approach for whole brain coverage, while taking advantage of multichannel radio frequency coils to increase signal-to-noise ratio in order to make whole brain spectroscopy a viable tool for metabolite detection. Dr Zhang, in the second paper, takes up multiple sclerosis (MS), which soon after the invention of MRI, was perceived to benefit the most among all CNS diseases from the high soft-tissue contrast of this technique. In spite of its ability to acquire exquisite images of demyelinating plaques, MRI has not fulfilled its undeclared promise of achieving imaging-based diagnosis in MS. In his review, Dr Zhang draws attention to the higher order sensitivities of MR images, that are, inner voxel signal variation or texture, to highlight the heterogeneous nature of MS. He offers an image analysis technique to pick up variations in plaque intensity

patterns that at the same time can quantify pathological changes. This technique could become a sensitive tool in search for the pathogenesis and in monitoring progression of this disease. Dr N. Robitaille and colleagues have taken up the issue of standardization of images acquired in multicenter studies. To take MRI to the next stage of validation of findings for clinical applications, isolated findings have to be reproduced by geographically-separated teams using different hardware. Such multicenter studies help strengthen the merits of findings and ultimately expand the scope of MRI in its neurological applications. These investigators offer an intensity standardization technique to correct scanner-dependent intensity variations, something that every participant of multicenter studies has suffered from and for which everyone will welcome a solution. Another paper of this issue offers a novel automatic technique, which is both simple and robust and able to use tissue-spatial intensity information to forge a more sensitive measure of brain-tissue assessment. The following paper is also from the Robitaille group and they bring into focus medical image segmentation. Finding the best way to combine different measures of the image labeling they survey the primary labeling techniques and forge a hybrid of them called SVS which is a "label fusion strategy" that they show being superior to any of the methods examined. They have used segmentations in challenging areas of the brain, that are, hippocampus and amygdala, to prove their point and validate their finding. A. Borogovac and I. Asllani chose an emerging MRI technology to highlight in this issue. Arterial spin labeling (ASL) is fast gaining acceptance in its ability to detect cerebral blood flow (CBF) and reliably establish correlation with brain function. ASL has proven to be able to offer meaningful fMRI-based surrogates of brain

function, manifested in terms of hemodynamics. This could serve research on healthy and diseased brain, in both neuroscience and neurology. While on the issue of fMRI, we will read our next contribution from T. Christensen et al. staging the power of this technique in assessing the nature of attention and its role in encoding information into memory. This group shows the power of fMRI in demonstrating how an imaging technique can quantitatively probe attention and what it takes to offer an account of contributions from recollection memory and incidental-perceptual memory. fMRI revealed that BOLD was modulated differentially by unintentionally encoded words compared to novel items. Such use of MRI in providing evidence in support of models of memory consolidation and retrieval is a confirmation of the power of this technique in exploring neuronal processes. Finally, Y. Liu et al. discuss their technique for automation of landmark selection in image registration. Automation of registration of MRI with histology images or EPI with high-resolution anatomical images will allow the images in one modality to benefit from information contained in another. Accurate registration achieved by automatically generated landmarks will help interpretation of images from high-field rodent images, as well as high-distortion images such as EPI acquired from human scanners.

This special issue hoped to highlight “MRI in Neuroscience.” Representative research reports covering the latest advancements in magnetic resonance acquisition or image processing, applications in neuroscience research, and clinical neurological applications are presented in this issue. These and similar works are rapidly expanding the use of MRI in neuroscience and through this encouraging more investigators to use this powerful tool in their research.

*Alayar Kangarlu
Ramin V. Parsey
Eric C. Bourekas*

Research Article

Tissue-Based MRI Intensity Standardization: Application to Multicentric Datasets

Nicolas Robitaille,¹ Abderazzak Mouiha,¹ Burt Crépeault,¹ Fernando Valdivia,¹
Simon Duchesne,^{1,2} and The Alzheimer's Disease Neuroimaging Initiative³

¹Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec, 2601 Chemin de la Canardière,
Québec, QC, Canada G1J 2G3

²Radiology Department, Faculty of Medicine, Laval University, Pavillon Ferdinand-Vandry, 1050 Avenue de la Médecine,
Québec, QC, Canada G1V 0A6

³Alzheimer's Disease Neuroimaging Initiative, 4150 Clement Street, Building 13 (114M), San Francisco, CA 94121, USA

Correspondence should be addressed to Nicolas Robitaille, nicolas.robitaille@crulrg.ulaval.ca

Received 4 May 2011; Accepted 16 November 2011

Academic Editor: Eric C. Bourekas

Copyright © 2012 Nicolas Robitaille et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Intensity standardization in MRI aims at correcting scanner-dependent intensity variations. Existing simple and robust techniques aim at matching the input image histogram onto a standard, while we think that standardization should aim at matching spatially corresponding tissue intensities. In this study, we present a novel automatic technique, called STI for *STandardization of Intensities*, which not only shares the simplicity and robustness of histogram-matching techniques, but also incorporates tissue spatial intensity information. STI uses joint intensity histograms to determine intensity correspondence in each tissue between the input and standard images. We compared STI to an existing histogram-matching technique on two multicentric datasets, Pilot E-ADNI and ADNI, by measuring the intensity error with respect to the standard image after performing nonlinear registration. The Pilot E-ADNI dataset consisted in 3 subjects each scanned in 7 different sites. The ADNI dataset consisted in 795 subjects scanned in more than 50 different sites. STI was superior to the histogram-matching technique, showing significantly better intensity matching for the brain white matter with respect to the standard image.

1. Introduction

Magnetic resonance images (MRIs) acquired with similar protocols but on different scanners will show dissimilar intensity values for the same tissue types [1]. These variations are machine-dependent and do not correspond to noise or bias field inhomogeneity, which both can be reduced with different image processing techniques (e.g., [2, 3], resp.). This problem becomes particularly severe in large, multicentric settings such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), in which longitudinal data is being acquired on more than 50 different platforms in the United States and Canada.

Image processing pipelines aimed at extracting tissue-based characteristics (e.g., grey matter/white matter identification) must be robust to these variations. Intensity standardization is therefore employed to reduce interscanner

differences in order for similar intensities to have similar tissue meaning in the standardized images, regardless of provenance. It has been shown that standardization improves segmentation [4, 5] and registration [6, 7]. However, scaling intensities with a simple linear transformation is not sufficient, since the influence of the MRI acquisition parameters on the image intensities is nonlinear [6]; a higher order transformation is thus needed.

Published standardization techniques generally propose matching image histograms. An algorithm proposed by Wang et al. [8] stretches or compresses a windowed part of the input image histogram with a multiplicative factor, found by minimizing the bin-count difference between the input and standard images histograms. The window is used to include only pixels of interest and remove, for example, the background; this makes the technique linear in the intensity range of interest. The technique developed by Nyúl et al.

[1, 9] matches input image histogram landmarks onto standard histogram landmarks, obtained during a learning process, linearly interpolating intensities between the landmarks. In particular, the variant in [1] uses percentile landmarks, which is simple and more robust. This landmark technique has been used in many studies [5, 7, 10–12]. Jäger et al. [13] extended this principle to two or more jointly used MRI sequences (e.g., T1 and T2), matching multidimensional joint histograms with nonlinear registration. As long as the MRI sequences are spatially aligned, which is assumed, no prior registration of the images is required for computing the joint histogram.

Other techniques use models with some form of *a priori* knowledge, such as the technique proposed by Hellier [6]. It approximates the input image histogram with a mixture of Gaussian functions and aligns their mean with those of the standard image through a polynomial function. Christensen [14] has proposed even-ordered derivatives to find the histogram peak corresponding to the characteristic value of brain white matter; the value is then used to normalize the global image intensity. Weisenfeld and Warfield [4] have proposed modeling the input image as a standardized image corrupted by a linear transformation. Their iterative algorithm then found the parameters of a linear model minimizing the Kullback-Leibler divergence between the standardized and the standard images, thus matching their histograms.

Bergeest and Jäger [15] compared four techniques' performances [1, 4, 6, 13] along with an earlier histogram-matching technique using dynamic histogram warping [16]. None clearly outperformed the others.

Further, in our view, histogram matching should not be the unique objective, as it does not guarantee the standardization of spatially corresponding *tissue* intensities. Towards this end, Leung et al. [17] have recently proposed a semiautomated segmentation technique that delineates cerebrospinal fluid (CSF), white matter (WM) and grey matter (GM) tissue components, for which they computed mean intensities. In a following step, they performed a linear regression between mean intensities and used the results of this regression to perform the standardization. However, this technique yields a linear transformation, which does not completely address the problem as mentioned above.

Thus, to our knowledge, techniques presented so far either matched histograms disregarding spatial correspondence or employed spatial correspondence and linear transformations. Our objective was to design a technique that would (1) use both histogram and tissue-specific intensity information; (2) provide a nonlinear intensity transformation between images; (3) share the simplicity and robustness of the Nyúl's landmark technique [1], while remaining fully automated.

In this study, we report the development of our *STandardization of Intensities* (STI) technique, which fulfills these requirements. We compare STI to the variant L_4 of Nyúl et al. [1], which matches foreground (FRG) intensity histograms using decile (10%) landmarks, in two different multicentric T1-weighted MRI datasets: the Pilot European ADNI (Pilot E-ADNI) study and the larger ADNI dataset.

2. Methodology

2.1. Pilot E-ADNI Dataset. The Pilot E-ADNI dataset was obtained with permission from the multicentric project [18]. Part of this dataset included data from three healthy volunteers, herein referred to as Subjects 1 to 3, scanned within the span of few weeks at seven different European centers (Sites 1 to 7), using the ADNI study 3D T1-weighted MP-RAGE protocol [19]. Information regarding the Pilot E-ADNI study can be found in [18]. Subjects were scanned two times in each center but some data from the first scan were not available. We thus used the data from the second scan only, giving a total of 21 images.

This dataset allowed us to evaluate the performance of standardization techniques by avoiding intersubject intensity variations and focusing only on interscanner differences. Making the reasonable hypothesis that subject tissue properties did not change between sites within the short study timeframe, a well-performing standardization technique should output similar tissue intensities independently of the scanning site.

2.2. ADNI Dataset. The second dataset was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.ucla.edu/>). It consisted in 795 baseline MRIs from control, mild cognitive impairment and probable Alzheimer's disease subjects, acquired on more than 50 different 1.5 T scanners (GE Healthcare, Philips Medical Systems, Siemens Medical Solutions) using the aforementioned protocol [19]. Ethics approval was obtained for each institution involved. MR parameters were standardized as per instructions provided by the ADNI MRI Core [19].

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see <http://www.adni-info.org/>.

2.3. Standard Image. The standard image used throughout this study is the reference image for the BrainWeb simulation software (<http://mouldy.bic.mni.mcgill.ca/brainweb/>) [20]. The standard image is a high resolution (1-mm³ isotropic), high signal-to-noise average of 27 T1-weighted images of a single human brain, with manually delineated CSF, GM, and WM tissue masks.

2.4. Preprocessing. All MRI volumes were preprocessed identically with the MINC image processing toolbox (<http://www.bic.mni.mcgill.ca/ServicesSoftware/MINC>) before standardization:

- (1) nonlocal means noise removal [2];
- (2) intensity inhomogeneity correction [3], performed prior to standardization as suggested in [10];
- (3) global linear registration (12 degrees of freedom) to the standard image [21], maximizing mutual information between the two volumes [22];
- (4) resampling to a 1-mm³ isotropic grid;
- (5) intensity clamping, which consisted in (a) setting to zero all intensity values below the percentile value 0.01, (b) setting to 100 all intensity values above the percentile value 99.99, and (c) linearly interpolating intensities between those limits. This step removed outliers of low and high intensities and rescaled image intensity between 0 and 100;
- (6) global nonlinear registration to the standard image [23], maximizing mutual information between the two volumes as in step 3.

Left of Figure 1 (red) summarizes the above preprocessing steps.

2.5. Intensity Standardization

2.5.1. STI. In the last preprocessing step, global *nonlinear* registration established spatial correspondence between tissues in the standard and input images. This spatial correspondence allowed us to compute a joint intensity histogram of the frequency distribution of intensity correspondences. From the most frequent tissue-specific correspondences, our aim was to compute an intensity transfer function mapping the *nonlinearly registered* input image (preprocessed with steps 1 to 6) onto the standard, which was then applied to the *linearly registered* input image (preprocessed with steps 1 to 5), as shown in Figure 1 (green), giving us the desired standardized image, in the standard reference space for further processing and/or comparison.

Since tissue intensities generally overlap, it was difficult to estimate tissue-specific correspondences from the global joint histogram. To refine its estimates, STI used the standard image tissue masks for (1) background (BKG), (2) WM and (3) GM, from the standard image. We chose to treat the background for two reasons. First, each image can be taken as a whole, with no background removal. Second, intensity corresponding to CSF is often treated in our downstream

processing pipelines. Since it is mostly similar to BKG, we found that it was more robust to correct it through BKG standardization.

For each tissue, STI performed the following steps:

- (1) kept only voxels contained in the applicable tissue mask for both nonlinearly registered input and standard images;
- (2) computed and smoothed, with a Gaussian low-pass filter, the standard-versus-input joint intensity histogram for the masked voxels. The width of the histogram bins was 0.25% in each dimension, that is, 400 bins covering the 0 to 100 intensity scale, and the full width at half maximum of the Gaussian filter was set to 10 bins;
- (3) found the mode, that is, maximum, in the joint histogram. The mode corresponds to the most frequent intensity correspondence between the nonlinearly registered input and standard images for the current tissue. This point determined a histogram landmark pair corresponding to the input-to-standard intensity mapping for the current tissue.

To the set of landmark pairs obtained with the tissue masks, STI added two extra pairs: the first, (0, 0), mapped both minimum intensities in the nonlinearly registered input and standard images, and the second, (100, 100), their maximum values. The resulting landmark set S_{STI} is then given by

$$S_{STI} = \{(0, 0)(m_{r,BKG}, m_{s,BKG})(m_{r,GM}, m_{s,GM})(m_{r,WM}, m_{s,WM})(100, 100)\}, \quad (1)$$

where $m_{r,X}$ and $m_{s,X}$ represent the intensity of the nonlinearly registered input and standard images, respectively, for tissue X .

STI completed the mapping function by linearly interpolating intensities between the landmark pairs (*piecewise* linear transformation) and finally applied this function to the *linearly* registered input image (preprocessed with steps 1 to 5) in order to create the standardized image.

We added an experimentally determined heuristic to this algorithm. Given large overlaps between tissue classes in some cases, we ordered the search from largest to smallest tissue component, reducing the voxel search space once a component mode was estimated. Practically, this resulted in estimating BKG first, as it generally had the largest mode. Once found, all voxels in an intensity range up to 10% above that mode were removed before estimating the WM mode. After the WM mode had been found, voxels in an intensity range 25% below as well as all voxels above this mode were removed before estimating the GM mode, thereby removing overlap between BKG/GM and GM/WM.

2.5.2. L_4 . We compared STI to the following implementation of the histogram-matching technique described in [1] as L_4 , which uses decile, that is, 10%, landmarks to match the histograms of nonlinearly registered input and standard images foreground (FRG). FRG is determined in each image via intensity thresholding. It corresponds to the set of voxels

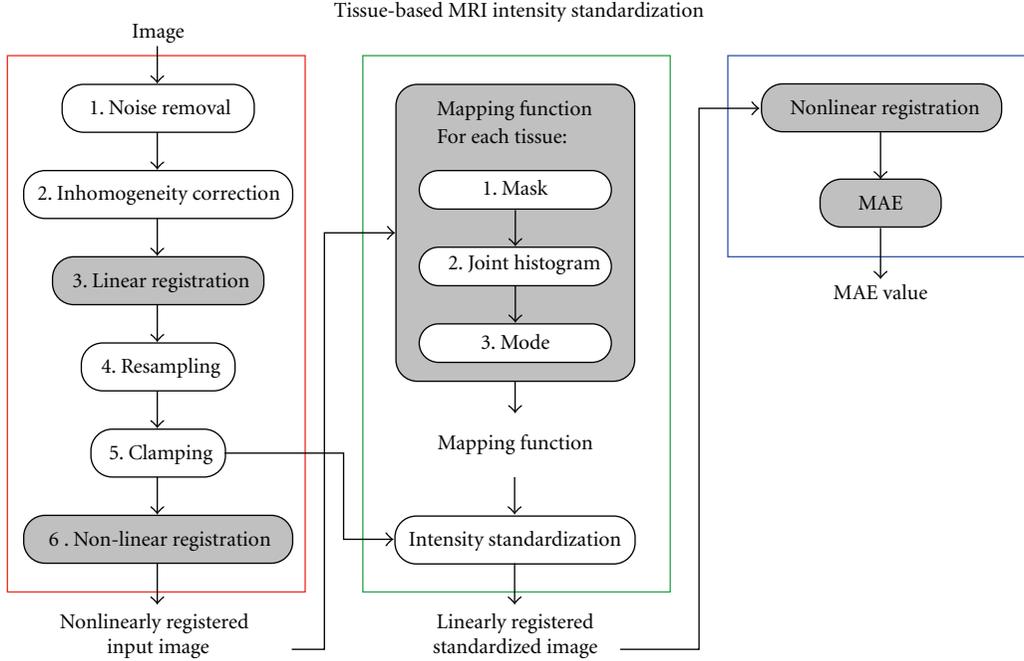


FIGURE 1: Flowchart showing the steps involved in image preprocessing (red), intensity standardization (green) and mean absolute error (MAE) computation for technique comparison (blue). The steps for which the standard image or its masks are required are shaded in grey. For the *mapping function* process, we summarize the three steps involved in STI.

for which the intensity is (1) higher than or equal to the mean intensity computed over the whole image and (2) lower than the intensity corresponding to the percentile value 99.8 obtained for the whole image. This operation thus crops the histogram of each image by removing, as assumed, the background and the intensity outliers, respectively. While the clamping step in the preprocessing phase already deals with outliers by clamping the image intensities between percentile values 0.01 and 99.99 to produce the input image, we chose to keep the additional less-tolerant limit of 99.8 to conform with [1].

For both the nonlinearly registered input and standard images, the intensity values corresponding to the percentile value 99.8 were first found, creating a first landmark pair. Decile landmarks, corresponding to the percentile values $\{10, 20, \dots, 90\}$, were then determined, within FRG only. This operation yielded nine more landmark pairs.

Two extra pairs, $(0, 0)$ and $(100, 100)$, were finally added to map, respectively, the minimum and the maximum values of the nonlinearly registered input and standard images. We thus obtained the following set of landmark pairs S_{L4} :

$$S_{L4} = \left\{ (0, 0) \left(m_{r,fg,10}, m_{s,fg,10} \right) \left(m_{r,fg,20}, m_{s,fg,20} \right) \right. \\ \left. \dots \left(m_{r,fg,90}, m_{s,fg,90} \right) \left(m_{r,99.8}, m_{s,99.8} \right) (100, 100) \right\}, \quad (2)$$

where $m_{r,fg,P}$ and $m_{s,fg,P}$ represent the intensities at percentile P for the nonlinearly registered input and standard image FRG, and $m_{r,99.8}$ and $m_{s,99.8}$ are intensity values at 99.8

percentile for the whole images. The mapping function is obtained by interpolating linearly between landmark pairs.

2.6. Technique Comparison. As mentioned, the intensity standardization mapping functions were *determined* using the preprocessed nonlinearly registered images (preprocessing steps 1 to 6). However, the obtained mapping functions were *applied* to the linearly registered images, preprocessed with steps 1 to 5, that is, prior to nonlinear registration.

To compare standardization techniques, we applied, as shown in Figure 1 (blue), a nonlinear registration to the latter images to match the standard image, using the same technique as in preprocessing step 6. We then used the standard image tissue masks to compute the voxelwise mean absolute error (MAE) on different voxel sets: (1) the standard image FRG, as defined in the L_4 procedure above, (2) WM and (3) GM. Applying standardization to the images prior to nonlinear registration allowed us, in this comparison scheme, to reduce any bias toward STI associated with the use of the standard masks.

The MAE was given by

$$MAE = \frac{1}{N} \sum_{v=1}^N |I_{o,v} - I_{s,v}|, \quad (3)$$

where N is the number of voxels, and $I_{o,v}$ and $I_{s,v}$ are intensity values for the nonlinearly registered standardized output and standard images, respectively, at voxel v . We note that since output and standard images intensity scales range from 0 to 100, MAE can be expressed in percentage.

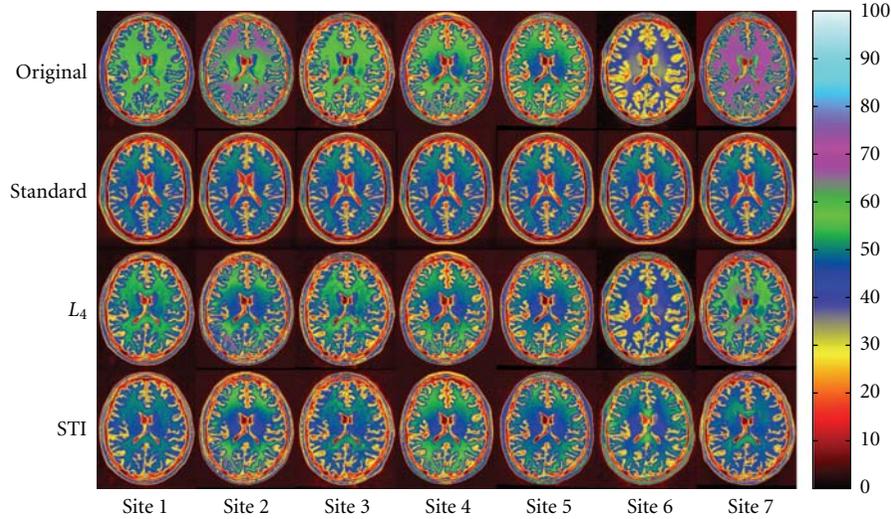


FIGURE 2: Standardized images for Subject 1 in the Pilot E-ADNI dataset. From top to bottom: linearly registered input images, standard image, and images standardized with L_4 and STI, respectively. From left to right: images for Sites 1 to 7. Color coding enhances intensity differences.

We performed t -tests to evaluate the statistical significance of MAE differences between L_4 and STI.

3. Results

3.1. Pilot E-ADNI Dataset. Figure 2 shows standardized linearly registered output images for Subject 1 in the Pilot E-ADNI dataset scanned at 7 different sites. We expected standardized intensities to be similar. Qualitatively, STI exhibited a better performance than L_4 , especially for WM, and in particular for Sites 6 and 7. We obtained similar results (not shown) for Subjects 2 and 3.

Figure 3 presents an example of joint intensity histograms computed before (a, c) and after (b, d) standardization. The grayscale images correspond to the natural logarithm of the joint histograms obtained for Subject 1 at Site 1, whose images are presented in Figure 2. In (a) and (c), we overlaid the intensity mapping functions obtained with L_4 and STI. We observe that STI gives a better intensity correspondence after standardization (d) than L_4 (b).

Figure 4 shows MAE boxplots of FRG, WM, and GM over the 21 images. We see that, compared to no standardization (Original), L_4 and STI clearly showed an improvement in terms of MAE for all voxel sets. However, STI offered better performance; the statistical test effectively showed that the difference between L_4 and STI was significant for WM ($P = 0.0075$), almost significant for GM ($P = 0.0674$), but not for FRG ($P = 0.2459$).

3.2. ADNI Dataset. Figure 5 shows MAE boxplots for the 795 different subjects for (a) FRG, (b) WM, and (c) GM. Compared to no standardization (Original), both L_4 and STI exhibited better MAE. STI significantly outperformed L_4 for

WM ($P < 0.0001$), with no difference for GM ($P = 0.3120$). However, L_4 was superior for FRG ($P = 0.0239$).

Figures 6 and 7 present ADNI image examples sorted according to MAE percentiles 100 (A), 90 (B), 75 (C), 50 (D), 25 (E), 10 (F), and 0 (G) for the FRG voxel set. Images (A) and (G) thus give the highest (worst) and lowest (best) MAE, respectively, for L_4 (Figure 6) and STI (Figure 7). FRG was selected, instead of WM or GM, to avoid any bias toward STI. In fact, selecting FRG would normally favor L_4 .

Qualitatively, although FRG MAE decreases from (A) to (G), a corresponding improvement in WM is not necessarily observed. This is also shown in Table 1, where FRG, WM and GM MAE values are given for each image of Figures 6 and 7. MAE values for GM do not necessarily follow the trend for FRG either.

We also see that L_4 and STI can both result in higher (worse) MAE than with no standardization (see Figures 7(A) and 7(C) for WM). In other words, the WM intensity of the nonstandardized image, in these cases, would be closer to the standard than the WM intensity given by L_4 and STI, according to MAE. Over the 795 images of the ADNI dataset, the percentages of images for which L_4 and STI gave higher MAE than without standardization for WM (worst case for both methods) were, respectively, 1.38% and 4.65%. The higher percentage obtained with STI is explained by multiple peaks or wider distributions in the joint intensity histograms of images similar to Figure 7(A). As mentioned, STI selects only the maximum peak for a given tissue. This point is further discussed below.

Finally, for the images presented in Figures 6 and 7, Table 1 reveals that STI gave the lowest MAE values in 26 cases (FRG: 7, WM: 10, GM: 9) versus 16 for L_4 (FRG: 7, WM: 4, GM: 5), even if selecting FRG as the sorting voxel set would have normally favored L_4 . It must be noted that this sample is not representative of the whole ADNI dataset,

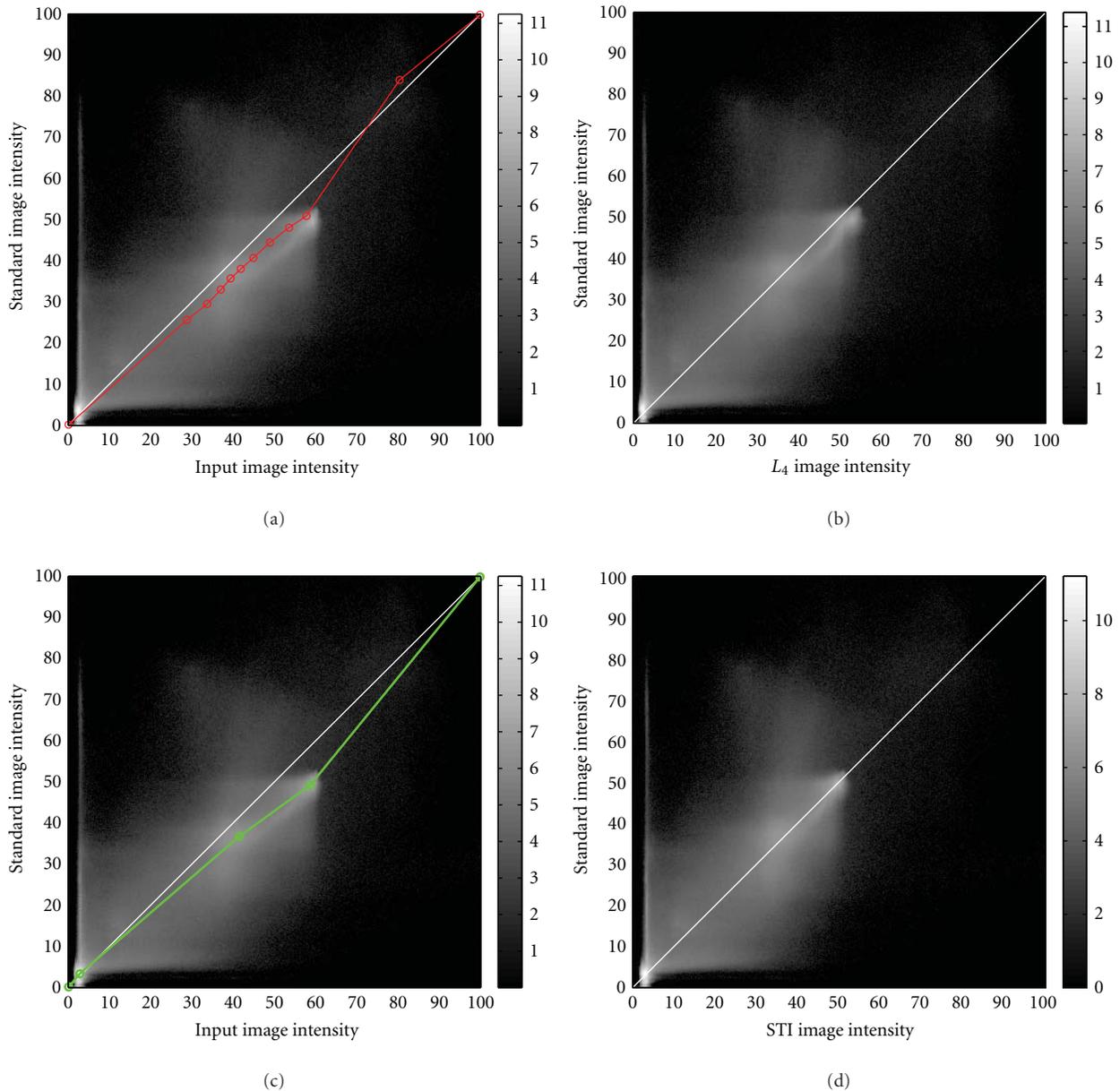


FIGURE 3: Natural logarithm of the joint intensity histograms obtained for Subject 1 at Site 1, whose images are presented in Figure 2, (a, c) before and (b, d) after standardization with (a, b) L_4 and (c, d) STI. In (a, c), we overlaid the intensity mapping functions obtained with L_4 (a) and STI (c). The histogram diagonals (white) represent perfect intensity correspondence.

as we artificially selected images to display at each MAE percentiles for each standardization technique. Yet, it is in accordance with boxplots shown in Figure 5 and statistical results detailed earlier.

4. Discussion

4.1. Methodological Considerations. STI uses spatial correspondence and joint intensity histograms between the input and standard images to find modes and use them as landmarks in the intensity mapping function. While the use of

joint histograms has been reported in [13], the authors computed joint histograms between different imaging modalities separately, rather than for the input and standard images. As demonstrated in this study, using such spatial correspondence improves the standardization quality in terms of MAE. This improvement in MAE can impact the final outcome of studies by reducing systematic errors, which in turn can reduce the number of subjects required to achieve a similar level of statistical significance or power.

In this study, the Pilot E-ADNI dataset allowed us to avoid intersubject intensity variations. Effectively, we should theoretically expect that, for a given subject, a similar image

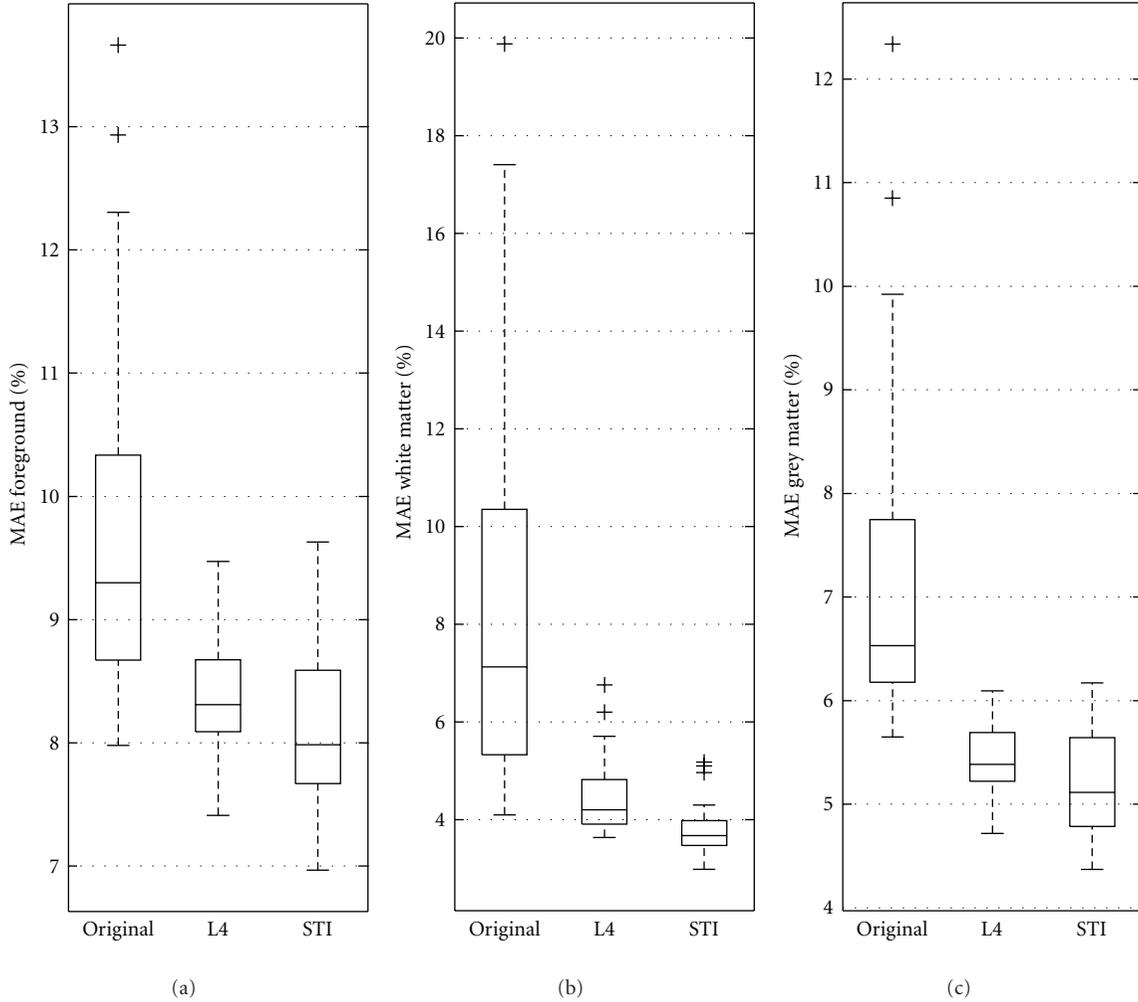


FIGURE 4: Boxplots of (a) FRG, (b) WM, and (c) GM mean absolute errors for the 21 images from Pilot E-ADNI. (Original) Nonlinearly registered input images, (L_4) nonlinearly registered output images standardized with L_4 , and (STI) nonlinearly registered output images standardized with STI.

be produced for all sites. Compared to L_4 , STI was nearer to this expectation, particularly for Sites 6 and 7. Effectively, we showed that STI was significantly superior to L_4 in WM, while differences were not significant in the other voxel sets.

For the ADNI dataset, STI again showed to be significantly better in WM than L_4 , while in FRG, L_4 was significantly superior. For FRG, however, we showed that better results did not necessarily correspond to better intensity correspondences for WM and/or GM. This suggests we should not rely on the results obtained in FRG, as long as we are mainly interested in brain GM and WM.

4.2. Limitations. We tested STI following linear registration only (results not presented); nonlinear registration yielded better performance. This reliance on registration however remains the main limitation of the technique.

STI is also designed to find one maximum, the mode, in the joint intensity histograms. In cases where two or more peaks are present or the intensity distribution is wider,

due to for example insufficient inhomogeneity correction, this might lead to discrepancies such as in Figure 7(A), for which we observe high intensity WM. Although a better preprocessing may solve some of these discrepancies, we plan to add further landmarks in the joint intensity histograms and thus produce better mapping functions in future versions. However, care must be taken to avoid “overstandardizing” intensity variations, especially when dealing with pathologies, for example, severe white matter diseases. We will have to further validate the influence of these pathologies on STI. However, due to the nonlinear registration step, STI should not be sensitive to volume changes, for example, associated to Alzheimer’s disease, as suggested by the results obtained with the ADNI dataset.

Another limitation is that STI has been developed for brain T1-weighted MRI. However, it could be easily applied to other sequences (e.g., T2-weighted images), provided a standard image for this acquisition and corresponding tissue masks.

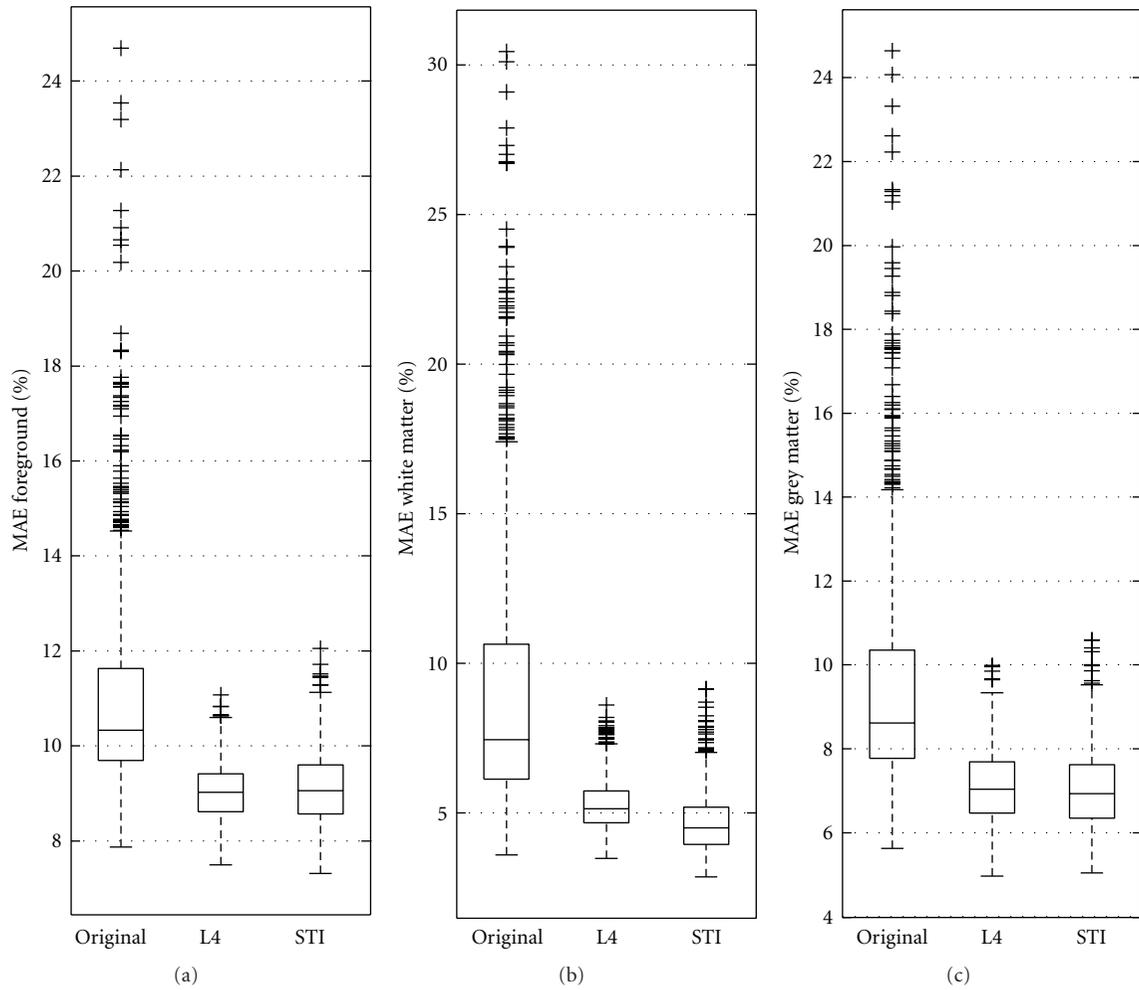


FIGURE 5: Boxplots of MAE obtained over the 795 images of the ADNI dataset for (a) FRG, (b) WM, and (c) GM. (Original) Nonlinearly registered input images, (L4) nonlinearly registered output images standardized with L_4 , (STI) nonlinearly registered output images standardized with STI.

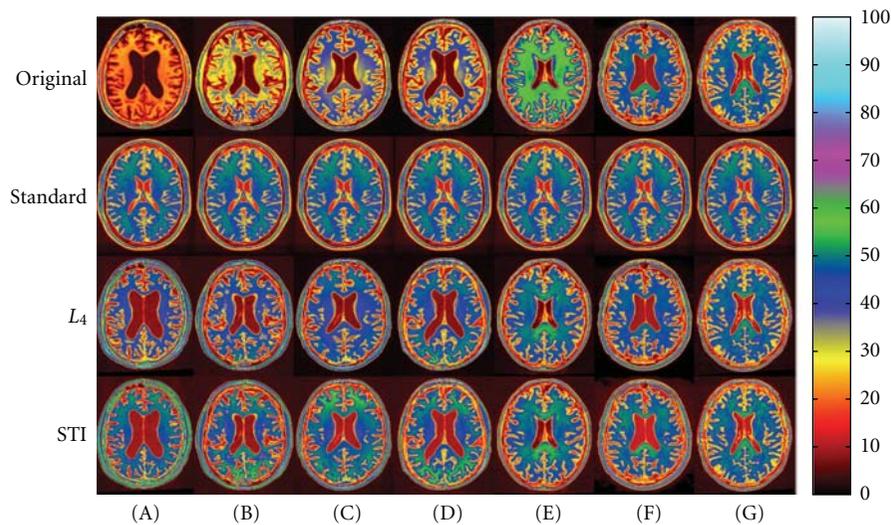


FIGURE 6: Standardized images of the ADNI dataset, sorted according to MAE percentiles (A) 100, (B) 90, (C) 75, (D) 50, (E) 25, (F) 10, and (G) 0 obtained for the FRG voxel set with L_4 . Images (A) and (G) correspond, respectively, to the highest (worst) and lowest (best) MAE obtained in FRG for L_4 . From top to bottom: linearly registered input images, standard image, and images standardized with L_4 and STI, respectively. MAE values are given in Table 1.

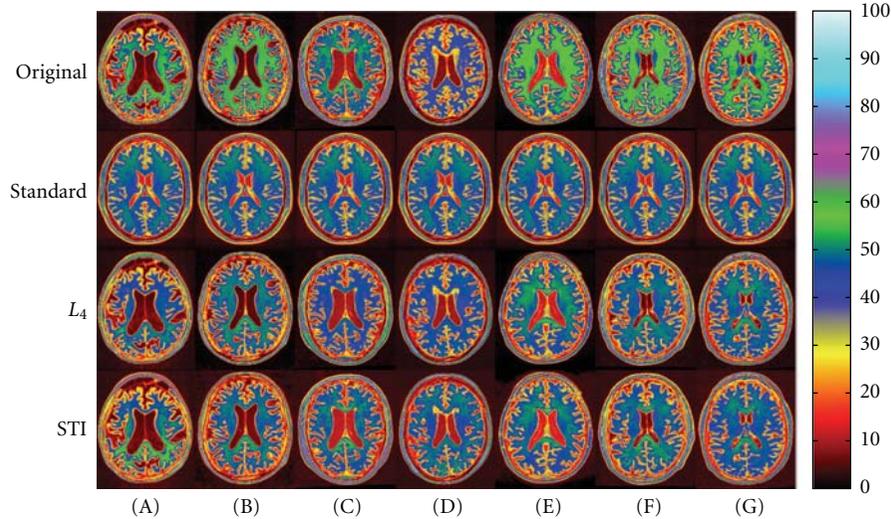


FIGURE 7: Standardized images of the ADNI dataset, sorted according to MAE percentiles (A) 100, (B) 90, (C) 75, (D) 50, (E) 25, (F) 10, and (G) 0 obtained for the FRG voxel set with STI. Images (A) and (G) correspond, respectively, to the highest (worst) and lowest (best) MAE obtained in FRG for STI. From top to bottom: linearly registered input images, standard image, and images standardized with L_4 and STI, respectively. MAE values are given in Table 1.

TABLE 1: MAE (%) of the ADNI images presented in Figures 6 and 7, obtained for FRG, WM, and GM.

Figure	Voxel set	Technique	(A)	(B)	(C)	(D)	(E)	(F)	(G)	
Figure 6	FRG	Original	21.27	13.93	11.94	12.14	9.90	8.85	7.87	
		L_4	11.07	9.84	9.41	9.02	8.62	8.30	7.50	
		STI	9.93	9.71	9.24	8.81	8.69	8.48	7.56	
	WM	Original	27.89	16.65	12.90	11.28	6.88	5.08	4.04	
		L_4	8.60	7.05	6.32	4.59	4.19	5.04	3.67	
		STI	4.42	4.86	4.03	4.04	4.40	4.34	3.64	
	GM	Original	22.61	14.31	11.24	10.77	7.58	7.20	5.62	
		L_4	9.99	8.58	7.52	6.65	6.04	6.70	4.97	
		STI	8.75	8.05	7.03	6.50	6.43	6.50	5.31	
	Figure 7	FRG	Original	12.14	11.77	10.10	10.90	10.13	10.29	8.93
			L_4	10.64	9.70	8.88	8.92	8.77	8.24	7.67
			STI	12.05	10.11	9.60	9.05	8.57	8.23	7.32
WM		Original	8.04	9.38	5.34	10.01	7.68	8.30	6.43	
		L_4	6.49	5.26	6.32	5.70	4.95	4.28	4.02	
		STI	8.24	5.94	5.00	4.11	4.89	4.32	3.69	
GM		Original	10.72	10.00	8.30	9.21	8.59	8.19	7.37	
		L_4	8.99	7.63	7.71	6.83	6.90	6.23	5.89	
		STI	10.59	8.12	7.69	6.70	6.79	6.29	5.49	

Best (lowest) MAE values are highlighted in bold characters.

5. Conclusion

We presented a new tissue-based standardization technique called STI. This technique uses spatial correspondence between an input image and a standard determined via global linear and nonlinear registration. Registration allows the use of joint histograms to determine intensity correspondence in each tissue, defined within voxel masks.

We compared STI to an existing histogram-matching technique and showed that STI was superior in terms of mean absolute error, particularly in the white matter, in two multicentric datasets. These results demonstrated that standardization techniques should not be aimed solely at matching histograms and that spatial information should also be incorporated. To our knowledge, it is the largest study on intensity standardization.

Abbreviations

2D:	Two-dimensional
ADNI:	Alzheimer's Disease Neuroimaging Initiative
BKG:	Background
CSF:	Cerebrospinal fluid
E-ADNI:	European ADNI
FRG:	Foreground
GM:	Grey matter
MAE:	Mean absolute error
MRI:	Magnetic resonance imaging
STI:	<i>STandardization of Intensities</i>
WM:	White matter.

Acknowledgments

The authors thank the Alzheimer's Disease Neuroimaging Initiative and the Pilot European ADNI (G.B. Frisoni) for access to MRI data. This work was supported by the Ministère du Développement Économique, de l'Innovation et de l'Exportation du Québec. The Pilot European ADNI study was funded thanks to an unrestricted grant by the Alzheimer's Association. Principal Investigator is Giovanni B. Frisoni (Italy), PIs of the Clinical program Bruno Vellas (France), of the MR Imaging program Frederik Barkof (The Netherlands), and of the Biological Marker program Harald Hampel (Ireland/Germany) and Kaj Blennow (Sweden). For the ADNI dataset, data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). Data were obtained from the ADNI database (<http://adni.loni.ucla.edu/>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorphix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (<http://www.fnih.org/>). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by

the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

References

- [1] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143–150, 2000.
- [2] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 425–441, 2008.
- [3] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *IEEE Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, 1998.
- [4] N. I. Weisenfeld and S. K. Warfield, "Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence," in *2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano*, pp. 101–104, April 2004.
- [5] Y. Zhuge and J. K. Udupa, "Intensity standardization simplifies brain MR image segmentation," *Computer Vision and Image Understanding*, vol. 113, no. 10, pp. 1095–1103, 2009.
- [6] P. Hellier, "Consistent intensity correction of MR images," in *Proceedings of the International Conference on Image Processing (ICIP '03)*, pp. 1109–1112, September 2003.
- [7] U. Bağcı, J. K. Udupa, and L. Bai, "The role of intensity standardization in medical image registration," *Pattern Recognition Letters*, vol. 31, no. 4, pp. 315–323, 2010.
- [8] L. Wang, H. M. Lai, G. J. Barker, D. H. Miller, and P. S. Tofts, "Correction for variations in MRI scanner sensitivity in brain studies with histogram matching," *Magnetic Resonance in Medicine*, vol. 39, no. 2, pp. 322–327, 1998.
- [9] L. G. Nyúl and J. K. Udupa, "On standardizing the MR image intensity scale," *Magnetic Resonance in Medicine*, vol. 42, no. 6, pp. 1072–1081, 1999.
- [10] A. Madabhushi and J. K. Udupa, "Interplay between intensity standardization and inhomogeneity correction in MR image processing," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 561–576, 2005.
- [11] Y. Zhuge, J. K. Udupa, J. Liu, and P. K. Saha, "Image background inhomogeneity correction in MRI via intensity standardization," *Computerized Medical Imaging and Graphics*, vol. 33, no. 1, pp. 7–16, 2009.
- [12] A. Madabhushi, J. K. Udupa, and G. Moonis, "Comparing MR image intensity standardization against tissue characterizability of magnetization transfer ratio imaging," *Journal of Magnetic Resonance Imaging*, vol. 24, no. 3, pp. 667–675, 2006.
- [13] F. Jäger, L. Nyúl, B. Frericks, F. Wacker, and J. Hornegger, "Whole body MRI intensity standardization," in *Bildverarbeitung für die Medizin*, A. Horsch, T. M. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, Eds., pp. 459–463, Springer, Berlin, Germany, 2007.
- [14] J. D. Christensen, "Normalization of brain magnetic resonance images using histogram even-order derivative analysis," *Magnetic Resonance Imaging*, vol. 21, no. 7, pp. 817–820, 2003.
- [15] J.-P. Bergeest and F. Jäger, "A comparison of five methods for signal intensity standardization in MRI," in *Bildverarbeitung*

- für die Medizin*, T. Tolxdorff, J. Braun, T. M. Deserno, A. Horsch, H. Handels, and H.-P. Meinzer, Eds., pp. 36–40, Springer, Berlin, Germany, 2008.
- [16] I. J. Cox, S. Roy, and S. L. Hingorani, “Dynamic histogram warping of image pairs for constant image brightness,” in *Proceedings of IEEE International Conference on Image Processing*, pp. 366–369, October 1995.
- [17] K. K. Leung, M. J. Clarkson, J. W. Bartlett et al., “Robust atrophy rate measurement in Alzheimer’s disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection,” *NeuroImage*, vol. 50, no. 2, pp. 516–523, 2010.
- [18] G. B. Frisoni, W. J. P. Henneman, M. W. Weiner et al., “The pilot European Alzheimer’s Disease Neuroimaging Initiative of the European Alzheimer’s Disease Consortium,” *Alzheimer’s and Dementia*, vol. 4, no. 4, pp. 255–264, 2008.
- [19] C. R. Jack Jr., M. A. Bernstein, N. C. Fox et al., “The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods,” *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.
- [20] B. Aubert-Broche, A. C. Evans, and L. Collins, “A new improved version of the realistic digital brain phantom,” *NeuroImage*, vol. 32, no. 1, pp. 138–145, 2006.
- [21] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, “Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space,” *Journal of Computer Assisted Tomography*, vol. 18, no. 2, pp. 192–205, 1994.
- [22] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
- [23] D. L. Collins and A. C. Evans, “Animal: validation and application of nonlinear registration-based segmentation,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 8, pp. 1271–1294, 1997.

Review Article

Arterial Spin Labeling (ASL) fMRI: Advantages, Theoretical Constrains and Experimental Challenges in Neurosciences

Ajna Borogovac and Iris Asllani

Department of Radiology, Columbia University, New York, NY 10032, USA

Correspondence should be addressed to Iris Asllani, ia2026@columbia.edu

Received 18 April 2011; Revised 27 September 2011; Accepted 11 October 2011

Academic Editor: Alayar Kangarlu

Copyright © 2012 A. Borogovac and I. Asllani. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cerebral blood flow (CBF) is a well-established correlate of brain function and therefore an essential parameter for studying the brain at both normal and diseased states. Arterial spin labeling (ASL) is a noninvasive fMRI technique that uses arterial water as an endogenous tracer to measure CBF. ASL provides reliable absolute quantification of CBF with higher spatial and temporal resolution than other techniques. And yet, the routine application of ASL has been somewhat limited. In this review, we start by highlighting theoretical complexities and technical challenges of ASL fMRI for basic and clinical research. While underscoring the main advantages of ASL versus other techniques such as BOLD, we also expound on inherent challenges and confounds in ASL perfusion imaging. In closing, we expound on several exciting developments in the field that we believe will make ASL reach its full potential in neuroscience research.

1. Introduction

Until we find a method that can directly and noninvasively measure the production and consumption of ATP, we must rely on measuring physiological correlates of ATP to study the brain at various functional states such as activation and disease. Cerebral blood flow (CBF) is one such correlate. CBF delivers glucose and oxygen to the brain to maintain basal ATP production and to replenish it during increased neuronal activity. Changes in CBF are concomitant with changes in neuronal activity, such as those occurring during task activation, or changes in metabolism that often indicate presence of disease [1]. Because of this close coupling with brain function, CBF is an essential physiological parameter, which is why much effort has been dedicated to developing reliable methods for measuring it.

All the major methods that have been developed for measuring CBF are based on the principles of compartmental modeling and tracer kinetics. These principles yield models that describe the dynamics of a tracer as it crosses the arterial tree into the brain's microvasculature (nondiffusible tracers) and into the tissue (diffusible tracers) prior to venous washout. Different methods use different types of tracers.

One of the main advantages of ASL fMRI is that, unlike most other methods, it uses arterial water as an endogenous tracer and thus does not require injection of exogenous tracers that can be uncomfortable and potentially harmful. Also, because ASL is noninvasive, it is safe to repeat over time and can therefore be used to track changes in CBF such as those due to disease progression or drug therapy. Important, ASL yields an absolute measurement of CBF and therefore any change in flow can be expressed in physiologically meaningful units rather than as a % change. Furthermore, ASL yields CBF images with higher spatial and temporal resolution than any other current technique.

And yet, despite these numerous advantages, ASL has yet to emerge as the technique of choice for measuring CBF at baseline or during task activation. The main motivation for this review was to postulate on the main reasons for this relatively slow-paced advancement of ASL. To this aim, we first start with a description of the general theoretical framework of ASL including various labeling implementations that have been developed to date. Then we highlight the key advantages of ASL versus other methods while expounding on its theoretical limitations and experimental challenges. We close with a review of recent developments in the field

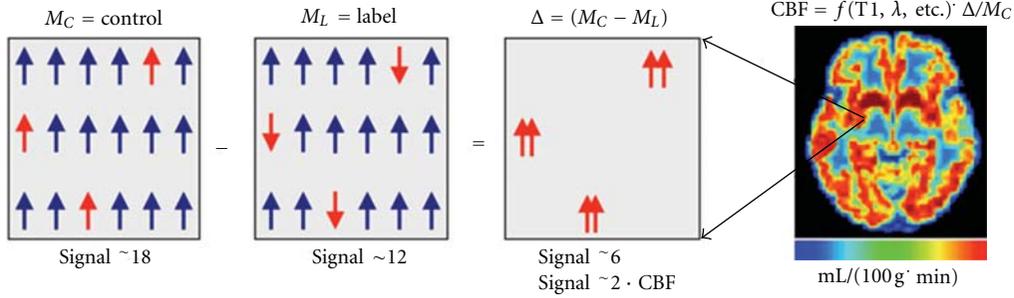


FIGURE 1: Schematic presentation of how ASL signal is obtained. The first three panels represent the signal from a single imaged voxel that originates from the control (left), label (center), and control-label difference (right) panels, respectively. The numbers are not meant to represent real flow. A real CBF image is shown in the rightmost panel. The color bar represents flow in $[0-107]$ mL/100 g · min range. Note that, as mentioned in text, the difference $\Delta = (M_C - M_L)$ image is converted to a single CBF image via a function that includes physiological and MR parameters such as relaxation rates, transit times, and blood tissue water partition coefficient, λ .

that have the promise of making ASL realize its full potential in brain research.

2. Theoretical Framework of ASL fMRI

In all ASL methods, the proton spins of the arterial water are labeled prior to reaching the imaged volume. “Labeling” refers to a change in the magnetic state of the inflowing spins by either saturation or inversion. Once the spins have been labeled, and after a time delay that allows for them to exchange with the tissue, an image, referred to as “labeled,” is acquired [2]. In this image, the blood water is in a different magnetization state from that of the static tissue water. If one models the static tissue as an upward, +1, vector then the labeled water is either 0 (saturation) or -1 (inversion). The signal from a given voxel in the labeled image represents a sum over both blood and tissue spins (Figure 1, left panel).

In addition to the “labeled” image, a “control” image is acquired in which, ideally, the arterial water spins have not been altered, and therefore both the static tissue spins and blood spins are in the same magnetic state at acquisition (Figure 1, middle panel). Now one can see (Figure 1, right panel) that at each imaged voxel, the difference between the control and labeled images is proportional to the amount of flow supplying that voxel.

The ASL signal is typically expressed as a fractional ratio (1) between the difference, $(M_C - M_L)$, image and the control, M_C , image, which represents the equilibrium magnetization [4]

$$\text{ASL}_{\text{signal}} = \frac{(M_C - M_L)}{M_C}. \quad (1)$$

A CBF image is computed by applying a set of measured or assumed physiological and MR parameters on the ASL-signal image to obtain voxelwise flow values in absolute physiological units of flow (Figure 1). Multiple pairs of labeled and control images are acquired to ensure that an average of several cardiac outputs has been obtained.

Since the original innovation of the basic ASL technique by Williams et al. in 1992 [4], numerous ASL sequences have

been developed. Based on how labeling is achieved, these sequences are commonly categorized as either continuous ASL (CASL), where labeling duration is relatively long (seconds), or pulsed ASL (PASL), where short (milliseconds) labeling pulses are used [5]. Recently, ASL sequences have been modified to include background suppression pulses that aim at suppressing the static tissue signal in order to improve the stability and SNR of the ASL [6, 7]. Thus, several permutations exist even within the same major ASL category. Moreover, recent developments such as pseudo-continuous ASL (pCASL) [8] and velocity selective ASL (VSASL) [9] have made this dichotomization into CASL and PASL somewhat obsolete. While recognizing the limitations of categorizations in general, in this review we highlight the basic principles as well as advantages and limitations of CASL, pCASL, PASL, and VSASL, separately.

2.1. Continuous Arterial Spin Labeling (CASL). In CASL, the inflowing arterial water spins are “continuously” labeled via inversion through a process called adiabatic fast passage (AFP) [2, 4, 10, 11]. For adiabatic inversion to occur, two conditions must be met: first, the entire labeling process needs to be faster than the relaxation times (hence the term “fast” in AFP), and, second, the orientation of the effective magnetic field, B_{eff} , needs to change at sufficiently slow rate so that the angle between B_{eff} and the net magnetization remains constant (hence the term “adiabatic”). Theoretically, the labeling pulse must be long enough for the steady state to be reached [4]. However, taking into consideration hardware and experimental restrictions, the labeling pulse is typically ~ 2 s. The inversion occurs at a thin plane referred to as the “labeling plane” generally positioned in the carotids where the average flow velocity ensures that both adiabatic conditions are met.

One of the major drawbacks of CASL is the requirement for a long labeling pulse to bring about the adiabatic inversion. This requirement complicates CASL on both theoretical and practical standpoints. Theoretically, long off-resonance RF pulses cause signal loss due to what are commonly referred to as magnetization transfer (MT) effects [12]. Because the control images do not require labeling pulses,

their signal would not be affected by MT. Consequently, the difference ($M_C - M_L$) would reflect not only the blood flow but also the loss in signal due to MT effects that are present in M_L but not M_C . The MT effects constrained the initial application of CASL to a single slice where in order to balance these effects, two labeled images were acquired, one above and one below the imaged slice [13]. To circumvent this constraint, Alsop and Detre applied an amplitude-modulated (AM) RF pulse with the same duration as the labeling pulse during the acquisition of the control image [14]. While causing approximately the same MT effects during the acquisition of the control images as during the labeled, the AM pulses affect a double inversion, that is, no net inversion, on the inflowing spins thus meeting the requirement of no change in magnetization state during the acquisition of the control image [14]. The introduction of the AM pulses moved CASL from single slice to whole-brain imaging thus opening new venues for its applications.

Practically, the requirement for long RF pulses is so taxing on the hardware that most commercially available MR scanners do not offer CASL as part of their software package. Several techniques have been developed to overcome this problem. They differ in the way label and control scans are realized [15]. In a recent study, Pohmann et al. investigated the sensitivity of four of these CASL techniques using both simulation and experimental data [15]. Briefly, the CASL implementations tested were the following:

- (i) dual-coil (DC-CASL) which works as the conventional CASL method described above except that it uses two separate coils for labeling and imaging, respectively [16, 17]. Since labeling is achieved independently from imaging, the MT effects are completely avoided. However, the need for an additional transmit channel increases the level of technical expertise and hardware requirements therefore making DC-CASL difficult for routine application.
- (ii) Almost continuous ASL (ACASL) where the labeling pulse is regularly and briefly interrupted thus alleviating the burden on the RF amplifier to produce long pulses. To ensure equal MT effects on both label and control images, two different variations of the control acquisition were considered: one restricted the imaged volume to a single slice, referred to as ss-ACASL, whereas the other allowed for multislice acquisition hence referred to as ms-ACASL.
- (iii) Pseudo-continuous ASL (pCASL) which instead of a long rectangular labeling pulse applies short and shaped pulses in combination with rephased gradients to adiabatically invert the inflowing spins [18]. An advantage of this sequence compared to DC-CASL is that it does not require an additional labeling coil and it can be implemented with the standard coils provided by the manufacturer.

As expected from theoretical considerations, ss-ACASL yielded higher SNR followed by the dual-coil DC-CASL acquisition. However, both these methods limit the application of CASL for routine CBF measurement as the first

is constrained to a single slice whereas the second requires additional personnel expertise and hardware work to be implemented. Realistically, pCASL is the optimum choice because it can be widely applied using standard hardware without substantial sacrifice in SNR. As discussed below, we believe that currently pCASL holds the best promise for routine application of ASL at higher magnetic fields.

2.2. Pseudo-Continuous ASL (pCASL). As mentioned above, one of the major drawbacks of CASL is the requirement for long RF labeling pulses, which, in addition to causing MT effects that confound the ASL signal, are also quite taxing on the hardware and therefore not widely available. Because PASL uses short RF pulses, it is less susceptible to MT effects. However, PASL suffers from low sensitivity compared to CASL; the SNR of PASL can be 30% to 50% lower than that of CASL [19–21].

Pseudo-continuous ASL was developed as an intermediate technique that takes advantage of CASL's superior SNR and PASL's high labeling efficiency without the need for long labeling pulses [18, 20, 22]. This is achieved by using a train of short RF pulses rather than continuous RF to bring about inversion of the water spins [22]. Given that the ASL signal represents the amount of labeled spins captured within an imaged voxel, a more efficient labeling pulse and decreased MT effects in pCASL should be reflected in higher SNR compared to conventional ASL. This theoretical statement was empirically evaluated by Wu et al. who compared optimized pCASL with standard PASL and CASL at 3T [20]. The authors found that pCASL provides a 50% improvement in SNR compared to PASL and an 18% increase in labeling efficiency compared to CASL (80% versus 68%, resp.) [20].

In addition to increased intrasubject SNR and feasibility, pCASL has been shown to have lower intersubject variability compared to standard ASL [23]. Gevers et al. compared the reproducibility and reliability of pCASL, conventional CASL, and PASL based on images acquired on 6 healthy volunteers who were scanned twice at three different imaging centers [23]. When compared to conventional CASL and PASL, pCASL with background suppression showed the least data dispersion and best reproducibility [23]. While Gevers et al. study was done on only healthy young volunteers, Xu et al. compared the reliability and precision of pCASL with those of ^{15}O PET on 8 young healthy subjects and 14 elderly, 2 of which were diagnosed with early Alzheimer's disease (AD) [24]. The authors found that compared to previous ASL and PET perfusion studies, pCASL offers as good or even better reliability in repeated measurements for both young and elderly subjects. The relationship between quantitative ASL CBF, age, and AD was found consistent with previous reports, further validating the approach [24].

Recently, Dai et al. have proposed a new approach for achieving pseudo-continuous labeling of the arterial spins with higher efficiency than CASL for the same RF power deposition at 3T [18]. The main new feature of this implementation is that it replaces the rectangular train of RF pulses with a more sophisticated Hanning pulse to achieve a more precise labeling profile without compromising the labeling

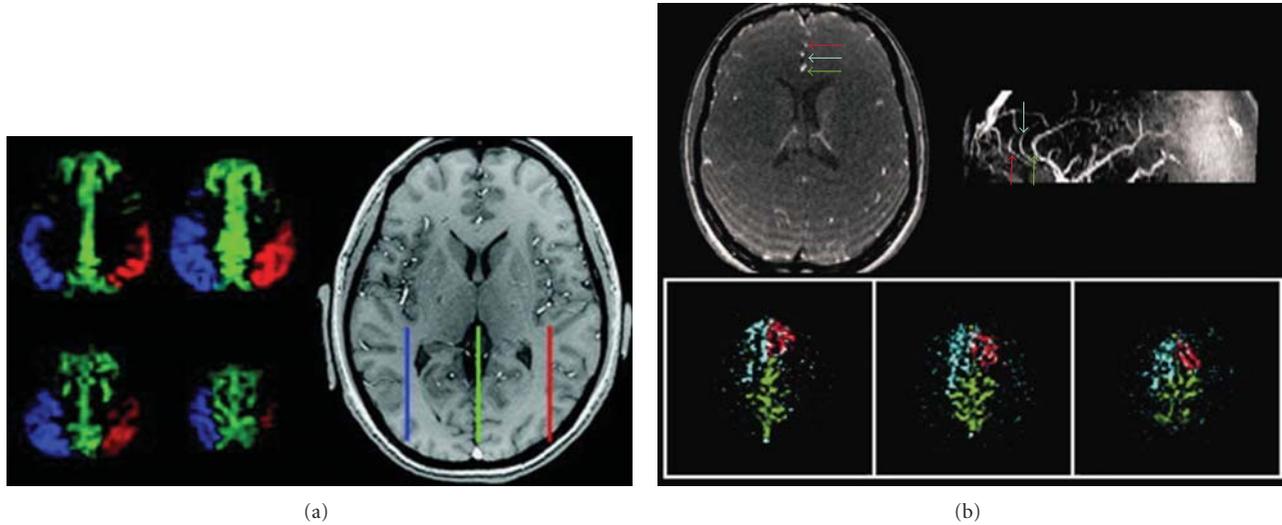


FIGURE 2: (a) Three-vessel encoding of vasculature above the Circle of Willis. In the labeling plane shown on the right, the ACA is well confined to the midline with the corresponding flow territory represented in pCASL perfusion maps as green. The territories supplied by the insular branches of the MCAs are also well labeled (shown in red and blue). Figure taken without modification from Wong [25]. (b) Selective encoding of three branches of the ACA using super-selective pCASL [26]. Top row shows TOF image and the corresponding sagittal maximum intensity projection with the branches of the ACA color coded. The bottom row shows perfusion-weighted images of the territories fed by each vessel. Figure was taken without modification from Helle et al. [26].

efficiency of the previous pCASL implementations [18]. In good agreement with the Wu study, labeling efficiency for the in vivo imaging with pCASL was 81% for the same power as standard CASL at 3T.

Other modifications of the pCASL sequence that seek to selectively label individual vessels have been introduced by Wong [25] and Helle et al. [26]. Wong combined vascular territory imaging [27–29] with the pCASL sequence [22] to simultaneously acquire perfusion images of two or more vascular territories with SNR close to that of standard ASL within the same total scan time (Figure 2(a)). Applying a different sequence modification, Helle et al. also used pCASL to selectively and independently encode vessels feeding into different perfusion territories (Figure 2(b)). As we discuss later in this review, these new developments in ASL that allow for independent labeling of different flow territories are especially relevant for clinical research.

2.3. Pulsed ASL (PASL). In contrast to CASL, labeling in PASL is achieved using more easily implemented short pulses (usually 10–15 ms) that invert spins in a specific region commonly referred to as the inversion slab (for a thorough review of PASL techniques cf. [5]). Depending on how labeling is applied with respect to the imaging volume, PASL techniques are divided into two main groups: symmetrical and asymmetrical. The original symmetrical PASL, called flow-sensitive alternating inversion recovery (FAIR), was developed in mid 1990s by Kwong et al. [30], Kim [31], and Schwarzbauer et al. [32]. The sequence consisted of two inversion recovery acquisitions: one involved slice-selective inversion, that is, the magnetization was inverted only at a selected slice, whereas the other was a non-slice-selective inversion. A delay was introduced after each inversion pulse

and before image acquisition. After the delay, the tissue magnetization of the imaged volume of the slice-selective inversion includes the signal from the inflow of the uninverted blood. On the other hand, the tissue magnetization after the nonselective inversion is expected to be approximately equal to that of the tissue (assuming T_1 of the blood to be approximately the same as the tissue T_1 , an assumption that works relatively well for gray matter). Similar to CASL, the difference between two consecutive images acquired using the two inversion types, respectively, results in a perfusion-weighted image. Although various versions of FAIR have been developed [5], the original FAIR sequence is still the most commonly used. Kim et al. extended the application of the original FAIR to multislice acquisition using a single inversion pulse to keep the temporal resolution sufficiently low for detecting motor activation [33]. Although FAIR is easy to implement and relatively straightforward in its application, multislice application remains problematic due to artifacts caused by imperfections in the profile of the inversion slice [5, 33].

The asymmetrical PASL sequence, called echo planar imaging and signal targeting with alternating radiofrequency (EPSTAR), was first proposed by Edelman et al. in 1998 [34, 35]. In this sequence, magnetization is inverted in a thick slab proximal to the imaging slice, followed by fast imaging (EPI) after a short delay that allows for the inverted magnetization to reach the imaged slice. The additional control image is similarly acquired after inverting magnetization in a slab symmetrically distal to the imaging slice and thus having the same MT effects. Proximal inversion with a control for off-resonance effects (PICOREs) [36] and transfer insensitive labeling technique (TILT) [37] sequences are based on the original EPSTAR technique.

As mentioned in the previous section, several ASL methods have been developed that, rather than labeling all of the feeding arteries, allow for selective imaging of individual perfusion territories. Regional perfusion imaging (RPI) technique developed by Van Laar et al. is based on concatenated TILT sequence labeling pulses and allows a labeling slab to be positioned at any angulation with respect to the imaging volume [38]. The RPI technique was the first to allow regional CBF measurement of individual feeding arteries. However, the sequence is very sensitive to magnetic field inhomogeneities and thus not best suited for high field imaging. A more recent sequence, pulsed star labeling of arterial regions (PULSARs), [29, 39] also based on the original EPSTAR sequence, has used an optimized water suppression pulse that presaturates the imaging volume thus increasing the sensitivity of the signal to flow [39]. Compared to the original RPI method, the PULSAR technique is less sensitive to field inhomogeneities, has better labeling efficiency and higher SNR. However, the sequences are difficult to implement and suffer from low sensitivity.

Combining the PULSAR labeling technique with a look-locker method for sampling at multiple time points and a periodic saturation scheme for clear definition of the arterial blood bolus, a quantitative STAR labeling of arterial regions (QUASAR) [40] technique was developed. Deconvolving the signal from multiple time points, the QUASAR method yields a simultaneous measurement of both arterial blood volume (aBV) and CBF [40].

Although it has yet to become routine in functional imaging of the brain, several other studies have reported on simultaneous measurement of aBV and CBF [3, 41, 42]. Simultaneous measurement of aBV and CBF using endogenous tracers may become an important tool in studying diseases in which the two physiological parameters may be dissociated.

2.4. Velocity Selective ASL (VSASL). While the pulse sequence of VSASL contains all the main elements of the conventional ASL acquisition of label and control images, the difference is that in VSASL labeling is achieved based on the velocity of the arterial water rather than its position. Using velocity selective pulses, a velocity cutoff, V_c , is imposed with the resulting labeled image containing, at least theoretically, only the spins whose velocity, V , meets the condition $V \leq V_c$. Assuming that the velocities in the arterial tree are monotonically decreasing, the amount of labeled blood in a given imaged voxel, that is, the ASL signal, is simply

$$\text{ASL}_{\text{signal}} = \text{PLD} \cdot \text{CBF}, \quad (2)$$

where CBF and PLD represent the amount of flow and the postlabeling delay for that voxel [9].

An implication of (2) is the interdependence of the ASL signal on V_c via CBF and PLD. Wu et al. performed a systematic evaluation of the interaction between PLD and V_c and showed that while the experimental data were in good agreement with the expected flow values in gray matter, a significant signal from large vessels persisted for velocities up to $V_c = 4 \text{ cm/s}$ [43]. Therefore, the authors recommended

a low cutoff $V_c = 4 \text{ cm/s}$ for quantitative measurement of tissue perfusion.

3. Which ASL Is Better?

With all the available ASL implementations, the obvious question is “which one is the best”? The answer is complex and perhaps warrants a review paper of its own, but the choice would depend on the application and should obviously involve, among others, considerations of availability of hardware, software, and technical expertise, as well as brain coverage needed and SNR assumed by the power analysis for the tested hypotheses. Unfortunately, the choice is often based on availability rather than scientific considerations. The need for technical expertise and sequence development, which to a certain degree depend on the type of scanner available, has hampered application-specific optimization of ASL imaging.

Generally, PASL has been more widely used because it is easier to implement and conceptually more straightforward than CASL. Also, because shorter labeling pulses are needed, PASL sequences are less affected by MT than the standard CASL sequence. However, drawbacks are still present such as low SNR, high sensitivity to transit times, and slice profile artifacts that can limit brain coverage. Although more difficult to implement, CASL, on the other hand, has been shown to yield higher SNR for whole brain imaging than PASL. While there is promise in VSASL, the technique is still relatively new, and more studies are needed to assess its sensitivity and applicability in disease and activation studies [44].

The recent development of pCASL, which draws on the respective advantages of CASL and PASL to provide reliable perfusion images with high SNR, has contributed to a substantial increase in applications of ASL at 3T. Due to its high efficiency, multislice capability, and relative ease of implementation without over-taxing the hardware, pCASL is becoming the best choice for a broad range of applications in brain research. Also, the potential of pCASL to selectively label vessels varying in size and orientation without compromising the SNR may prove invaluable in studying disease diagnosis, progression, and treatment.

With the concomitant advances in parallel imaging and fast acquisition pulses, ASL is primed to become the essential fMRI method for brain research. However, as mentioned above, the ASL signal is transformed into a physiological unit of CBF using a set of known or assumed MR and physiological parameters such as relaxation times, partition coefficient, transit times, inversion efficiency, and so forth [45]. It follows that any error in the estimation or assumption of these parameters would affect the absolute quantification of CBF. Furthermore, ASL images typically go through a processing algorithm that involves realignment, tissue segmentation, and normalization to the MNI or Talairach space for group analyses [46]. A detailed description of how each of these parameters and steps can affect quantification of flow is beyond the scope of this review. Here we focus on the basic principles of CBF quantification and describe recent analytical methods that have been developed to

increase ASL's sensitivity for detecting changes in CBF while minimizing the effects of confounds such as partial volume effects (PVEs) and arterial transit times (ATTs).

4. From Signal to Absolute Quantification of CBF

ASL is based on the theory of tracer kinetics, which was first applied for measuring CBF in humans by Kety and Schmidt in 1948 [47]. In ASL, the tracer is the magnetically labeled arterial water, which is a diffusible endogenous tracer. The theory provides the mathematical tools that describe the relationship between the arterial concentration of the labeled water and the resulting tissue concentration. These mathematical tools were the basis of the "general kinetic model for quantitative perfusion imaging with ASL" developed by Buxton et al. [48] in which the Bloch's equation for longitudinal magnetization was modified to include delivery and clearance terms proportional to local blood flow as shown in (3):

$$\frac{dM_T(t)}{dt} = \frac{M_T^0 - M_T(t)}{T_1} + f \cdot \left(\frac{\lambda M_A(t) - M_T(t)}{\lambda} \right), \quad (3)$$

where M_T^0 is the equilibrium magnetization of tissue, λ is the partition coefficient for water, and M_T and M_A represent the time-dependent longitudinal magnetizations of tissue and arterial blood, respectively, [48].

Based on this generic model, several solutions have been suggested for both CASL and PASL techniques [39, 45]. Also, based on the timing of the acquisition parameters, models have been constructed to mathematically describe the signal from the various compartments within the brain [2, 21, 45]. For a short postlabeling delay, most of the label is assumed to be in the arterial compartment whereas for longer delays a two-compartment model separates the tissue signal from that of the arterial blood. Although numerous studies have shown good agreement of CBF values with the more conventional flow measurement techniques such as autoradiography, microsphere method, and PET [49–51], there are several confounds that affect the absolute quantification of CBF with ASL. In the next few sections, we review some of these confounds and describe new analytical methods that have been developed to minimize their impact on CBF quantification.

4.1. Partial Volume Effects. One of the constraints in ASL imaging is the need for fast image acquisition to ensure that the signal from the labeled blood is captured before it relaxes to its equilibrium state. Fast imaging is done at the expense of spatial resolution, which means that the signal from a given voxel will reflect a mixture of signals generated from all the three main brain tissues—gray matter (GM), white matter (WM), and CSF—comprising the voxel [52]. Because the flow values from each of these tissues are different, a difference in flow values between two voxels could be simply due to a difference in tissue heterogeneity rather than a true difference in flow. These cross-tissue contamination effects,

known as partial volume effects (PVEs), are a direct consequence of limited spatial resolution in imaging in general. In ASL, PVEs are exacerbated by the nonlinear dependency of its signal on tissue heterogeneity via contributions from GM, WM, and CSF in the control image, M_C , in the denominator of (1), (cf. [52] for details).

PVEs can be quite appreciable in cortical regions where GM can be as thin as 2 mm. To give a sense of the magnitude of these effects, a voxel containing 80% GM and 20% CSF would be generally assumed and analyzed as a GM voxel [52]. For such a voxel, a simple calculation based on (1) and assuming the relative tissue magnetization intensities for SE-EPI to be $M_{CSF} : M_{GM} : M_{WM} \sim 1.6 : 1.2 : 1.0$, CBF in GM would be underestimated by $\sim 24\%$ [52].

Recently, postprocessing algorithms have been developed to correct for PVEs in ASL imaging [52, 53]. In the original method developed by our group, a linear regression algorithm is spatially applied on the difference, $(M_C - M_L)$, image as well as the control, M_C , image. The algorithm assumes that within a given spatially selective kernel, the equilibrium magnetization values and blood flow are uniformly distributed [52]. For the M_C image, the algorithm models the equilibrium magnetization of a given voxel as a weighted sum of contributions from each tissue and each voxel within the kernel. For the $(M_C - M_L)$ image, that is, the perfusion-weighted image, the intensity at each voxels is expressed as a weighted sum of flow distributions from GM and WM within the kernel, independently. In both cases, the weighting coefficients are the tissue's fractional volumes obtained as posterior probability values from the segmentation of a high-resolution images [52].

This method yields a measure of flow, referred to as "pure flow" or "flow density" that is independent of voxel's tissue content. In other words, one can compute a *pure* gray matter CBF ($CBF_{d,GM}$) and a *pure* white matter CBF ($CBF_{d,WM}$) image, independently. For each voxel, CBF_d from a given tissue represents the amount of blood flow the voxel would have were it comprised entirely of that tissue [52]. As we discuss in some detail below, this novel parameter, CBF_d , has been shown to be more sensitive in detecting changes in CBF over time than the net CBF obtained with conventional ASL [54, 55].

A disadvantage of this method is that the linear regression is applied spatially thus causing an inherent spatial smoothing of the raw data. Due to SNR considerations, the larger the spatial kernel of the linear regression the higher the SNR. On the other hand, the larger the kernel, the larger the smoothing effect of the PVE correction (PVEc) algorithm. This could be detrimental to detecting localized changes in CBF such as those found in stroke or in highly localized activation paradigms. To circumvent this drawback, Chappell et al. have implemented the above method in the time domain by acquiring multiple ASL images with varying delay times [53]. Because it is applied in time rather than space, this PVEc method protects spatial features of CBF, thus avoiding the introduction of added smoothness to the boundaries of the regions of hypo- or hyperperfusion (Figure 3). In this case, the drawback is the time needed to acquire the data, which

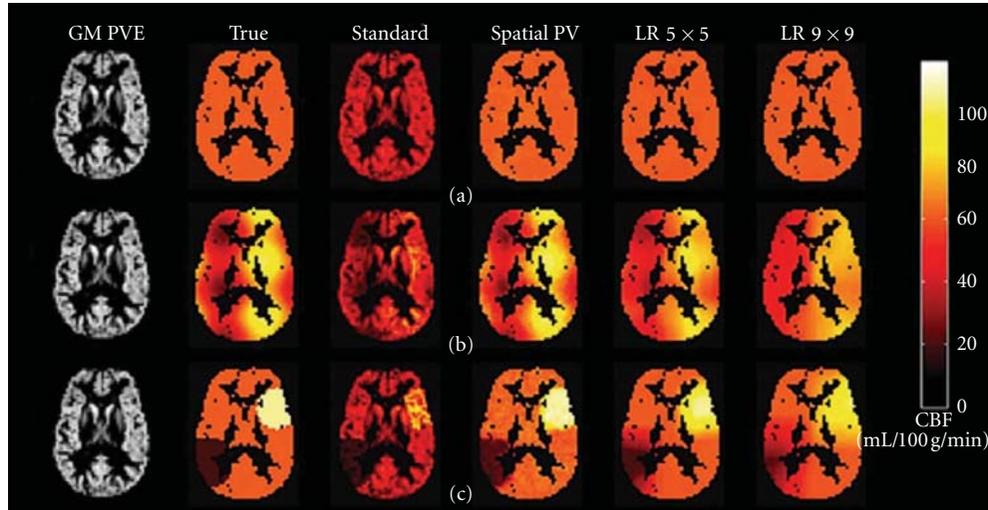


FIGURE 3: CBF images from three simulation data sets: (a) homogenous gray matter CBF, (b) superimposed spatially sinusoidal fluctuation, and (c) localized regions of hypo- and hyperperfusion. CBF images from conventional CASL (3rd column) are compared with those from PVE correction performed in the time domain [41] (4th column) [40], and PVE correction done spatially with a small and a large kernel size (columns 5th and 6th, resp.). Note that the time-domain PVE (4th column) retains the spatial features of the true hypo/hyperperfused regions (2nd column) whereas the spatially applied PVE method has a smoothing effect that increases with the size of the kernel. Figure was taken without modification from Chappell et al. [53].

is an impediment especially for activation studies or studies where the patient's time in the scanner is restricted.

4.2. The Transit Times Confound. For all ASL techniques, blood is labeled at a location distal from the region of interest. Therefore, if acquisition were to take place immediately following the labeling pulse, not all of the labeled blood would have made it into the tissue, and consequently CBF would be misestimated; a voxel containing arterial blood with labeled spins destined for another voxel would have its CBF overestimated whereas a voxel that was imaged prior to all the labeled blood having reached it would have its CBF value underestimated. To describe the transit times of the blood from the labeling location to a given imaged voxel, two physiological parameters have been defined: arterial transit time (ATT), which represents the average time it takes the blood to cross the vasculature from the labeling plane to the microvasculature in the region of interest, and tissue transit time (TTT), which is the time it takes labeled blood to exchange with region's tissues [2]. The postlabeling delay is inserted at the end of the labeling pulse to allow for the labeled blood to reach the volume of interest and exchange with the tissue. However, a compromise needs to be made between the length of the delay and loss of signal due to relaxation processes. For regions with long transit times, the delay will still not be sufficient, and the interpretation of data becomes more complicated [2].

For certain applications where CBF needs to be measured at a localized region, the distance between the labeling location and the region of interest can be shortened (within the limits imposed by the MT effects) and the effect of transit times minimized. However, for many studies, hypotheses involve whole brain acquisition of the CBF in which the CASL techniques with labeling plane positioned in the carotid

would be the method of choice. In this case, different areas of the brain will have different transit times therefore quantification would require knowledge or estimation of these times for each region. For studies in which the one-compartment arterial model is employed, that is, when the labeled blood is assumed to be mostly in macrovasculature, ATT is the main parameter that needs to be estimated; for the two-compartment model that includes the tissue microvasculature, estimation of TTT becomes essential.

By acquiring multiple ASL images at varying postlabeling delay values, ATT can be estimated via a parametric fit of the curves representing the fractional ASL signal versus time [2]. Because this step requires relatively long scanning times, ATT is not routinely measured in ASL imaging. Instead, when computing the CBF, ATT values are generally assumed to be either homogeneous throughout the brain or uniformly distributed within an acquisition slice and varying linearly with the ascending slice positions.

Recently, there have been two developments in estimation of ATT: first, improvements in technology and increased SNR at higher fields have allowed for voxelwise and ROI-wise parametric fitting of the multiple-PLD curves. Second, a method that varies labeling duration rather than postlabeling delay has been developed [55]. This method has been shown to be ~30% shorter than the multiple-PLD method [55]. Results have shown substantial heterogeneity in mean ATT values across the brain and across subjects even for healthy young volunteers (Figure 4), [55].

To estimate TTT, Wang et al. proposed a method that involved the use of flow encoding bipolar gradients to obtain the ratio of the perfusion signals in the vascular and microvascular compartments as a function of postlabeling delay [56]. The global mean tissue transit time was estimated at 1100 and 1400 ms for two conditions of the bipolar

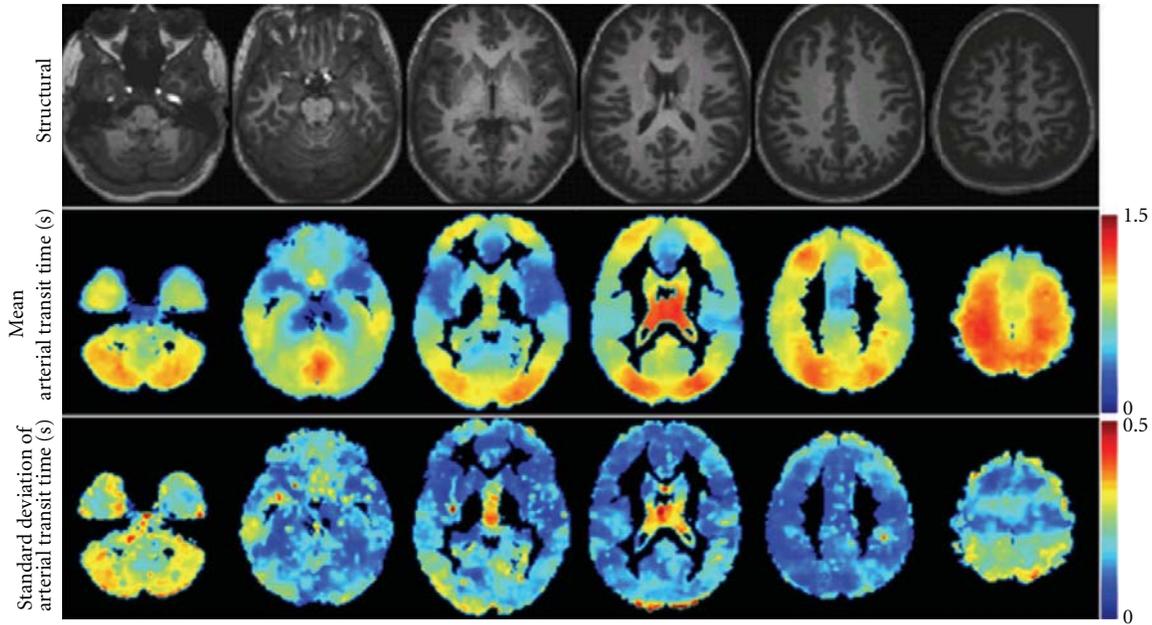


FIGURE 4: Voxelwise estimation of ATT values using a multiple-labeling-duration acquisition described in Borogovac et al. [55]. The units in the color bars are in seconds. Note regional heterogeneity in group mean ATT shown in the 2nd row. Also, across subjects standard deviation maps (3rd row) indicate variability in ATT especially in the posterior regions. This variability is expected to be higher in disease. Figure taken with permission from Borogovac et al. [55].

gradients with encoding velocity of 29 and 8 mm/sec, respectively. The mean TTT measured within cerebral vascular territories was shortest in the deep middle cerebral artery (MCA). As proof of concept, the method was applied on two patients with cerebrovascular disease where prolonged tissue transit times were demonstrated in the affected hemisphere [56]. However, the method suffers from dependency on the specific encoding velocity, and its routine application is hampered by the need for multiple-PLD acquisition.

With the above general background on the theoretical basics of ASL fMRI and its confounding factors in assessing CBF, we proceed with reviewing the advantages and some of the experimental challenges for applications of ASL in basic and clinical research.

5. ASL fMRI: Better Than BOLD?

Several imaging methods have been developed that exploit the neurovascular coupling of neuronal activity to local changes in CBF cerebral blood volume (CBV), and other physiological correlates [57]. The most pervasively used has been blood oxygenation level-dependent (BOLD) MRI, which, since its discovery in the early 1990s has been extensively used to map regions in the brain that respond to task-specific activation [58].

BOLD is a susceptibility-based method that creates “functional,” $T2^*$ -weighted images by exploiting local inhomogeneities in the magnetic field due to changes in the relative concentrations of oxygenated and deoxygenated hemoglobin (dHb) accompanying brain activation [58].

In contrast with the nuclear medicine methods, both BOLD and ASL MRI use endogenous tracers and therefore

are completely noninvasive and more readily available. Because BOLD has higher SNR and higher temporal resolution than ASL, it is more suited for event-related designs, especially when absolute quantification is not essential to the hypothesis being tested. Also, BOLD is easier to implement and does not usually require any additional programming of the RF and gradient pulses already provided by the manufacturer.

However, ASL offers several advantages over BOLD, especially in applications where slow varying changes in brain function are investigated:

(1) *Spatial Localization.* Because the BOLD effect originates from an intricate interplay between changes in CBF, CBV and oxygen consumption, its signal is composite in nature and unable to pinpoint to a single correlate of neuronal activity [59–61]. Furthermore, because the signal comes primarily from the intravascular dHb, the spatial correlation to the actual site of activation is relatively poor with considerable spatial spreading onto the venous structures [62]. In contrast, the ASL signal is straightforward to interpret because it reflects, at least theoretically, a single physiological process, namely, CBF. Consequently, the task-specific patterns mapped with ASL yield better spatial correlations with the actual site of regional involvement than BOLD [63].

(2) *Signal Quantification.* The conclusions from BOLD studies have been mainly qualitative in nature as the baseline values are commonly unaccounted for, and the signal is typically expressed in percent change [64]. The effect of baseline variability in BOLD fMRI data has been experimentally

shown by Cohen et al. [65] and Brown et al. [66]. Both studies reported a mismatch between the change in baseline CBF and the corresponding BOLD response on the same subjects and for the same stimulus [65, 66].

ASL, on the other hand, yields a physiologically quantifiable measure thus allowing baseline levels to be directly compared before and after activation [21].

(3) *Power Spectrum.* The power spectrum of the BOLD signal shows higher amplitudes at low frequencies in what has been described as the $1/f$ noise. This temporal autocorrelation makes BOLD fMRI unsuitable for application in experimental designs with fundamental frequency below 0.01 Hz, that is, for task events spaced by more than ~90 seconds apart [67]. In contrast, due to pairwise subtraction of adjacent time points, the power spectrum of ASL is essentially frequency independent, which makes it ideal for tracking slow varying changes in the brain such as those due to emotional responses, mood changes, disease, drug therapy, and so forth. In a recent study, Borogovac et al. [55] used PVEc ASL fMRI to compare changes in CBF due to motor-visual activation within the same session and across two sessions separated by 1 month (Figure 5). The study underscored the utility of the CBF_d parameter mentioned above in detecting longitudinal changes in CBF. Because this physiological parameter is relatively independent of tissue heterogeneity across subjects, it was more stable across time and ~60% more sensitive in detecting changes due to activation [55].

(4) *Susceptibility Effects.* Because BOLD is a susceptibility-based technique, gradient echo (GE) EPI is commonly employed to achieve maximum sensitivity. Consequently, BOLD is prone to artifacts in the areas with high susceptibility such as those around tissue-bone or tissue air boundaries, especially at high fields. ASL, on the other hand, can be combined with spin echo (SE) imaging to reduce bulk susceptibility artifacts thus yielding greater sensitivity in lower brain regions and more precise localization [68]. However, as higher field scanners become more available, the feasibility of using spin-echo-based BOLD fMRI is also increasing.

6. ASL fMRI in Aging and Disease

The BOLD response is a sensitive indicator of *where* neural activity occurs, but it is very difficult to interpret the magnitude of the BOLD response as a quantitative reflection of underlying physiology. The effect of the baseline state is perhaps the most serious issue for interpreting BOLD measurements in disease. For example, in a recent study of subjects at risk for AD, Fleisher and colleagues have found a reduced BOLD response in the hippocampus to a memory task in the at-risk subjects compared with controls [69]. However, by including ASL measurements as well, they found that, during the performance of the task, the two groups had similar absolute levels of flow, but that flow in the baseline state was elevated in the at-risk group [69].

According to a 2008 review paper by Deibler et al. [70], during a period of one year more than 3000 ASL procedures

were performed as part of routine clinical brain MRI evaluation at 1.5 T and 3.0 T. As mentioned above, much of the value of ASL imaging comes from its noninvasive nature and the fact that it can be acquired within a routine MR scan commonly prescribed to patients.

In general, ASL applications in the clinical realm can be divided into two main groups: vascular diseases such as stroke and carotid occlusive diseases [71–73], and “functional” diseases, including normal aging [54], Alzheimer’s disease (AD) [74], and schizophrenia [75]. This dichotomization is not meant to be inclusive, but it serves to underscore the physiological basis for the observed CBF measure. In vascular diseases, changes in CBF are to a response in structural changes in the brain, such as carotid occlusions, hematomas, tumors, or the advent of a stroke and other ischemic events. In “functional” diseases, changes in CBF (in time or as compared to healthy populations) can occur independently of structural changes in the brain or precede them.

In their recent review of ASL applications in routine clinical practice, Deibler et al. described the use of ASL for a range of diseases where hyperperfusion can be detected both focally, as in luxury perfusion, spontaneous recanalization, seizure activity, tumors, among others, and globally, as in young populations, or during conditions of hypercapnia, and reported cases of postcarotid endarterectomy [70, 76].

It is important to emphasize that the ASL confounds described in the section above become even more relevant in clinical applications. For example, in studies of stroke and carotid occlusive diseases, estimation of transit times is of primary importance [71, 72]. In this case, concomitant measurement of CBF would increase the reliability of the transit time measurements and would make the interpretation of the results more straightforward.

In studies that involve comparison of CBF between young and elderly populations, PVE becomes a main confound because of the atrophy present in the latter [54]. Recently, our group has applied the PVEc algorithm to ASL data acquired on young and elderly populations. The largest PVE contribution was found in the frontal lobe and accounted for an additional 10% and 12% increase in the age-related CBF difference between men and women, respectively, [54].

7. Future Directions in ASL fMRI

There are two significant challenges in ASL imaging that continue to hamper its routine application in brain research and comprise the main focus of current ASL development research: low SNR and relatively low temporal resolution. Our numerical simulations, based on tissue relaxation times at 3 T, assuming average GM CBF_d of 100 mL/(100 g × min) [77] and Gaussian distribution of noise, have shown that the highest achievable SNR is ~4%. High field imaging is beneficial for ASL because, in addition to the expected increase in the SNR due to field considerations, there is an increase in SNR that is due to longer T1 values at higher fields; increased T1 translates into less labeling loss, that is, signal loss, due to relaxation. Wang et al. showed that, for PASL, the SNR and CNR increased 2.3x and 2.7x, respectively, for resting state perfusion at 4 T compared to 1.5 T [21]. However,

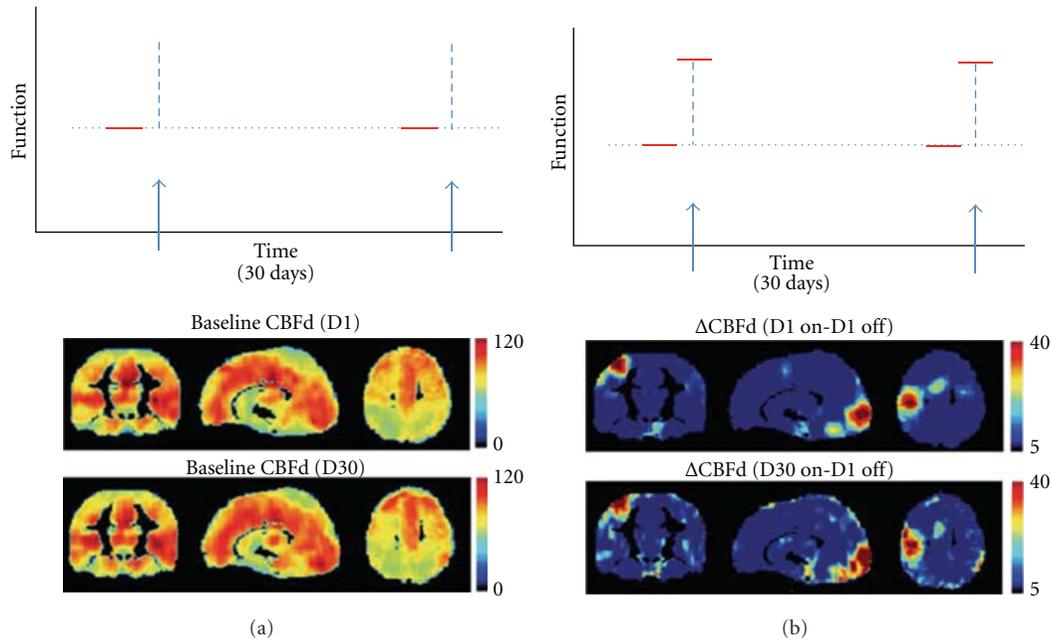


FIGURE 5: Tracking functional changes over one month. (a) Comparing baseline CBF on day 1 (indicated by first red horizontal line in upper panel, illustrating experimental design), and baseline CBF on day 30 (second red line in upper panel) shows stability over time. Middle panel shows the whole brain maps on day 1, and lower panel shows whole brain maps on day 30. (b) Comparing acute CBF changes induced by visual or motor stimulation on day 30 to baseline CBF on day 1 (as illustrated in the upper panel) is similar to acute changes induced by visual or motor stimulation on day 1 to baseline CBF on day 1. Middle panel shows the whole brain maps of day 1 stimulation to day 1 baseline, and lower panel shows whole brain maps of day 30 stimulation to day 1 baseline. Maps show similar motor and visual cortex activations. (Note that this is a modified version of Figure 3 in Borogovac et al. [55]).

there was no significant improvement in sensitivity for detecting changes in CBF due to motor activation [21], which the authors attributed to increased physiological noise and susceptibility-related artifacts at 4 T. Because of the need for fast sequential scanning of control and labeled images, ASL has generally relied on EPI imaging, which is problematic in higher fields due to field homogeneity imperfections that introduce distortions in regions of high magnetic susceptibility. One solution has been to combine fast three-dimensional (3D) sequences with ASL imaging to provide higher SNR while reducing image distortions [24, 78]. Another approach for increasing SNR in ASL is the use of a phase array receiver coils, which allow for image acquisition with shorter echo times; a decrease in echo time is beneficial both in terms of SNR and in reducing distortions due to susceptibility artifacts [79].

As mentioned above, temporal resolution is also inherently poor in ASL, especially for detecting fast changes in brain function due to activation. This is a direct consequence of the pairwise acquisition in ASL; to obtain one CBF image, two images, control and label, have to be acquired, thus doubling the effective TR, which generally varies between 4 s to 8 s. So far there have been two emerging methods for improving temporal resolution in ASL: turbo-ASL [9] and single-shot ASL [80]. Because of the complexity of signal quantification in both techniques, they are restricted to applications in which absolute quantification is not of primary importance. Hernandez-Garcia et al. combined a two-coil approach with turbo CASL for detecting perfusion responses

in both block-design and event-related experiments [81]. The higher temporal resolution was achieved by collecting the control and labeled images after a single labeling period. With the advantage of optimum SNR of CASL and increased temporal resolution, the authors reported satisfactory sensitivity for detecting perfusion response to an event-related paradigm [81].

It has become clear that, given the complexity of questions in brain research today, no single technique can be the panacea of the experimental challenges we face in answering them. The solution relies on combining the advantages of various imaging techniques with advances in analytical methods for better evaluation of the physiological parameters that underlie brain function at various states. In this regard, ASL development is branched in three directions: first, developing new implementations of the technique that are more suited to applications in higher fields and can increase the spatial and temporal resolution of CBF imaging. At the moment, pCASL combined with novel fast imaging sequences holds the best promise.

The second is combining baseline ASL CBF measurement with BOLD, in what is known as calibrated BOLD fMRI [61]. With a separate measurement of blood flow with ASL, it is possible to calculate how much the oxygen metabolism would have to change to give the measured BOLD response [82]. However, until an alternative is found to the requirement for measurement under hypercapnic condition, calibrated BOLD has yet to find wide application in the clinical realm.

Third, because ASL suffers from low SNR, advances in analytical methods that boost the sensitivity of the method are imperative. Perhaps more important, sophisticated analytical methods allow us to ask more sophisticated questions about brain function. For example, recently ASL fMRI has been combined with multivariate analysis to detect covariate CBF patterns that could distinguish AD patients from healthy controls with 95% specificity and 100% sensitivity [83].

These recent insights and technical developments suggest that ASL fMRI is on the cusp of realizing its full potential for brain research.

Acknowledgments

The authors are greatly indebted to Scott Small, M.D., for insightful discussions regarding the application of fMRI in disease. They also thank the anonymous reviewer for his/her thorough critique and invaluable suggestions. Partial funding for research included in this review was provided by NIH/NIMH R21MH082308 grant.

References

- [1] R. Buxton, *Introduction to Functional Magnetic Resonance Imaging*, Cambridge University Press, 2009.
- [2] D. C. Alsop and J. A. Detre, "Reduced transit-time sensitivity in noninvasive magnetic resonance imaging of human cerebral blood flow," *Journal of Cerebral Blood Flow and Metabolism*, vol. 16, no. 6, pp. 1236–1249, 1996.
- [3] D. H. Kim and D. M. Spielman, "Reducing gradient imperfections for spiral magnetic resonance spectroscopic imaging," *Magnetic Resonance in Medicine*, vol. 56, no. 1, pp. 198–203, 2006.
- [4] D. S. Williams, J. A. Detre, J. S. Leigh, and A. P. Koretsky, "Magnetic resonance imaging of perfusion using spin inversion of arterial water," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 1, pp. 212–216, 1992.
- [5] X. Golay, J. Hendrikse, and T. C. C. Lim, "Perfusion imaging using arterial spin labeling," *Topics in Magnetic Resonance Imaging*, vol. 15, no. 1, pp. 10–27, 2004.
- [6] D. M. Garcia, G. Duhamel, and D. C. Alsop, "Efficiency of inversion pulses for background suppressed arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 54, no. 2, pp. 366–372, 2005.
- [7] K. S. Lawrence, J. A. Frank, P. A. Bandettini, and F. Q. Ye, "Noise reduction in multi-slice arterial spin tagging imaging," *Magnetic Resonance in Medicine*, vol. 53, no. 3, pp. 735–738, 2005.
- [8] A. C. Silva and S. G. Kim, "Pseudo-continuous arterial spin labeling technique for measuring CBF dynamics with high temporal resolution," *Magnetic Resonance in Medicine*, vol. 42, no. 3, pp. 425–429, 1999.
- [9] E. C. Wong, M. Cronin, W. C. Wu, B. Inglis, L. R. Frank, and T. T. Liu, "Velocity-selective arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 55, no. 6, pp. 1334–1341, 2006.
- [10] M. Sardashti, D. G. Schwartzberg, G. P. Stomp, and W. T. Dixon, "Spin-labeling angiography of the carotids by presaturation and simplified adiabatic inversion," *Magnetic Resonance in Medicine*, vol. 15, no. 2, pp. 192–200, 1990.
- [11] D. G. Norris, "Adiabatic radiofrequency pulse forms in biomedical nuclear magnetic resonance," *Concepts in Magnetic Resonance B*, vol. 14, no. 2, pp. 89–101, 2002.
- [12] S. D. Wolff and R. S. Balaban, "Magnetization transfer contrast (MTC) and tissue water proton relaxation in vivo," *Magnetic Resonance in Medicine*, vol. 10, no. 1, pp. 135–144, 1989.
- [13] J. A. Detre, J. S. Leigh, D. S. Williams, and A. P. Koretsky, "Perfusion imaging," *Magnetic Resonance in Medicine*, vol. 23, no. 1, pp. 37–45, 1992.
- [14] D. C. Alsop and J. A. Detre, "Multisection cerebral blood flow MRI imaging with continuous arterial spin labeling," *Radiology*, vol. 208, no. 2, pp. 410–416, 1998.
- [15] R. Pohmann, J. Budde, E. J. Auerbach, G. Adriany, and K. Ugurbil, "Theoretical and experimental evaluation of continuous arterial spin labeling techniques," *Magnetic Resonance in Medicine*, vol. 63, no. 2, pp. 438–446, 2010.
- [16] G. Zaharchuk, P. J. Ledden, K. K. Kwong, T. G. Reese, B. R. Rosen, and L. L. Wald, "Multislice perfusion and perfusion territory imaging in humans with separate label and image coils," *Magnetic Resonance in Medicine*, vol. 41, no. 6, pp. 1093–1098, 1999.
- [17] W. Zhang, A. C. Silva, D. S. Williams, and A. P. Koretsky, "NMR measurement of perfusion using arterial spin labeling without saturation of macromolecular spins," *Magnetic Resonance in Medicine*, vol. 33, no. 3, pp. 370–376, 1995.
- [18] W. Dai, D. Garcia, C. De Bazelaire, and D. C. Alsop, "Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields," *Magnetic Resonance in Medicine*, vol. 60, no. 6, pp. 1488–1497, 2008.
- [19] E. C. Wong, R. B. Buxton, and L. R. Frank, "A theoretical and experimental comparison of continuous and pulsed arterial spin labeling techniques for quantitative perfusion imaging," *Magnetic Resonance in Medicine*, vol. 40, no. 3, pp. 348–355, 1998.
- [20] W.-C. Wu, M. Fernández-Seara, J. A. Detre, F. W. Wehrli, and J. Wang, "A theoretical and experimental investigation of the tagging efficiency of pseudocontinuous arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 58, no. 5, pp. 1020–1027, 2007.
- [21] J. Wang, D. C. Alsop, L. Li et al., "Comparison of quantitative perfusion imaging using arterial spin labeling at 1.5 and 4.0 Tesla," *Magnetic Resonance in Medicine*, vol. 48, no. 2, pp. 242–254, 2002.
- [22] D. M. Garcia, C. Bazelaire, and D. Alsop, "Pseudo-continuous flow driven adiabatic inversion for arterial spin labeling," in *Proceedings of the 13th Meeting in International Society of Magnetic Resonance in Medicine*, 2005.
- [23] S. Gevers, M. J. Van Osch, R. P.H. Bokkers et al., "Intra- and multicenter reproducibility of pulsed, continuous and pseudo-continuous arterial spin labeling methods for measuring cerebral perfusion," *Journal of Cerebral Blood Flow and Metabolism*, vol. 31, no. 8, pp. 1706–1715, 2011.
- [24] G. Xu, H. A. Rowley, G. Wu et al., "Reliability and precision of pseudo-continuous arterial spin labeling perfusion MRI on 3.0 T and comparison with 15O-water PET in elderly subjects at risk for Alzheimer's disease," *NMR in Biomedicine*, vol. 23, no. 3, pp. 286–293, 2010.
- [25] E. C. Wong, "Vessel-encoded arterial spin-labeling using pseudocontinuous tagging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1086–1091, 2007.
- [26] M. Helle, D. G. Norris, S. Rüfer, K. Alfke, O. Jansen, and M. J. P. Van Osch, "Supersensitive pseudocontinuous arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 64, no. 3, pp. 777–786, 2010.

- [27] P. J. Van Laar, J. Hendrikse, X. Golay, H. Lu, M. J. P. Van Osch, and J. Van Der Grond, "In vivo flow territory mapping of major brain feeding arteries," *NeuroImage*, vol. 29, no. 1, pp. 136–144, 2006.
- [28] J. Hendrikse, J. Van Der Grond, H. Lu, P. C. M. Van Zijl, and X. Golay, "Flow territory mapping of the cerebral arteries with regional perfusion MRI," *Stroke*, vol. 35, no. 4, pp. 882–887, 2004.
- [29] R. Werner, D. G. Norris, K. Alfke, H. M. Mehdorn, and O. Jansen, "Continuous artery-selective spin labeling (CASSL)," *Magnetic Resonance in Medicine*, vol. 53, no. 5, pp. 1006–1012, 2005.
- [30] K. K. Kwong, D. A. Chesler, R. M. Weisskoff et al., "MR perfusion studies with T1-weighted echo planar imaging," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 878–887, 1995.
- [31] S. G. Kim, "Quantification of relative cerebral blood flow change by flow-sensitive alternating inversion recovery (FAIR) technique: application to functional mapping," *Magnetic Resonance in Medicine*, vol. 34, no. 3, pp. 293–301, 1995.
- [32] C. Schwarzbauer, S. P. Morrissey, and A. Haase, "Quantitative magnetic resonance imaging of perfusion using magnetic labeling of water proton spins within the detection slice," *Magnetic Resonance in Medicine*, vol. 35, no. 4, pp. 540–546, 1996.
- [33] S. G. Kim, N. V. Tsekos, and J. Ashe, "Multi-slice perfusion-based functional MRI using the FAIR technique: comparison of CBF and BOLD effects," *NMR in Biomedicine*, vol. 10, no. 4-5, pp. 191–196, 1997.
- [34] R. R. Edelman and Q. Chen, "EPISTAR MRI: multislice mapping of cerebral blood flow," *Magnetic Resonance in Medicine*, vol. 40, no. 6, pp. 800–805, 1998.
- [35] R. R. Edelman, B. Siewert, D. G. Darby et al., "Qualitative mapping of cerebral blood flow and functional localization with echo-planar MR imaging and signal targeting with alternating radio frequency," *Radiology*, vol. 192, no. 2, pp. 513–520, 1994.
- [36] E. C. Wong, R. B. Buxton, and L. R. Frank, "Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling," *NMR in Biomedicine*, vol. 10, no. 4-5, pp. 237–249, 1997.
- [37] X. Golay, M. Stuber, K. P. Pruessmann, D. Meier, and P. Boesiger, "Transfer insensitive labeling technique (TILT): application to multislice functional perfusion imaging," *Journal of Magnetic Resonance Imaging*, vol. 9, no. 3, pp. 454–461, 1999.
- [38] P. J. Van Laar, J. Van Der Grond, and J. Hendrikse, "Brain perfusion territory imaging: methods and clinical applications of selective arterial spin-labeling MR imaging," *Radiology*, vol. 246, no. 2, pp. 354–364, 2008.
- [39] X. Golay, E. T. Petersen, and F. Hui, "Pulsed star labeling of arterial regions (PULSAR): a robust regional perfusion technique for high field imaging," *Magnetic Resonance in Medicine*, vol. 53, no. 1, pp. 15–21, 2005.
- [40] E. T. Petersen, T. Lim, and X. Golay, "Model-free arterial spin labeling quantification approach for perfusion MRI," *Magnetic Resonance in Medicine*, vol. 55, no. 2, pp. 219–232, 2006.
- [41] M. J. Brookes, P. G. Morris, P. A. Gowland, and S. T. Francis, "Noninvasive measurement of arterial cerebral blood volume using Look-Locker EPI and arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 58, no. 1, pp. 41–54, 2007.
- [42] T. Kim and S. G. Kim, "Quantification of cerebral arterial blood volume and cerebral blood flow using MRI with modulation of tissue and vessel (MOTIVE) signals," *Magnetic Resonance in Medicine*, vol. 54, no. 2, pp. 333–342, 2005.
- [43] W. C. Wu and E. C. Wong, "Feasibility of velocity selective arterial spin labeling in functional MRI," *Journal of Cerebral Blood Flow and Metabolism*, vol. 27, no. 4, pp. 831–838, 2007.
- [44] G. Duhamel, C. De Bazelaire, and D. C. Alsop, "Evaluation of systematic quantification errors in velocity-selective arterial spin labeling of the brain," *Magnetic Resonance in Medicine*, vol. 50, no. 1, pp. 145–153, 2003.
- [45] J. Wang, Y. Zhang, R. L. Wolf, A. C. Roc, D. C. Alsop, and J. A. Detre, "Amplitude-modulated continuous arterial spin-labeling 3.0-T perfusion MR imaging with a single coil: feasibility study," *Radiology*, vol. 235, no. 1, pp. 218–228, 2005.
- [46] I. Asllani, A. Borogovac, C. Wright, R. Sacco, T. R. Brown, and E. Zarahn, "An investigation of statistical power for continuous arterial spin labeling imaging at 1.5 T," *NeuroImage*, vol. 39, no. 3, pp. 1246–1256, 2008.
- [47] S. S. Kety and C. F. Schmidt, "The nitrous oxide method for the quantitative determination of cerebral blood flow in man: theory, procedure and normal values," *The Journal of Clinical Investigation*, vol. 27, pp. 476–483, 1948.
- [48] R. B. Buxton, L. R. Frank, E. C. Wong, B. Siewert, S. Warach, and R. R. Edelman, "A general kinetic model for quantitative perfusion imaging with arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 40, no. 3, pp. 383–396, 1998.
- [49] E. G. Walsh, K. Minematsu, J. Leppo, and S. C. Moore, "Radioactive microsphere validation of a volume localized continuous saturation perfusion measurement," *Magnetic Resonance in Medicine*, vol. 31, no. 2, pp. 147–153, 1994.
- [50] F. Q. Ye, K. F. Berman, T. Ellmore et al., "H215O PET validation of steady-state arterial spin tagging cerebral blood flow measurements in humans," *Magnetic Resonance in Medicine*, vol. 44, no. 3, pp. 450–456, 2000.
- [51] J. R. Ewing, L. Wei, R. A. Knight et al., "Direct comparison of local cerebral blood flow rates measured by MRI arterial spin-tagging and quantitative autoradiography in a rat model of experimental cerebral ischemia," *Journal of Cerebral Blood Flow and Metabolism*, vol. 23, no. 2, pp. 198–209, 2003.
- [52] I. Asllani, A. Borogovac, and T. R. Brown, "Regression algorithm correcting for partial volume effects in arterial spin labeling MRI," *Magnetic Resonance in Medicine*, vol. 60, no. 6, pp. 1362–1371, 2008.
- [53] M. A. Chappell, A. R. Groves, B. J. Macintosh, M. J. Donahue, P. Jezard, and M. W. Woolrich, "Partial volume correction of multiple inversion time arterial spin labeling MRI data," *Magnetic Resonance in Medicine*, vol. 65, pp. 1173–1183, 2011.
- [54] I. Asllani, C. Habeck, A. Borogovac, T. R. Brown, A. M. Brickman, and Y. Stern, "Separating function from structure in perfusion imaging of the aging brain," *Human Brain Mapping*, vol. 30, no. 9, pp. 2927–2935, 2009.
- [55] A. Borogovac, C. Habeck, S. A. Small, and I. Asllani, "Mapping brain function using a 30-day interval between baseline and activation: a novel arterial spin labeling fMRI approach," *Journal of Cerebral Blood Flow and Metabolism*, vol. 30, no. 10, pp. 1721–1733, 2010.
- [56] J. Wang, D. C. Alsop, H. K. Song et al., "Arterial transit time imaging with flow encoding arterial spin tagging (FEAST)," *Magnetic Resonance in Medicine*, vol. 50, no. 3, pp. 599–607, 2003.
- [57] R. Buxton, *Introduction to Functional MRI*, Cambridge University Press, 1st edition, 2002.
- [58] S. Ogawa, D. W. Tank, R. Menon et al., "Intrinsic signal changes accompanying sensory stimulation: functional brain

- mapping with magnetic resonance imaging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 13, pp. 5951–5955, 1992.
- [59] S.-G. Kim and K. Ugurbil, "Comparison of blood oxygenation and cerebral blood flow effects in fMRI: estimation of relative oxygen consumption change," *Magnetic Resonance in Medicine*, vol. 38, no. 1, pp. 59–65, 1997.
- [60] B. M. Ances, O. Leontiev, J. E. Perthen, C. Liang, A. E. Lansing, and R. B. Buxton, "Regional differences in the coupling of cerebral blood flow and oxygen metabolism changes in response to activation: implications for BOLD-fMRI," *NeuroImage*, vol. 39, no. 4, pp. 1510–1521, 2008.
- [61] T. L. Davis, K. K. Kwong, R. M. Weisskoff, and B. R. Rosen, "Calibrated functional MRI: mapping the dynamics of oxidative metabolism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 4, pp. 1834–1839, 1998.
- [62] F. G. C. Hoogenraad, P. J. W. Pouwels, M. B. M. Hofman, J. R. Reichenbach, M. Sprenger, and E. M. Haacke, "Quantitative differentiation between BOLD models in fMRI," *Magnetic Resonance in Medicine*, vol. 45, no. 2, pp. 233–246, 2001.
- [63] G. K. Aguirre, J. A. Detre, and J. Wang, "Perfusion fMRI for functional neuroimaging," *International Review of Neurobiology*, vol. 66, pp. 213–236, 2005.
- [64] R. G. Shulman, D. L. Rothman, and F. Hyder, "A BOLD search for baseline," *NeuroImage*, vol. 36, no. 2, pp. 277–281, 2007.
- [65] E. R. Cohen, E. Rostrup, K. Sidaros et al., "Hypercapnic normalization of BOLD fMRI: comparison across field strengths and pulse sequences," *NeuroImage*, vol. 23, no. 2, pp. 613–624, 2004.
- [66] G. G. Brown, L. T. E. Zorrilla, B. Georgy, S. S. Kindermann, E. C. Wong, and R. B. Buxton, "BOLD and perfusion response to finger-thumb apposition after acetazolamide administration: differential relationship to global perfusion," *Journal of Cerebral Blood Flow and Metabolism*, vol. 23, no. 7, pp. 829–837, 2003.
- [67] G. K. Aguirre, J. A. Detre, E. Zarahn, and D. C. Alsop, "Experimental design and the relative sensitivity of BOLD and perfusion fMRI," *NeuroImage*, vol. 15, no. 3, pp. 488–500, 2002.
- [68] J. Wang, L. Li, A. C. Roc et al., "Reduced susceptibility effects in perfusion fMRI with single-shot spin-echo EPI acquisitions at 1.5 Tesla," *Magnetic Resonance Imaging*, vol. 22, no. 1, pp. 1–7, 2004.
- [69] A. S. Fleisher, M. Donohue, K. Chen, J. B. Brewer, and P. S. Aisen, "Applications of neuroimaging to disease-modification trials in Alzheimer's disease," *Behavioural Neurology*, vol. 21, no. 1-2, pp. 129–136, 2009.
- [70] A. R. Deibler, J. M. Pollock, R. A. Kraft, H. Tan, J. H. Burdette, and J. A. Maldjian, "Arterial spin-labeling in routine clinical practice—part 2: hypoperfusion patterns," *American Journal of Neuroradiology*, vol. 29, no. 7, pp. 1235–1241, 2008.
- [71] R. P. H. Bokkers, M. J. P. Van Osch, C. J. M. Klijn, L. J. Kappelle, and J. Hendrikse, "Cerebrovascular reactivity within perfusion territories in patients with an internal carotid artery occlusion," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 82, no. 9, pp. 1011–1016, 2011.
- [72] R. P. H. Bokkers, F. J. Wessels, H. B. Van Der Worp, J. J. M. Zwanenburg, W. P. Th. M. Mali, and J. Hendrikse, "Vasodilatory capacity of the cerebral vasculature in patients with carotid artery stenosis," *American Journal of Neuroradiology*, vol. 32, no. 6, pp. 1030–1033, 2011.
- [73] M. Viallon, S. Altrichter, V. M. Pereira et al., "Combined use of pulsed arterial spin-labeling and susceptibility-weighted imaging in stroke at 3T," *European Neurology*, vol. 64, no. 5, pp. 286–296, 2010.
- [74] D. C. Alsop, W. Dai, M. Grossman, and J. A. Detre, "Arterial spin labeling blood flow MRI: its role in the early characterization of Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 20, no. 3, pp. 871–880, 2010.
- [75] L. Scheef, C. Manka, M. Daamen et al., "Resting-state perfusion in nonmedicated schizophrenic patients: a continuous arterial spin-labeling 3.0-T MR study," *Radiology*, vol. 256, no. 1, pp. 253–260, 2010.
- [76] A. R. Deibler, J. M. Pollock, R. A. Kraft, H. Tan, J. H. Burdette, and J. A. Maldjian, "Arterial spin-labeling in routine clinical practice—part 3: hyperperfusion patterns," *American Journal of Neuroradiology*, vol. 29, no. 8, pp. 1428–1435, 2008.
- [77] I. Asllani, A. Borogovac, and J. Hirsch, "Constraints in absolute quantification of CBF using arterial spin labeling fMRI," in *Proceedings of the 15th Annual Meeting of the Organization for Human Brain Mapping*, Poster Presentation, San Francisco, Calif, USA, June 2009.
- [78] M. A. Fernández-Seara, Z. Wang, J. Wang et al., "Continuous arterial spin labeling perfusion measurements using single shot 3D GRASE at 3 T," *Magnetic Resonance in Medicine*, vol. 54, no. 5, pp. 1241–1247, 2005.
- [79] R. L. Wolf and J. A. Detre, "Clinical neuroimaging using arterial spin-labeled perfusion magnetic resonance imaging," *Neurotherapeutics*, vol. 4, no. 3, pp. 346–359, 2007.
- [80] J. H. Duyn, C. X. Tan, P. Van Gelderen, and M. N. Yongbi, "High-sensitivity single-shot perfusion-weighted fMRI," *Magnetic Resonance in Medicine*, vol. 46, no. 1, pp. 88–94, 2001.
- [81] L. Hernandez-Garcia, G. R. Lee, A. L. Vazquez, and D. C. Noll, "Fast pseudo-continuous arterial spin labeling for functional imaging using a two-coil system," *Magnetic Resonance in Medicine*, vol. 51, no. 3, pp. 577–585, 2004.
- [82] F. Hyder, I. Kida, K. L. Behar, R. P. Kennan, P. K. Maciejewski, and D. L. Rothman, "Quantitative functional imaging of the brain: towards mapping neuronal activity by BOLD fMRI," *NMR in Biomedicine*, vol. 14, no. 7-8, pp. 413–431, 2001.
- [83] I. Asllani, C. Habeck, N. Scarneas, A. Borogovac, T. R. Brown, and Y. Stern, "Multivariate and univariate analysis of continuous arterial spin labeling perfusion MRI in Alzheimer's disease," *Journal of Cerebral Blood Flow and Metabolism*, vol. 28, no. 4, pp. 725–736, 2008.

Research Article

Metabolite Mapping with Extended Brain Coverage Using a Fast Multisection MRSI Pulse Sequence and a Multichannel Coil

Zhengchao Dong,^{1,2} Feng Liu,¹ Alayar Kangarlu,¹ and Bradley S. Peterson¹

¹ Department of Psychiatry, Columbia University College of Physicians & Surgeons, New York State Psychiatric Institute, New York, NY 10032, USA

² Department of Psychiatry, Columbia University and New York State Psychiatric Institute, 1051 Riverside Drive, No. #74, New York, NY 10032, USA

Correspondence should be addressed to Zhengchao Dong, zhengchaodong@gmail.com

Received 28 July 2011; Accepted 16 November 2011

Academic Editor: Eric C. Bourekas

Copyright © 2012 Zhengchao Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multisection magnetic resonance spectroscopic imaging is a widely used pulse sequence that has distinct advantages over other spectroscopic imaging sequences, such as dynamic shimming, large region-of-interest coverage within slices, and rapid data acquisition. It has limitations, however, in the number of slices that can be acquired in realistic scan times and information loss from spacing between slices. In this paper, we synergize the multi-section spectroscopic imaging pulse sequence with multichannel coil technology to overcome these limitations. These combined techniques now permit elimination of the gaps between slices and acquisition of a larger number of slices to realize the whole brain metabolite mapping without incurring the penalties of longer repetition times (and therefore longer acquisition times) or lower signal-to-noise ratios.

1. Introduction

The applications of proton magnetic resonance spectroscopic imaging (1H MRSI) of the brain can benefit from technical developments in pulse sequences and hardware advances to overcome various limitations of MRSI, including low SNR, long acquisition times, and lipid contamination. In pulse sequence development, 2D PRESS-MRSI [1] and 2D STEAM-MRSI [2] have been developed to address several of these limitations and are now widely used. By exciting only a small region of interest within the brain, both sequences permit reduction in the field of view (FOV), which in turn permits a reduction in the number of phase encoding steps and thereby reduces the scan time required to achieve a given spatial resolution [3]. These pulse sequences also reduce contamination of the metabolite signals by lipid signals from the scalp. These and other advantages have motivated the extension of PRESS- and STEAM-MRSI from conventional uses to 3D or multiple 2D applications, or to their combination with other techniques, such as echo-planar spectroscopic imaging [4] and spiral MRSI [5]. PRESS and STEAM

MRSI, however, also have several disadvantages, such as relatively small brain coverage and scan times that are still long for human applications, especially when used in 3D or multiple 2D modes [6].

Another development in fast MRSI sequences has been multi-section MRSI [7]. This sequence features two advantages over other PRESS- or STEAM-based fast sequences. (1) Multiple slices are consecutively excited and sampled within one repetition time (TR), whereas lipid signals from subcutaneous fat are suppressed through the application of oblique saturation bands placed by the user on localizer images [3]. (2) Each slice is dynamically shimmed, and therefore better spectral quality can be achieved than with global shimming. In addition, multi-section MRSI also possesses the following features. (1) Full echoes, instead of free induction decays (FIDs), are acquired, so that magnitude spectra can be used without employing phase correction. Acquiring full echoes offers an additional advantage for the reconstruction of MRSI signals sampled using a multichannel receiver RF coil, in that the residual water signals can be used as sensitivity references when combining signals from the various channels

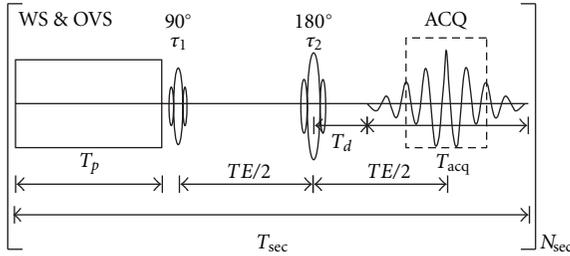


FIGURE 1: Pulse sequence diagram and timing for multi-section MRSI. The T_p is the time for the preparation period, which includes water suppression and outer volume suppression; τ_1 and τ_2 are the durations of 90° and 180° pulses, respectively; TE is the spin echo time; T_d is the time delay before data acquisition, which is used for slice selection in the z -direction and phase encoding in the x - y plane of the slice (not shown); T_{acq} is the acquisition time; T_{sec} is the time for one slice, and the total repetition time TR for N_{sec} slices is $TR = T_{sec} \times N_{sec}$. WS: water suppression; OVS: outer volume suppression; ACQ: acquisition.

of the coil [8]. (2) Spacing between slices avoids the signal cancellation caused by “crosstalk” between adjacent slices.

The conventional implementation of multisection MRSI technique [7] also has attendant problems, however. The spacing required between slices fails to acquire information throughout the entire imaging volume, not only losing important information but also creating difficulties for the coregistration of high-resolution MRI images with the MRSI data, which is important for the segmentation of MRSI slices and the analysis of the MRSI data. Although the use of a 288 ms echo time (TE) is necessary to accommodate the full number of echoes for 512 data points at a sampling rate of 2000 Hz, it also lowers the SNR and requires a long TR of 2.3 seconds to sample phase-encoded signals from 4 slices. Because each TR is “full” (Figure 1), any further increase in the number of slices is possible only by linearly increasing the TR , which undermines the “fast” feature of the pulse sequence. The limited number of slices also limits the volume coverage, requiring an increase in slice thickness and a trade-off with compromising spatial resolution to cover a larger volume with the same 4 slices.

Hardware development, and the use of multichannel coils in particular, has benefitted the applications of 1H MRSI. The initial and conventional application of multichannel coils in MRS was to improve SNR [9–12], as the array of surface coils provides superior sensitivity compared with more conventional quadrature volume coils [10]. In recent years, parallel MRSI has employed multichannel coils to accelerate data acquisition [13]. The gains in SNR or the reduction in scan time, though valuable in their own right, can be further traded for other potential advantages, such as spatial resolution of MRSI [14].

We note that the limitations of multi-section MRSI are not caused by intrinsic disadvantages of the sequence, but rather are the consequence of technical limitations and compromises that can be addressed powerfully using multichannel coil technology. We hypothesize that the use of multichannel coil technology for multi-section spectroscopic im-

aging can eliminate the need for gaps between slices and permit an increase in the number of slices without resorting to longer acquisition times or lower spatial resolution for improved brain coverage. Therefore, our aim in this report is to synergize the multi-section spectroscopic imaging (SI) pulse sequence [7] with multichannel coil technology to realize whole brain metabolite mapping by removing the gaps between slices and increasing the number of slices without resorting to trading off longer TR s and therefore longer acquisition times or poorer spatial resolution for more brain coverage.

2. Methods

2.1. MRSI Sequence. The timing of the simplified RF pulse sequence for multi-section MRSI is schematically shown in Figure 1. T_p is the time for the preparation period, which includes water suppression and outer volume suppression; τ_1 and τ_2 , are the durations of 90° and 180° pulses, respectively; TE is the spin echo time; T_d is the time delay before data acquisition, which is used for slice selection in the z -direction and phase encoding in the x - y plane of the slice (not shown); T_{acq} is the acquisition time; T_{sec} is the time for one slice; the total repetition time TR for N slices is $TR = T_{sec} \times N_{sec}$. The timing of the sequence is determined by hardware capacities and practical considerations. The times of T_p and T_d , as well as τ_1 and τ_2 for example, relate to the hardware and are optimized to the shortest possible times during the design of the pulse sequence. They are treated as constants in this application. T_{acq} and N_{sec} are variables selected by the users. T_{acq} equals N/SW , where N is the number of data points in the echo and SW is the spectral width, both of which are determined by practical considerations and compromises. For example, one way to reduce scan time, T_{sec} , is to reduce the TE , which is possible by reducing T_{acq} . Increasing spectral width can reduce T_{acq} , but does so at the expense of increasing noise. Reducing N reduces T_{acq} , but it will reduce the spectral resolution and produce truncation effects, as the acquisition may begin after the echo signal is fully built up and end before the echo signal is fully decayed, as shown in the dashed box in Figure 1. Although zero-padding the truncated echo prior to Fourier transformation may improve the spectral resolution, it will also produce wiggling in the spectrum. Both low digital resolution and wiggling will hinder spectral fitting in the frequency domain [15]. All these factors considered, SW and N were conventionally designated to be 1000 Hz and 256, respectively, for a 1.5T scanner [7] and 2000 Hz and 512 for 3T scanners. With these parameters, and for $PE = 32 \times 32$ and $N_{sec} = 4$, the TE is 280 ms, yielding a T_{sec} of 575 ms, $TR = 2300$ ms, and a scan time of 30 minutes. Together with preparation time, including slice and outer volume suppression (OVS) band prescription, shimming, and prescanning, the scan time totals 50 min. Further increasing the number of slices will proportionally increase the scan time.

We reduced the number of data points in the echoes acquired on a 3T scanner from 512 to 256, so that 7 slices will be covered in a TR of 2300 ms. To avoid the disadvantage of

low spectral resolution and the truncation effects on spectral fitting in the frequency domain associated with acquisition of a limited number of data points, we adopted an algorithm of spectral fitting in the time domain. Those details will be described below under *Data Processing section*.

2.2. Computer Simulation. We performed computer simulations to compare the effects of spectral fitting algorithms on signal truncation in the time domain and spectral resolution or sinc wiggles in the frequency domain. The signal simulated the 3 singlets of Ch, Cr, and NAA with amplitudes of 24, 36, and 48 (a.u.), respectively. The linewidth was 10 Hz for either a Lorentzian or Gaussian lineshape. For a Voigtian lineshape, the Lorentzian decay was 2.5 Hz, and the Gaussian decay was 7.5 Hz for all three lines. Spectral width was 2000 Hz, the number of data points in the echo was 256, and it was zero-filled to 512 or 1024. We also conducted a Monte Carlo simulation for signal fitting in the time domain and compared the standard deviations of the estimated amplitudes with their Cramer Rao Lower Bounds (CRLBs), a benchmark for assessing the accuracy of spectral fitting algorithms.

2.3. MR Data Acquisition. We carried out all MR measurements on a spectroscopic phantom and on 3 human volunteers, respectively, using a whole body 3T scanner (Signa HDx 3.0T, GE Healthcare, Waukesha, WI), equipped with a quadrature transmit/receive head coil, and an 8-channel receive-only head coil. First, we acquired scout images using a commercial gradient recalled echo-based 3-planar MRI sequence and then prescribed the localizer images of the MRSI slices, which were in axial plane for phantom scans and in an oblique axial plane parallel to the anterior commissure-posterior commissure line in human subjects. Then, we localized the MRSI slices by copying the location of the localizer images and acquired MRSI data using the multiplanar MRSI pulse sequence [7]. The number of slices, the slice thickness, and spacing varied in accord with those of different MRSI sessions (*vide infra*). However, typical parameters of the MRSI pulse sequence were as follows for both the phantom and human subjects: number of slices = 7; slice thickness = 10 mm; spacing between slices = 4 or 0 mm; nominal number of phase encodings (PEs) = 16×16 or 32×32 ; $TR/TE = 2300/144$ ms; spectral width = 2000 Hz; number of data points in the echo = 256. The duration of an MRSI scan was 8 or 30 minutes, depending on the number of PEs. Total scan time including MRI localizer, MRSI slice prescription, OVS band placement, and field shimming was about 26 or 50 minutes. When repeating the MRSI scans with different slice spacings, we changed only the spacing but not the position of the central (4th) slice. Therefore, we used signals from this slice to assess the effects of differing slice spacings on the ‘‘crosstalk’’ between slices. For each MRSI scan with a differing spacing between slices, we performed an autoprescan for field shimming and transmitter gain optimization. Then receiver gains (RGs) were manually adjusted if needed to retain the same RG values for all MRSI scans. The protocol was approved by the

Institutional Review Board of the New York State Psychiatric Institute. Written informed consent was obtained from each human participant.

2.4. Data Processing

2.4.1. Combination of Multichannel Data. The k -space MRSI data from individual coil elements were transformed to the image domain using a 2D spatial Fourier transform after spatial filtering that used a Hamming window function. The data from the multiple coil elements were combined using the following procedures.

(1) *Water Signal Removal.* We used a matrix-pencil-method-based procedure [16] to decompose the signal, identify water components by their frequencies, and remove them from the signal [16]. This method was able to remove water signal almost completely (>98%) without interfering with the metabolite signals of interest.

(2) *Removal of Corrupted Points.* We replaced the first 6 echo data points [17], which were corrupted by the activation of the analogue-to-digit converter, with 6 extrapolated points derived from signal parameters that were estimated from the uncorrupted data points using the matrix pencil method.

(3) *Data Apodization.* We next multiplied the cleaned echo data by a Gaussian function, $G(t) = e^{-\beta t^2}$, to suppress noise and reduce baseline distortion, albeit at the expense of line broadening. The line broadening was 10 Hz for phantom data and 4 Hz for *in vivo* data.

(4) *Phase Alignment.* We eliminated voxelwise phase differences in echoes from individual coil elements by subtracting the phases at the top of their echoes.

(5) *Weighted Summation.* We summed the phase-aligned echoes using weighting factors that were proportional to the echo amplitudes and inversely proportional to the noise levels of the coil elements. Noise levels were determined by measuring the standard deviations of the data points in the signal-free regions of the frequency domain spectra of a phantom.

2.4.2. Spectral Fitting. We quantified the spectral components using the following general model function to fit the echoes:

$$S(t) = \sum_{m=1}^M A_m e^{i(2\pi f_m t + \varphi_m)} e^{-\alpha_m |t| - \beta_m t^2}, \quad (1)$$

where A_m , f_m , φ_m , and α_m represent the amplitude, frequency, phase, and Lorentzian decay of peak m , respectively; β_m is the Gaussian decay. Note that the t runs from $-dt \cdot N/2$ to $dt \cdot (N/2 - 1)$, where dt is dwell time and N is the number of data points in the echo. Note also that when fitting the spectrum with a pure Lorentzian model, we set β_m to be zeros; when

TABLE 1: Comparisons of signal amplitudes acquired on a phantom with or without spacing between slices. Means and SD were calculated using (2). n is the number of selected voxels in the 4th slice. The TR was 2.3 s for tests 1 and 2, and 3.0 s for test 3. The labeled concentrations for NAA, Cr, and Cho were 12.5, 10.0, and 3.0 mM, respectively.

Test	NAA (mean \pm SD)	Cr (mean \pm SD)	Cho (mean \pm SD)
1 ($n = 47$)	1.26% \pm 4.16%	1.71% \pm 3.23%	1.05% \pm 5.82%
2 ($n = 39$)	0.51% \pm 4.22%	0.07% \pm 6.15%	-0.47% \pm 4.17%
3 ($n = 49$)	0.47% \pm 3.29%	-1.73% \pm 4.43%	-0.84% \pm 4.82%

fitting the spectrum with pure Gaussian model, we set the α_m to be zeros; when fitting the spectrum with a Voigtian model, we set the β_m to be the same for all M peaks.

Further notes are warranted for the process of spectral fitting. The signal parameters in (1) were determined using a nonlinear least squares fitting routine in Matlab (2007a, The MathWorks, Natick, Massachusetts). To ensure that these parameters were real numbers, the real parts and imaginary parts of both the model function in (1) and the measured echo signal were concatenated, respectively, to form real-number series. It is important to determine accurately the initial values for A_m , f_m , and α_m , as well as the number of peaks, to ensure the robustness of the fitting. They were estimated from the magnitude spectrum obtained by zero-padding the echo to 4096 points and then performing FFT, whereas the global initial phase was obtained by the phase of the top point of the echo.

2.4.3. Comparing Data Acquired with or without Slice Spacing.

The influence of acquiring contiguous slices on signal intensity was determined by comparing signal amplitudes from the two scans with or without spacing between slices. We selected voxels from a region within the brain in the 4th slice whose location was the same for the scans with or without spacing. The amplitudes of the m -th signal components (peak areas in the frequency domain), A_m , were used to calculate the relative differences of the signals obtained with or without spacing

$$d_m = \frac{(A_{m,w\cdot sp} - A_{m,wo\cdot sp})}{A_{m,w\cdot sp}}. \quad (2)$$

We used the means and standard deviations (S.D.) of the d 's to evaluate the reductions in signal caused by the ‘‘crosstalk’’ introduced between slices when spacing was removed.

3. Results

3.1. Comparing Signals with and without Spacing between Slices. Tables 1 and 2 show the relative differences of the amplitudes, calculated using (2), of the phantom and *in vivo* MRS signals acquired with and without spacing between slices. The phantom data did not show significant differences in these values (Table 1). The mean values of the differences in amplitudes (or the signal reduction) caused by the

TABLE 2: Comparisons of signal amplitudes acquired on a human subject with or without spacing between slices. Means and SD were calculated using (2). n is the number of selected voxels in the 4th slice. The TR was 2.3 s for tests 1 and 2, and 3.0 s for test 3.

Test	NAA (mean \pm SD)	Cr (mean \pm SD)	Cho (mean \pm SD)
1 ($n = 34$)	2.50% \pm 4.47%	7.30% \pm 4.09%	2.94% \pm 9.71%
2 ($n = 35$)	7.33% \pm 4.24%	1.56% \pm 6.23%	5.63% \pm 6.64%
3 ($n = 38$)	2.80% \pm 5.95%	3.04% \pm 8.07%	2.52% \pm 10.0%

TABLE 3: Relative errors (%) of spectral fitting of the simulated signal in the frequency domain, caused by truncation effects and by wiggles, respectively.

Lineshape	Original 256 points			Zero-padding to 1024 points		
	NAA	Cr	Ch	NAA	Cr	Ch
Lorentzian	4.72	2.04	4.34	4.29	3.23	0.82
Gaussian	4.06	0.19	1.72	4.38	2.10	0.44
Voigtian	8.45	15.31	6.07	7.00	9.33	0.59

‘‘crosstalk’’ in the *in vivo* data was $<7.5\%$ for the first two tests, when $TR = 2.3$ s. On the average, the signal reduction was smaller for the third test when $TR = 3.0$ s (Table 2).

3.2. Comparison of Spectral Fitting in the Time and Frequency Domains. Errors can be as large as 4% when fitting the noise-free original data in the frequency domain with a Lorentzian or Gaussian lineshape, or up to 15% for fitting data with a Voigtian lineshape (Table 3). These errors are the consequence of low spectral resolution and a (Voigtian) model mismatch in the frequency domain. When fitting the zero-padded signals (from the original 256 to 1024 points) in the frequency domain, errors were of the same order, whereas errors were reduced to zero when fitting the spectrum with the original 1024 points. This finding indicates that errors using padded signals are caused by wiggles (Figure 2). In contrast, when fitting the original 256 points of noise-free signals in the time domain, the signals can be perfectly recovered, regardless of whether the line shapes are Lorentzian, Gaussian, or Voigtian. Monte Carlo simulations with 400 noise realizations added to the 256 points time domain signals revealed that the estimated amplitudes of NAA, Cr, and Ch approximated the true values for all three lineshapes. The SDs of the Lorentzian and Voigtian lineshapes were less than 2 times the Cramér-Rao Lower Bounds (CRLBs), whereas the SDs of the Gaussian line shapes were approximately 1.5 times the CRLB (Table 4).

3.3. Combination of Multichannel Signals. The phase differences caused by the differing positions of coil elements were eliminated, yielding perfect alignment of the array signals in the frequency domain and thereby enhancing SNR (Figure 3).

3.4. Spectral Fitting and Metabolite Maps. The performance of spectral fitting of the data of 256 points in the time domain

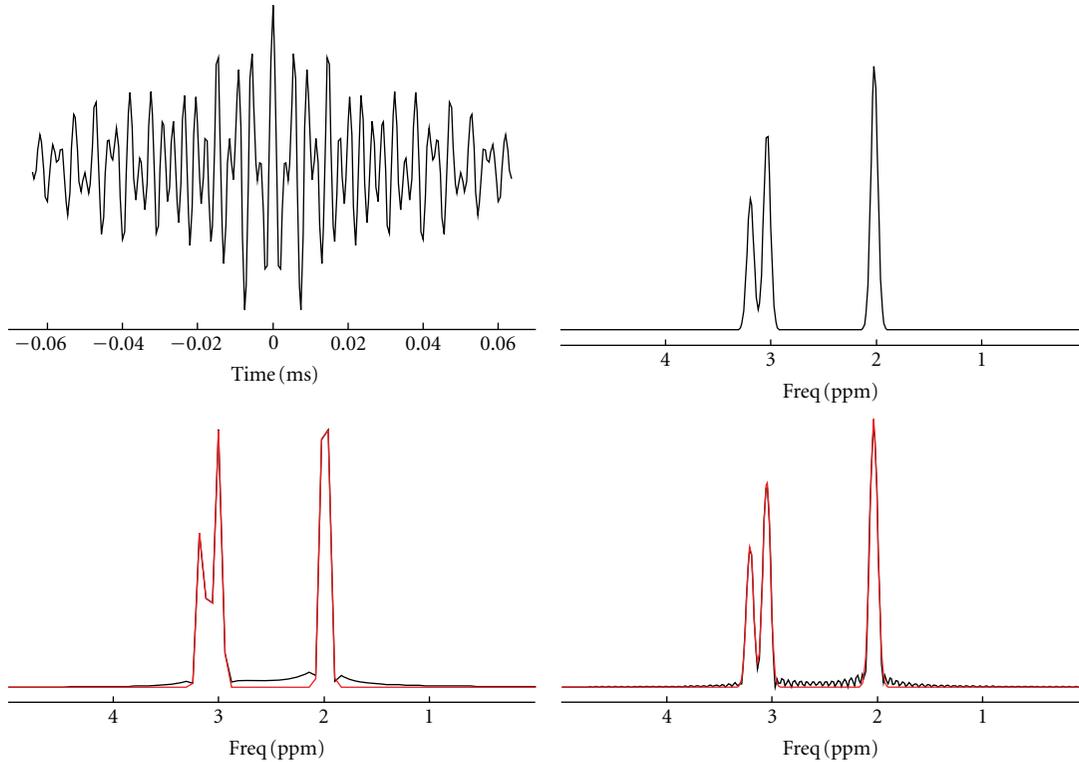


FIGURE 2: Simulated echo with 256 data points and its spectrum obtained with 1024 data points (upper). The time and frequency domain fitting algorithms can fully recover each of them. However, when fitting the spectrum with 256 data points, or when the spectrum was zero-filled from 256 to 1024 data points, sizeable errors are evident as a consequence of either low spectral resolution or truncation effects, which manifest as wiggles in the spectrum after zero-filling (lower).

TABLE 4: Results of Monte Carlo study of the estimated amplitudes (mean \pm SD). The true values of NAA, Cr, and Ch are 48, 36, and 24, respectively; the CRLBs are 0.1574, 0.1288, and 0.1858 for Lorentzian, Gaussian, and Voigtian, respectively.

Lineshape	NAA	Cr	Ch
Lorentzian	47.99 \pm 0.27	35.99 \pm 0.31	24.02 \pm 0.30
Gaussian	48.00 \pm 0.18	36.00 \pm 0.19	24.00 \pm 0.19
Voigtian	48.00 \pm 0.37	36.00 \pm 0.40	24.00 \pm 0.38

is further demonstrated using *in vivo* data. An example of the spectral fitting is shown in Figure 4, in the form of an absolute spectrum. The whole brain MRSI of NAA is shown in Figure 5, overlaid on their localizer images. Similar results were obtained for the other two subjects.

4. Discussion

We reported herein improvements of the well-known and widely used multi-section MRSI technique. The modified pulse sequence has several important advantages over its standard implementation: (1) it eliminates spacing between slices; (2) it allows increase in the number of slices (in our study from 4 to 7) without the expense of increasing scan time; (3) it permits a reduction in slice thickness and therefore improves spatial resolution, which can be achieved

without the expense of reducing overall volume coverage because of the increased number of slices that are acquired; (4) it employs a multichannel coil for data acquisition to improve SNR. It also employs fitting of the severely truncated full echo in the time domain. These improvements make the multi-section MRSI technique more diagnostically valuable.

Compared with the performance of the original MRSI pulse sequence, each modification individually can have its own unique advantages, limitations, and challenges. Spacing between slices, for example, was introduced in the original implementation of the sequence to avoid “crosstalk” between adjacent slices, a phenomenon of signal interference caused by the imperfect slice profiles in which edges are not clearcut but instead interlace with one another. “Crosstalk” can cause signal loss and thus reduce SNR. Slice spacing, however, comes at the expense of information loss from the volumes between slices that are not imaged, which can be as large as 30% of the total MRSI imaging volume. This reduced volume of coverage also reduces overall spatial resolution. The reduction in signal caused by eliminating spacing between slices was less than 10% *in vivo* using our modified MRSI sequence (Table 2). This loss of signal would be regarded as substantial if the SNR of the original signal was low, but it is in fact an inconsequential loss when using an 8-channel multichannel coil array, which typically doubles SNR compared with use of a standard quadrature coil [8, 11, 12]. Therefore, even with a 10% signal reduction

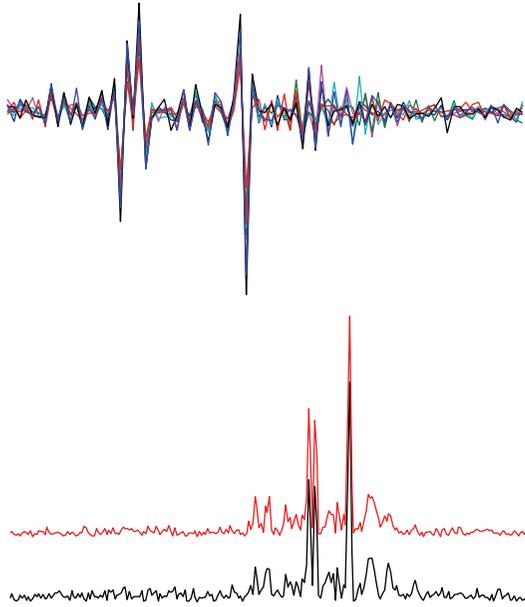


FIGURE 3: Examples of the combination of *in vivo* signals from a multichannel coil array. Upper: phase alignment displays the real parts of the spectra without phase correction. Lower: a combined spectrum (red) and a channel spectrum with the highest SNR, shown in absolute mode. The SNRs of the combined spectrum and the channel spectrum with the highest SNR were 77.8 and 46.1, respectively.

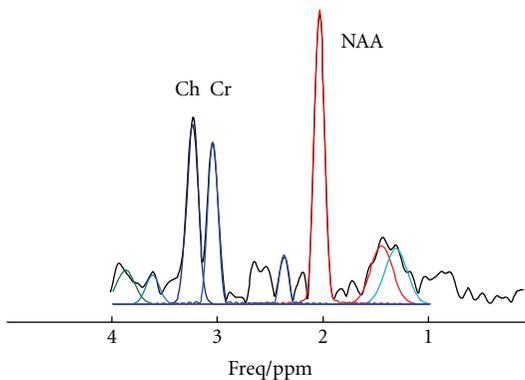


FIGURE 4: An example of spectral fitting of the *in vivo* data in the time domain for 256 data points, displayed in the frequency domain and absolute mode. Shown in colors are the individual fitted spectral lines overlaid on the measured spectrum in black. The excellent fitting supports the merit of spectral fitting in the time domain.

caused by “crosstalk” from contiguous slices, the SNR using a multichannel coil in conjunction with our modified sequence is still much higher than the SNR when using a standard quadrature head coil without “crosstalk” (Figure 4). We note, in fact, that increasing the number of slices from 4 to 7 can actually help to reduce the “crosstalk” across slices, because interleaved slice excitations are separated temporally by at most one interval if the number of slices is 4 (i.e., 1-3-2-4), whereas with 7 slices, the temporal interleave can be 2

or 3 intervals for adjacent slices (e.g., 1-3-5-7-2-4-6, where slices 1 and 2 are temporally separated by 3 intervals, slices 4 and 5 are separated by 2 intervals, and so on). In addition, the reduction in slice thickness theoretically would lead directly to a proportional reduction in SNR when shimming is perfect. In reality, however, and especially in regions near air-tissue interfaces, the effect of reduced volumes in each slice on SNR is not linear because smaller volumes come with narrower line widths. Consequently, we did not observe significant signal drop-out in regions of air-tissue interface within the lower frontal lobe (Figure 5). Thus, the advantages that our modified multi-section MRSI sequence provides are the whole brain coverage with contiguous slices, potentially improved spatial resolution, and improved spectral line width.

The reduction in number of data points in the echo, which is the core of the current modification, poses severe challenges for spectral fitting and spectral quantification. Computer simulation (Table 3) showed that fitting a spectrum with low spectral resolution in the frequency domain can produce sizeable errors. Conversely, increasing the spectral resolution can improve the accuracy of spectral fitting in the frequency domain. When the number of data points in the original echo was 1024, the frequency domain fitting algorithm perfectly recovered the spectrum. However, significant errors remained if the spectrum was obtained by zero-filling the 256 data points to 1024 (Table 3), reflecting the detrimental effects of truncation or wiggle artifacts on spectral fitting. Spectral fitting in the time domain, in contrast, accurately fits the echo using 256 data points, suggesting that time domain spectral fitting is preferable for signals with fewer data points, a possibility that Monte Carlo simulation of the signal (Table 4) and spectral fitting of the *in vivo* data (Figures 4 and 5) are verified.

Our implementation of whole brain metabolite mapping using a multiple 2D MRSI sequence also affords distinct advantages over 3D MRSI, in which the number of phase encoding (PE) steps in the 3rd dimension is typically 8 [18, 19]. This small number of data points entails pronounced, long-range signal contamination across slices because of the effects of the point spread function when reconstructing the slice data directly using FFT. Spatial filtering with Hamming, Hanning, or Kaiser window functions must be applied prior to FFT to suppress this signal contamination, but at the expense of increasing the amount of signal bleeding between adjacent slices. Therefore, the effective slice thickness of 3D MRSI is approximately 1.4 times that of the nominal slice thickness, which significantly degrades spatial resolution. The effective thickness of slices using multiple 2D MRSI, on the other hand, is close to its nominal value, despite the fact that the slice profile is not ideal and slight “crosstalk” is present between slices. Another limitation of 3D MRSI using PRESS localization is that the first and last slices cannot be used, reducing the effective number of slices by 2 [18, 19]. Our pulse sequence, in contrast, provides high-quality MRSI images in all 7 slices (Figure 5). 3D MRSI has a distinct advantage over multi-section MRSI, however, in that the location of slices in 3D MRSI can be shifted to specific

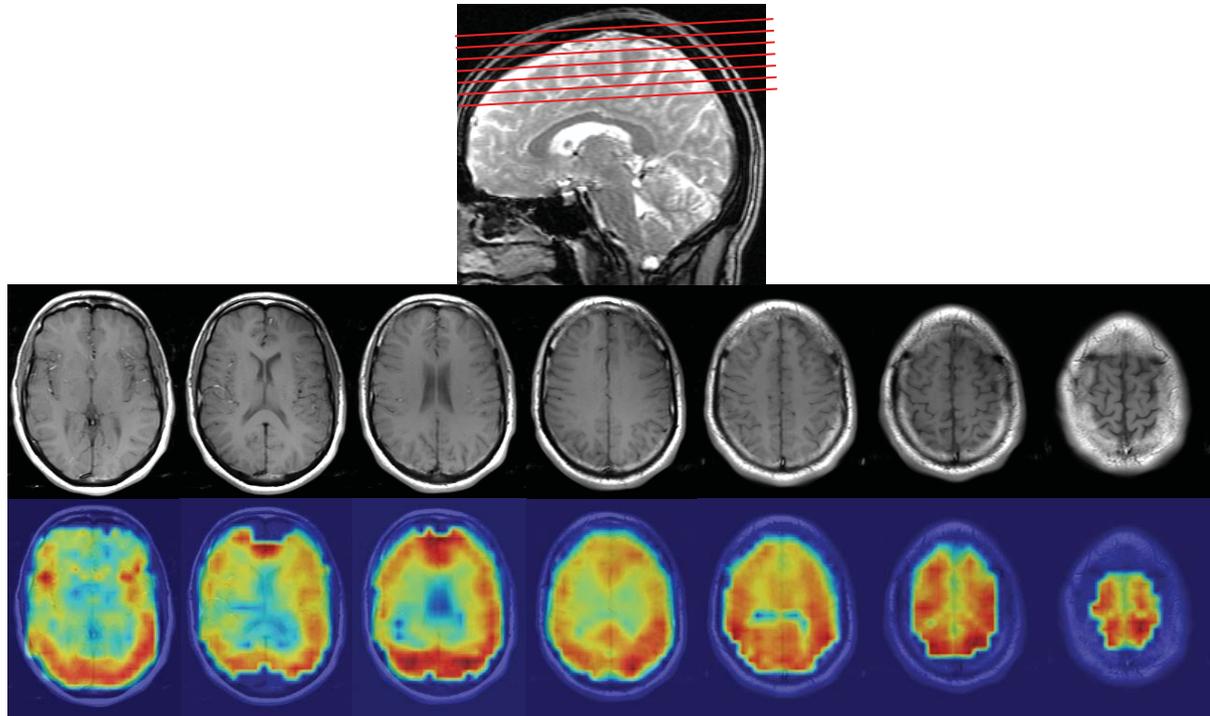


FIGURE 5: Slice prescription (top), localizer images (middle), and NAA images overlaid on localizers (bottom). The slices are contiguous and cover a large volume of the brain.

anatomical regions of interest by employing a phase shift of the Fourier transform.

Comparing signal losses in phantom and *in vivo* MRSI data in the absence of spacing between slices showed that signal loss caused by “crosstalk” is related to the slice profile, the spatial and temporal separations between successive excitations, and the longitudinal relaxation times of the subjects. Our experimental procedures ensured that the spatial and temporal separations of excitations were identical for the phantom and *in vivo* scans. The longitudinal relaxation times of molecules in the phantom were significantly longer than those in the human brain, and therefore comparatively larger signal loss could be expected if all other conditions remained the same. The fact that signal loss was negligible in the phantom (Table 1) but was approximately 10% in the human brain (Table 2) when spacing was eliminated as compared with 4 mm spacing can be attributed to a better slice profile in the phantom. Comparing spectral fitting in the time domain with fitting in the frequency domain has been of long-standing interest in MRS [20–25]. As the time domain and frequency domain signals are related by the Fourier transform, they theoretically have the same information content. However, fitting in the time and frequency domains may have their unique advantages and disadvantages depending on the properties of the signal, such as noise level and phase or baseline distortions [20–25]. Time domain methods are preferable in the presence of distortions in the measured signals, including truncation [20, 23]. Our computer simulation provided a numerical example for the detrimental effects of truncation in the frequency domain

and showed that fitting in the time domain is immune to those effects.

In conclusion, we have presented a realization of extended brain metabolite mapping using a multiple 2D MRSI pulse sequence in conjunction with use of a multichannel RF coil and spectral fitting in the time domain. These combined techniques have permitted an increase in the number of slices from 4 to 7, without sacrificing scan time or SNR. The extended brain coverage, reduced slice thickness, and increased SNR can potentially make the sequence more clinically valuable.

Grant Support

This study was funded in part by NIMH grants K02 74677, MH089582, MH068318, 1P50MH090966, NIEHS grant ES015579, and NIDA grants DA027100, and DA017820.

Acknowledgments

The authors would like to thank Drs. Jan Willem van der Veen and Dikoma Shungu for their help in updating the multi-section MRSI sequence to allow the acquisition of more than 4 slices. They would like to thank the anonymous reviewers for their insightful comments and suggestions that help improve the paper.

References

- [1] P. Bottomley, *Selective Volume Method for Performing Localized NMR Spectroscopy*, 1984.
- [2] J. Frahm, H. Bruhn, M. L. Gyngell, K. D. Merboldt, W. Hanicke, and R. Sauter, "Localized high-resolution proton NMR spectroscopy using stimulated echoes: initial applications to human brain in vivo," *Magnetic Resonance in Medicine*, vol. 9, no. 1, pp. 79–93, 1989.
- [3] P. B. Barker and D. D. M. Lin, "In vivo proton MR spectroscopy of the human brain," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 49, no. 2, pp. 99–128, 2006.
- [4] S. Posse, C. DeCarli, and D. Le Bihan, "Three-dimensional echo-planar MR spectroscopic imaging at short echo times in the human brain," *Radiology*, vol. 192, no. 3, pp. 733–738, 1994.
- [5] E. Adalsteinsson, P. Irrazabal, S. Topp, C. Meyer, A. Macovski, and D. M. Spielman, "Volumetric spectroscopic imaging with spiral-based k-space trajectories," *Magnetic Resonance in Medicine*, vol. 39, no. 6, pp. 889–898, 1998.
- [6] M. L. Zierhut, E. Ozturk-Isik, A. P. Chen, I. Park, D. B. Vigneron, and S. J. Nelson, "1H spectroscopic imaging of human brain at 3 Tesla: comparison of fast three-dimensional magnetic resonance spectroscopic imaging techniques," *Journal of Magnetic Resonance Imaging*, vol. 30, no. 3, pp. 473–480, 2009.
- [7] J. H. Duyn, J. Gillen, G. Sobering, P. C. M. Van Zijl, and C. T. W. Moonen, "Multisection proton MR spectroscopic imaging of the brain," *Radiology*, vol. 188, no. 1, pp. 277–282, 1993.
- [8] Z. Dong and B. Peterson, "The rapid and automatic combination of proton MRSI data using multi-channel coils without water suppression," *Magnetic Resonance Imaging*, vol. 25, no. 8, pp. 1148–1154, 2007.
- [9] T. Schäffter, P. Börnert, C. Leussler, I. C. Carlsen, and D. Leibfritz, "Fast 1H spectroscopic imaging using a multi-element head-coil array," *Magnetic Resonance in Medicine*, vol. 40, no. 2, pp. 185–193, 1998.
- [10] S. M. Wright and L. L. Wald, "Theory and application of array coils in MR spectroscopy," *NMR in Biomedicine*, vol. 10, no. 8, pp. 394–410, 1997.
- [11] M. A. Brown, "Time-domain combination of MR spectroscopy data acquired using phased-array coils," *Magnetic Resonance in Medicine*, vol. 52, no. 5, pp. 1207–1213, 2004.
- [12] T. Prock, D. J. Collins, A. S. K. Dzik-Jurasz, and M. O. Leach, "An algorithm for the optimum combination of data from arbitrary magnetic resonance phased array probes," *Physics in Medicine and Biology*, vol. 47, no. 2, pp. N39–N46, 2002.
- [13] U. Dydak, M. Weiger, K. P. Pruessmann, D. Meier, and P. Boesiger, "Sensitivity-encoded spectroscopic imaging," *Magnetic Resonance in Medicine*, vol. 46, no. 4, pp. 713–722, 2001.
- [14] D. Xu, A. P. Chen, C. Cunningham, J. A. Osorio, S. J. Nelson, and D. B. Vigneron, "Spectroscopic imaging of the brain with phased-array coils at 3.0 T," *Magnetic Resonance Imaging*, vol. 24, no. 1, pp. 69–74, 2006.
- [15] J. Keeler, "Elimination of truncation artifacts from NMR spectra. Application to carbon-13 multiplicity determination by two-dimensional spectroscopy," *Journal of Magnetic Resonance*, vol. 56, no. 3, pp. 463–470, 1984.
- [16] Z. Dong, W. Dreher, and D. Leibfritz, "Toward quantitative short-echo-time in vivo proton MR spectroscopy without water suppression," *Magnetic Resonance in Medicine*, vol. 55, no. 6, pp. 1441–1446, 2006.
- [17] H. Barkhuijsen, R. de Beer, and W. M. M. J. Bovee, "Communication: application of linear prediction and singular value decomposition (LPSVD) to determine NMR frequencies and intensities from the FID," *Magnetic Resonance in Medicine*, vol. 2, no. 1, pp. 86–89, 1985.
- [18] G. Goelman, S. Liu, D. Hess, and O. Gonen, "Optimizing the efficiency of high-field multivoxel spectroscopic imaging by multiplexing in space and time," *Magnetic Resonance in Medicine*, vol. 56, no. 1, pp. 34–40, 2006.
- [19] R. Otazo, S. Y. Tsai, F. H. Lin, and S. Posse, "Accelerated short-TE 3D proton echo-planar spectroscopic imaging using 2D-SENSE with a 32-channel array coil," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1107–1116, 2007.
- [20] A. Van Den Boogaart, M. Ala-Korpela, F. A. Howe, L. M. Rodrigues, M. Stubbs, and J. R. Griffiths, "Magnetic resonance spectroscopy data analysis: time or frequency domain?" *Magnetic Resonance Materials in Physics, Biology, and Medicine*, vol. 2, no. 3, pp. 479–482, 1994.
- [21] C. Elster, A. Link, F. Schubert, F. Seifert, M. Walzel, and H. Rinneberg, "Quantitative MRS: comparison of time domain and time domain frequency domain methods using a novel test procedure," *Magnetic Resonance Imaging*, vol. 18, no. 5, pp. 597–606, 2000.
- [22] J. B. Poulet, D. M. Sima, and S. Van Huffel, "MRS signal quantitation: a review of time- and frequency-domain methods," *Journal of Magnetic Resonance*, vol. 195, no. 2, pp. 134–144, 2008.
- [23] R. Bartha, D. J. Drost, and P. C. Williamson, "Factors affecting the quantification of short echo in-vivo 1H MR spectra: prior knowledge, peak elimination, and filtering," *NMR in Biomedicine*, vol. 12, no. 4, pp. 205–216, 1999.
- [24] S. Mierisova and M. Ala-Korpela, "MR spectroscopy quantitation: a review of frequency domain methods," *NMR in Biomedicine*, vol. 14, no. 4, pp. 247–259, 2001.
- [25] L. Vanhamme, T. Sundin, P. Van Hecke, and S. Van Huffel, "MR spectroscopy quantitation: a review of time-domain methods," *NMR in Biomedicine*, vol. 14, no. 4, pp. 233–246, 2001.

Research Article

Label Fusion Strategy Selection

Nicolas Robitaille¹ and Simon Duchesne^{1,2}

¹ Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec, 2601, Chemin de la Canardière, QC, Canada G1J 2G3

² Radiology Department, Faculty of Medicine, Laval University, Pavillon Ferdinand-Vandry, 1050, Avenue de la médecine, QC, Canada G1V 0A6

Correspondence should be addressed to Nicolas Robitaille, nicolas.robitaille@crulrg.ulaval.ca

Received 26 April 2011; Revised 8 September 2011; Accepted 25 September 2011

Academic Editor: Alayar Kangarlu

Copyright © 2012 N. Robitaille and S. Duchesne. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Label fusion is used in medical image segmentation to combine several different labels of the same entity into a single discrete label, potentially more accurate, with respect to the exact, sought segmentation, than the best input element. Using simulated data, we compared three existing label fusion techniques—STAPLE, Voting, and Shape-Based Averaging (SBA)—and observed that none could be considered superior depending on the dissimilarity between the input elements. We thus developed an empirical, hybrid technique called SVS, which selects the most appropriate technique to apply based on this dissimilarity. We evaluated the label fusion strategies on two- and three-dimensional simulated data and showed that SVS is superior to any of the three existing methods examined. On real data, we used SVS to perform fusions of 10 segmentations of the hippocampus and amygdala in 78 subjects from the ICBM dataset. SVS selected SBA in almost all cases, which was the most appropriate method overall.

1. Introduction

Label fusion is a process used in medical image segmentation. Its aim is to produce a single, discrete element or *label* from a combination of multiple independent inputs. The merged result is potentially more accurate, with respect to the exact, sought segmentation, than each individual input due to the reduction of uncorrelated errors. Labels can be obtained by combining inputs from different raters or automated segmentations [1, 2].

A long-term goal of our research program is to obtain accurate, automated segmentations of neuroanatomical structures, primarily the hippocampus (HC). Our primary motivation stems from our work in Alzheimer's disease, for which HC volume and atrophy measurements are putative disease markers (see reviews in [3–6]). Of the multiple HC segmentation approaches available (see [7] for review), novel template-based paradigms propose the use of template libraries [8]. In such approaches, a single label is found by combining multiple individually segmented HC through label fusion [2].

To reach our goal, we thus decided to investigate different fusion processes. To suit our research context, we restricted

our analysis to techniques that depend solely on given input labels. We disregarded techniques that depend on intensity images [9, 10], since these images may sometimes be unavailable or noisy. We also ignored techniques that depend on object-specific training, i.e. that have geometric or topological prior.

Our first objective was to characterize applicable label fusion strategies. The first approach is the Vote method (or sum rule), which has been widely used and described by virtue of its simplicity [1, 9, 11–13]. The second is also a well-known technique called Simultaneous Truth and Performance Level Estimation (STAPLE), initially proposed by Warfield et al. [14, 15], and used in a variety of studies [9, 16]. The third approach is referred to as Shape-Based Averaging (SBA), which incorporates spatial information [17].

While testing the implementations of these three approaches on simulated data, we observed that the technique with a result closest to the ground truth was not the same depending on the dissimilarity between raters' input labels, as detailed below. Therefore, the second objective of our study was to propose an empirical, hybrid STAPLE-Vote-SBA (SVS) technique that automatically selects the right label fusion approach based on this dissimilarity.

We report results of comparison tests on the four label fusion methods for simulated two-dimensional (2D) and three-dimensional (3D) data as well as HC and amygdala (AG) labels obtained from magnetic resonance images (MRI). All images used in this study were binary. For the real data, we performed label fusion on HC and AG independently.

2. Materials and Methods

2.1. Mathematical Notation. Our mathematical notation is as follows. We consider an image of N pixels or voxels ($x = 1, 2, \dots, N$) for which K raters ($k = 1, 2, \dots, K$) each produces a binary label segmentation e_k . To each element of e_k , i.e. each pixel/voxel x , is assigned a label $i(e_k(x) = i)$ equal to 0 or 1, for background and segmented object, respectively. A decision matrix E is formed with all the e_k vectors, $E = [e_1 e_2 \dots e_K]$ with size $N \times K$, and fed to a label fusion algorithm to obtain an estimate of the true segmentation T .

2.2. Data. For evaluating the performance of SVS with respect to STAPLE, Vote and SBA, our data consisted of 2D and 3D simulated data as well as real data.

2.2.1. Two-Dimensional (2D) Simulated Data. We created two simulated 2D data sets: one for training SVS and one for testing the label fusion approaches. The SVS version trained with 2D data is hereafter referred to as SVS-2D.

The data consisted of multiple binary images created from a ground-truth object, shown in Figure 1(a), which was an ellipse geometry defined by eight control points interpolated with cubic splines.

We generated individual, simulated rater images by moving the control points of the ground-truth ellipse and reinterpolating with cubic splines. We moved the control points in random directions, following a uniform distribution, with random distances from their original coordinates. The random distance followed a normal distribution of zero mean with a standard deviation adjusted so that it could be modified by a normalized deformation factor f_σ (between 0 and 1) to create images with a relative difference area v_D ranging from 0% to 50%, where v_D is given by

$$v_D = \frac{v_{k|T}}{V_{\text{TRUTH}}}, \quad (1)$$

where V_{TRUTH} corresponds to the area in pixels of the ground-truth ellipse. $v_{k|T}$ represents the number of pixels in the image that are different between decision e_k of rater k and the ground truth T :

$$v_{k|T} = \#\{x \mid e_k(x) \neq T(x)\}. \quad (2)$$

In other words, $v_{k|T}$ is the total number of false positives and false negatives with respect to T . Figures 1(b) and 1(c) show two rater images corresponding to v_D values of 25% and 50%, respectively.

For each of the training and testing sets, we created 625 label fusion tests, each consisting of 10 deformed images, for a total of 6,250 images in the training set and 6,250 different images in the testing set. Each test was created by varying f_σ of the test images according to a given Gaussian distribution. For each test, different mean and standard deviation were used for f_σ , ranging both from 0 to 1 with 25 linearly spaced points each, making a total of 625 Gaussian distributions, one for each test. Negative values of f_σ and values higher than 1 were clamped to 0 and 1, respectively. We performed the label fusion of the 10 deformed images in each of the 625 tests of the testing set.

2.2.2. Three-Dimensional (3D) Simulated Data. As for the 2D case, we created two simulated 3D sets: one for training SVS and one for testing the label fusion techniques. The SVS version trained with 3D data is hereafter referred as SVS-3D. An SVS version was also trained with the combination of 2D and 3D training sets. It is referred as SVS-2D&3D.

The 3D data consisted of binary volume images created from a ground-truth ellipsoid. To produce the ground truth, we first created a cubic regular grid volume. This volume was then warped along each axis by dividing each voxel coordinate by its corresponding ground-truth ellipsoid radius, creating a warped grid. By applying this warping transformation, the ellipsoidal space became a spherical space. A ground-truth sphere was created by regularly sampling the angles θ and ϕ in the spherical-coordinate space (r, θ, ϕ) , giving a set of 26 control points (r_c, θ_c, ϕ_c) .

To produce the ground-truth image, the control points were projected into a Cartesian space with the following axes: $x = \theta$, $y = \phi$, and $z = r$. We transformed the warped grid into spherical coordinates (r_g, θ_g, ϕ_g) and performed a cubic interpolation of (θ_g, ϕ_g) on (r_c, θ_c, ϕ_c) to find r^* at each point (θ_g, ϕ_g) . For each grid voxel, if $r_g < r^*$, the voxel was considered inside the sphere and was labeled accordingly. The warped grid (spherical space) was then unwarped into the regular grid (ellipsoidal space) to give the desired ground-truth ellipsoid image shown in Figure 1(d).

While appearing complex, this process in fact simplified the creation of the deformed ellipsoid images. We randomly moved the control points of the ground-truth sphere along r , modifying r_c , reinterpolated to find r^* for the warped grid, performed the labeling by thresholding (i.e. $r_g < r^*$), and unwarped the grid to obtain the deformed ellipsoid.

As for the 2D sets, the random distance followed a normal distribution of zero mean. The standard deviation was adjusted so that it could be modified by f_σ to create deformed ellipsoids with relative difference in volume, v_D (1), ranging between 0% and 50% with respect to the ground truth.

Figures 1(e) and 1(f) show two examples of deformed images with v_D of 25% and 50%, respectively. As for the 2D data, we produced a training set and a testing set, each consisting of 625 label fusion tests. Each test was created as previously described and comprised 10 deformed images. Each of the training and testing sets thus consisted of 6,250 images. We performed the label fusion of the 10 deformed images in each of the 625 tests of the testing set.

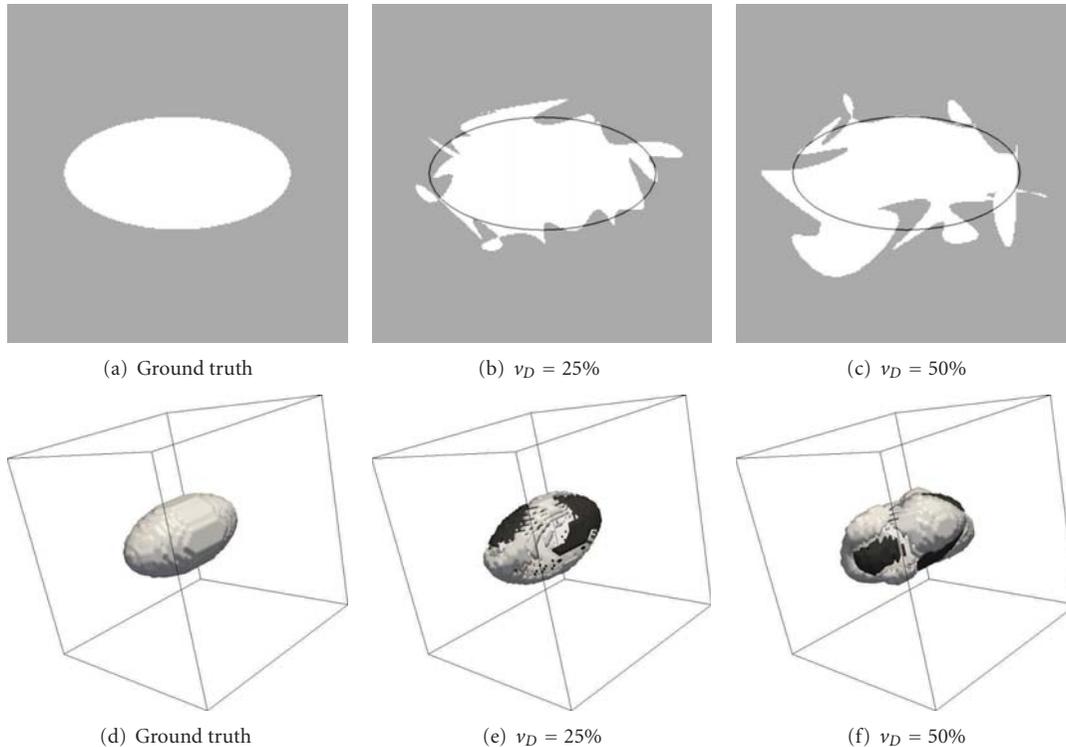


FIGURE 1: (a, b, and c) 2D and (d, e, and f) 3D simulated images showing the ground truth (a, d), and images with v_D of 25% (b, e) and 50% (c, f). White and black surfaces (e, f) represent, respectively, voxels added to or missing from the ground truth. In 2D, the ground truth was an ellipse geometry of radius 1 AU (arbitrary units) along the x -axis and 0.5 AU along the y -axis, consisting of eight control points, located at constantly separated angles, between which the ellipse was interpolated with cubic splines. We then mapped this geometry on a grid of 256×256 pixels between -1.5 and 1.5 AU along both the x - and y -axes. In 3D, the ground-truth image was an ellipsoid geometry of radius 1 AU along the x -axis and 0.5 AU along both the y - and z -axes, consisting of 26 control points. See text for construction details. The geometry was mapped in a grid of $64 \times 64 \times 64$ voxels between -1.5 and 1.5 AU along each of the three axes.

2.2.3. Real MRI Data. The real MRI data consisted of intensity images and segmented left and right HC and AG labels of 78 young, neurologically healthy subjects part of the ICBM database [18]. Subjects were scanned in Montréal (Québec, Canada) on a Philips Gyroscan 1.5T scanner (Philips Medical, Best, Netherlands) using a T1-weighted fast gradient echo sequence (sagittal acquisition, TR = 18 ms, TE = 10 ms, 1-mm^3 voxels, flip angle = 30°).

The ground truth consisted of left and right HC and AG manual labels, presented in a previous study [19], with a reported intraclass reliability coefficient of 0.900 and 0.925 for interrater and intrarater reliability, respectively.

The labels available for fusion were obtained using a template-based segmentation algorithm [2]. In this approach, each subject's image is compared in turn to a library of other such images; the 10 images with highest match (e.g., highest normalized mutual information) are selected and then nonlinearly aligned with the original subject image. Given that each image in the library has an associated label, inverse warping allows the transfer of label in the original subject's space, where they must be fused to provide a single object. In our dataset, we received 10 labels for each subject, obtained with this technique, for each of the four following

regions: left HC, right HC, left AG, and right AG. Label fusions were then performed independently for each region, giving a total of 312 label fusions ($78 \text{ subjects} \times 4 \text{ regions}$). We assessed the performance of the fusions using the manual segmentations as "ground truths".

2.3. Label Fusion Strategies. The next sections present the three existing label fusion strategies that we used in this study: STAPLE, Vote, and SBA. We implemented all label fusion methods, including SVS, in MATLAB (MathWorks, Natick, MA, USA).

It is important to note that all approaches were applied to the disputed pixels/voxels only. Pixels/voxels for which all the raters unanimously agreed on their label were not considered; the label was automatically assigned. Working with only disputed pixels/voxels speeded up computation for all methods and significantly improved the results given by STAPLE (see [16]).

2.3.1. STAPLE. STAPLE is an expectation-maximization (EM) algorithm that iteratively estimates (1) the true segmentation from the raters' performance (E-step) and (2) the raters' performance (sensitivity and specificity) from this

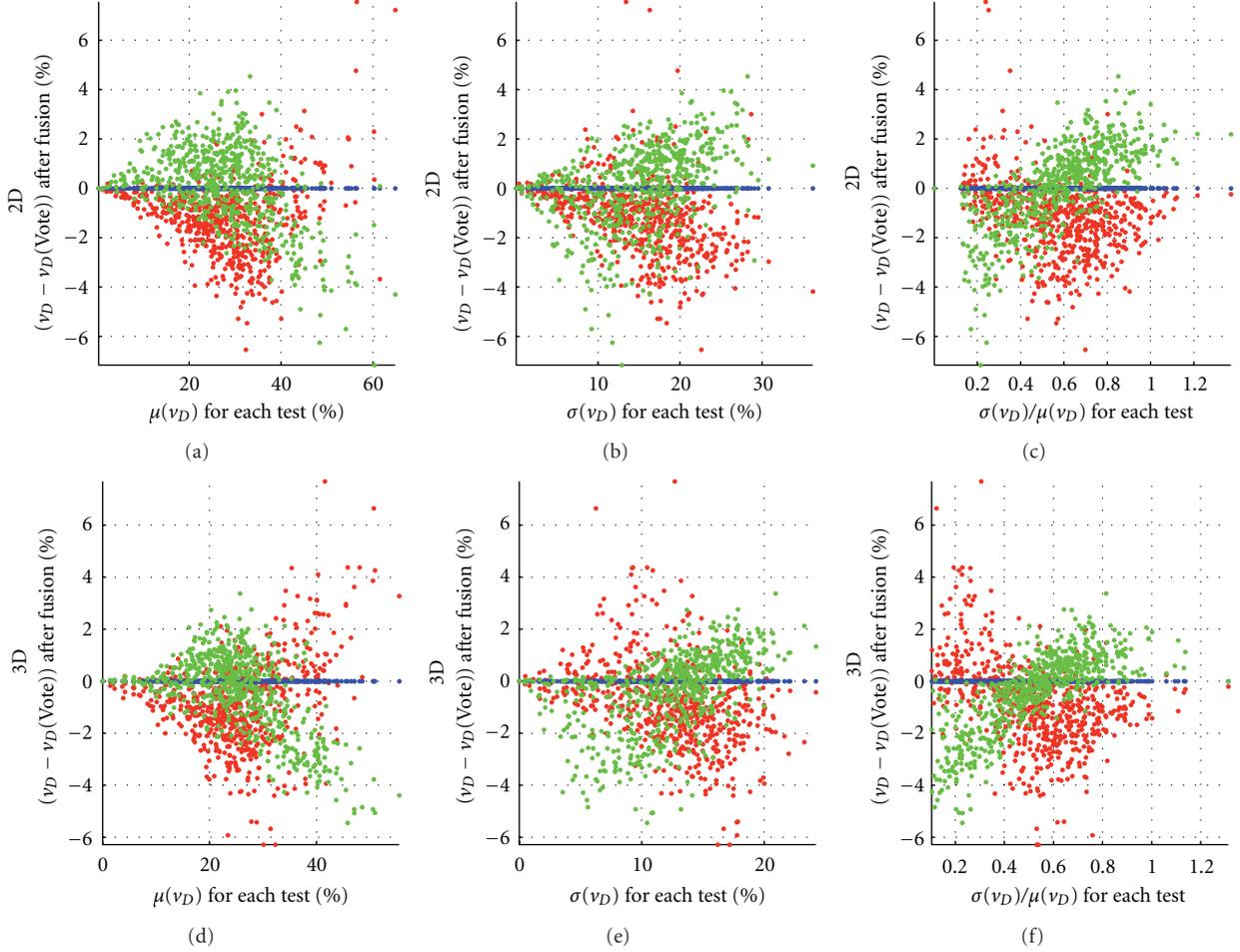


FIGURE 2: Scatter plots showing v_D centered on the Vote’s values, i.e. $(v_D - v_D(\text{Vote}))$, obtained after label fusion with STAPLE (red), Vote (blue), and SBA (green) of the 625 tests of the (a, b, and c) 2D and (d, e, and f) 3D training sets, with respect to each test’s (a, d) mean $(\mu(v_D))$, (b, e) standard deviation $(\sigma(v_D))$, and (c, f) coefficient of variation $(\sigma(v_D)/\mu(v_D))$ of v_D , calculated over the input labels for each test. The centered v_D corresponds to v_D minus the v_D evaluated for Vote. We note that $\sigma(v_D)/\mu(v_D)$ better discriminates the label fusion methods than $\sigma(v_D)$.

true segmentation estimate (M-step). We implemented STAPLE following the mathematical description in [20].

2.3.2. Vote. The Vote method consists of summing for each pixel/voxel x and label i , the occurrences of label i among the raters, and assigning the most occurring label to x .

2.3.3. SBA. SBA is a voting scheme where each vote is weighted by the signed Euclidean distance computed for each input label. In this study, SBA is the only method that incorporates spatial information in the label fusion process. We implemented this method following the mathematical description in [17].

2.4. Label Fusion Strategy Selection: SVS. SVS is a strategy that selects the most appropriate method among STAPLE, Vote, and SBA, based solely on the input labels and their

dissimilarity. We point out that SVS is not limited to these three label fusion methods. It could easily be extended to include further methods.

2.4.1. Experimental Observations. We developed SVS after observing, during our simulations, that the performance of STAPLE, Vote, and SBA was dependent on the distribution of v_D in the input labels of each label fusion test. This can be observed in the scatter plots of Figure 2 obtained for the 2D (a, b, and c) and 3D training sets (d, e, and f). The scatter plots show v_D centered on the Vote’s values, i.e. $(v_D - v_D(\text{Vote}))$, after label fusions performed with STAPLE (red), Vote (blue), and SBA (green), as a function of the mean $\mu(v_D)$ (a, d), standard deviation $\sigma(v_D)$ (b, e), and coefficient of variation $\sigma(v_D)/\mu(v_D)$ (c, f) of v_D , calculated over the input labels for each test.

We note that $\sigma(v_D)$ and $\sigma(v_D)/\mu(v_D)$ give an idea of how differently the raters perform between themselves, while

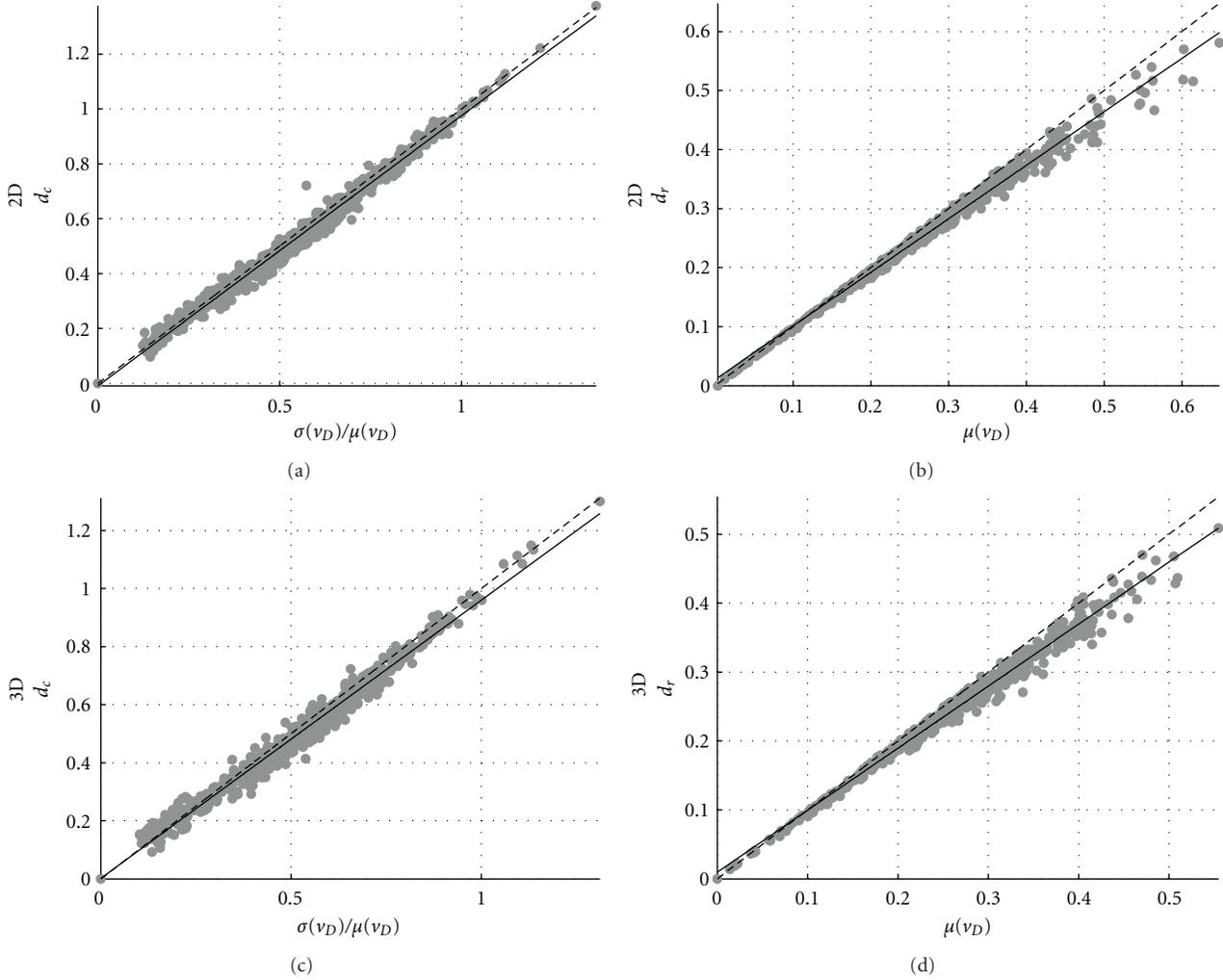


FIGURE 3: (a, c) d_c versus $\sigma(v_D)/\mu(v_D)$ and (b, d) d_r versus $\mu(v_D)$ for all the 625 tests of the (a, b) 2D and (c, d) 3D training sets. A linear fit was performed for both scatter plot types (continuous line) showing the quasi-one-to-one relationship between the theoretical values $\sigma(v_D)/\mu(v_D)$, $\mu(v_D)$ and their estimates d_c and d_r . The one-to-one relationship is represented by the dashed line.

$\mu(v_D)$ measures how bad the raters are overall. These measures thus describe, in a way, the dissimilarity in the raters' input labels.

As can be seen, none of STAPLE, Vote, and SBA can be considered superior to the others. The choice of the best method seems to depend on the distribution of v_D . For low values of $\sigma(v_D)/\mu(v_D)$, which better discriminates the label fusion methods than $\sigma(v_D)$, SBA seems better (i.e. with lower values of v_D after label fusion), while, for higher values, STAPLE would be a better choice. Focusing on the results with respect to $\mu(v_D)$, STAPLE seems better at lower values, and SBA, at higher values. We also observe that in none of the cases does Vote clearly outperform the other methods.

These observations thus suggested that $\sigma(v_D)/\mu(v_D)$ and $\mu(v_D)$ could be used to determine the appropriate label fusion method.

2.4.2. Dissimilarity Factors. The measures $\sigma(v_D)/\mu(v_D)$ and $\mu(v_D)$ cannot be used in practice since the computation of v_D

depends on $v_{k|T}$ (1) and V_{TRUTH} , and thus requires to know the ground truth, which is what we try to estimate with label fusion. We thus needed to find estimates for $\sigma(v_D)/\mu(v_D)$ and $\mu(v_D)$.

We overcame this problem by using the following scheme. For $v_{k|T}$, we first computed the frequency of occurrence $f(x, i)$, between 0 and 1, of each label i for each pixel/voxel x over all raters:

$$f(x, i) = \frac{\#\{k \mid e_k(x) = i\}}{K}. \quad (3)$$

We then computed, for each rater k and each pixel/voxel x , the estimated probability that rater k misclassifies pixel/voxel x , i.e. that the assigned label was a false positive or a false negative:

$$p_k(x) = 1 - f(x, e_k(x)). \quad (4)$$

For each estimated rater's probability $p_k(x)$, we then performed a Bernoulli trial with B experiments to compute

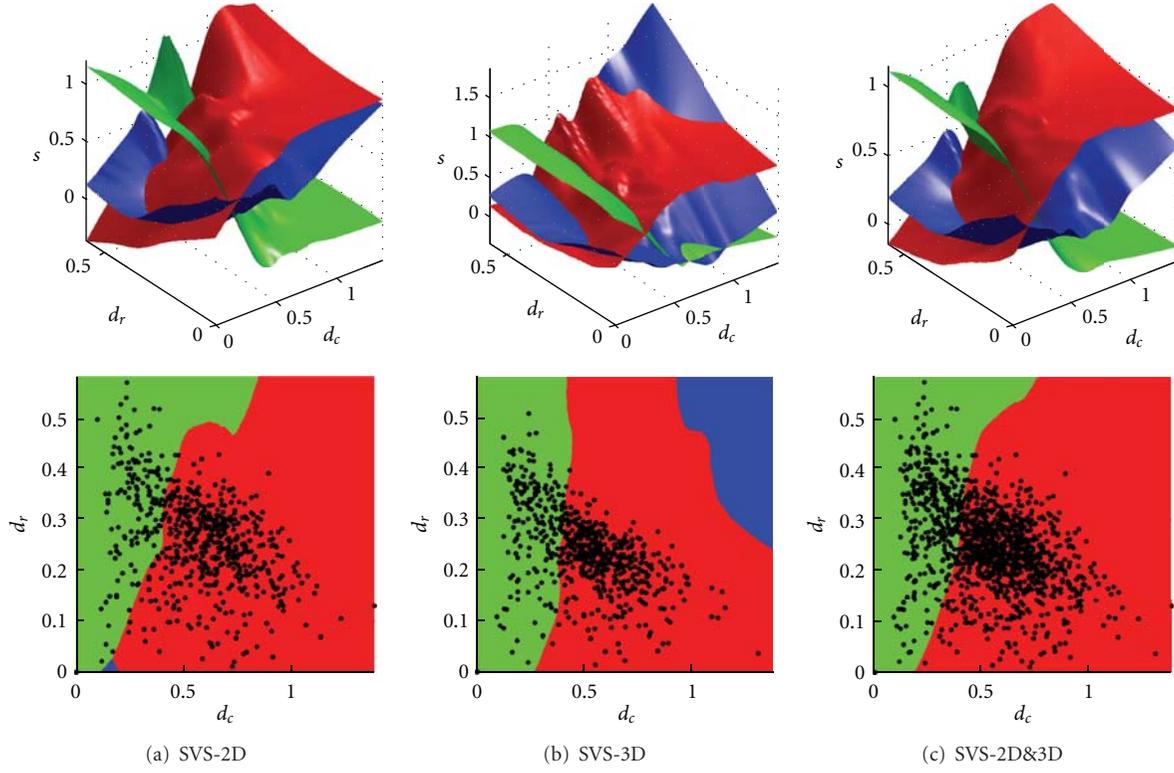


FIGURE 4: (Top) Scoring surface functions in the space (d_c, d_r, s) and (bottom) SVS selection regions in the space (d_c, d_r) , where each method among STAPLE (red), Vote (blue), and SBA (green) gives the highest score, for (a) SVS-2D, (b) SVS-3D, and (c) SVS-2D&3D. The bottom images correspond to the top views of the surfaces presented above. The overlaid scatter plot represents the (d_c, d_r) values of the tests for each SVS version's training set.

the probability $P_k(x)$ that a majority of B “virtual” raters misclassified pixel/voxel x , according to $p_k(x)$:

$$P_k(x) = \sum_{i=\lceil(B+1)/2\rceil}^B \binom{B}{i} p_k^i(x) (1 - p_k(x))^{B-i}. \quad (5)$$

This last equation corresponds to a cumulative sum of the upper half of the probability mass function of a binomial distribution. In this study, we used $B = 99$ so that i ranged from 50 to 99. An odd number for B was used to separate the binomial probability mass function equally into a lower and an upper part, the latter corresponding to a clear majority.

From (5), we were able to compute an estimate v_k of $v_{k|T}$ by summing $P_k(x)$ over all pixels/voxels:

$$v_k = \sum_{x=1}^N P_k(x). \quad (6)$$

To estimate V_{TRUTH} , we used (3) in a similar Bernoulli trial approach. For each pixel/voxel x , we computed a probability that a majority of $B = 99$ “virtual” raters classifies pixel/voxel x as being part of the segmented region, i.e. with label 1, according to $f(x, 1)$:

$$F(x) = \sum_{i=\lceil(B+1)/2\rceil}^B \binom{B}{i} f^i(x, 1) (1 - f(x, 1))^{B-i}. \quad (7)$$

We then summed $F(x)$ over all pixels/voxels to obtain an estimate V of V_{TRUTH} :

$$V = \sum_{x=1}^N F(x). \quad (8)$$

From v_k and V , we defined two empirical factors: the *dissimilarity coefficient* d_c , estimating $\sigma(v_D)/\mu(v_D)$, and the *dissimilarity ratio* d_r , estimating $\mu(v_D)$. These factors are respectively given by

$$\begin{aligned} d_c &= \frac{\sigma(v_k)}{\mu(v_k)}, \\ d_r &= \frac{\mu(v_k)}{V}. \end{aligned} \quad (9)$$

In Figure 3, we demonstrate the performance of these estimates by showing that d_c (a, c) and d_r (b, d) match, with a quasi-one-to-one relationship, their theoretical values $\sigma(v_D)/\mu(v_D)$ and $\mu(v_D)$, respectively, for both the 2D (a, b) and 3D (c, d) training sets.

2.4.3. SVS Training. To perform its selection, SVS finds a score s , from the dissimilarity factors d_c and d_r , for each of STAPLE, Vote, and SBA, i.e. $s_{\text{STAPLE}}(d_c, d_r)$, $s_{\text{VOTE}}(d_c, d_r)$ and $s_{\text{SBA}}(d_c, d_r)$, and selects the label fusion method that gives the

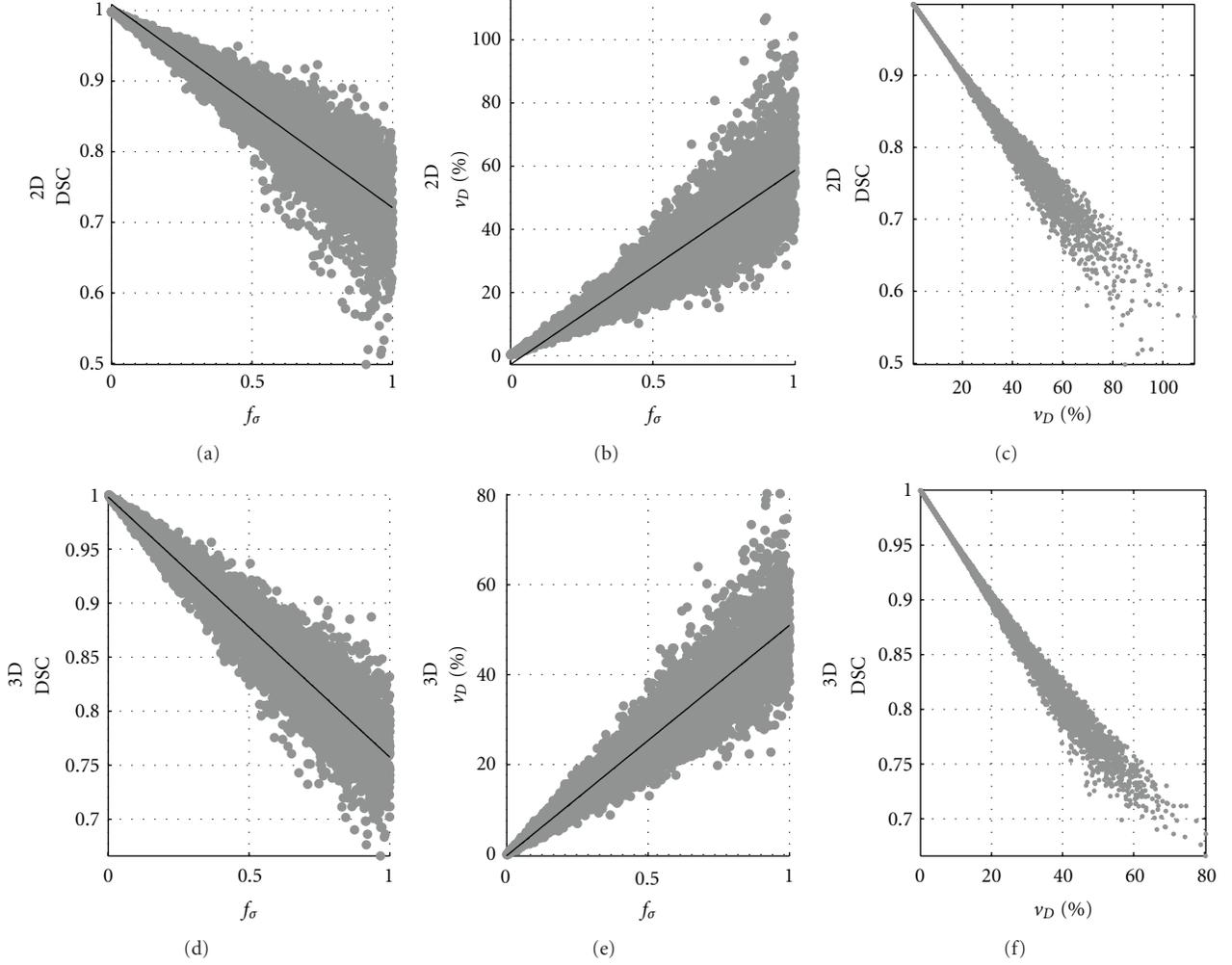


FIGURE 5: (a, d) DSC and (b, e) v_D for all the 6250 rater images of the (a, b) 2D and (d, e) 3D testing sets as a function of f_σ . A linear fit was performed for both scatter plot types (black) showing the quasilinear relationship between f_σ and both DSC and v_D . (c, f) DSC plotted as a function of v_D . We see that the scatter plots also follow a quasilinear trend. This graph demonstrates that, compared to a point with given v_D and DSC, a neighbor point with a higher (worse) v_D can still give a higher (better) or similar DSC, especially for high v_D , questioning the validity of DSC as a performance measure for label fusion.

highest score. The following training procedure was used to determine the scoring functions $s_{\text{STAPLE}}(d_c, d_r)$, $s_{\text{VOTE}}(d_c, d_r)$, and $s_{\text{SBA}}(d_c, d_r)$.

- (1) For each label fusion test t of a given training set, we computed d_c and d_r , according to the approach presented in the last section.
- (2) After performing label fusion with STAPLE, Vote, and SBA, we first summed, for each label fusion method m and test t , the number of pixels/voxels v_m that were different between the label fusion result T_m and the ground truth T , i.e. the number of false positives and false negatives:

$$v_m = \#\{x \mid T_m(x) \neq T(x)\}. \quad (10)$$

For each test t , we assigned a score s of 1 to the label fusion method with the lowest v_m , corresponding to

the best method, 0 to the method with the highest v_m , corresponding to the poorest method, and we linearly interpolated the score value for the remaining method.

- (3) Following the last two steps of the training procedure, we had, for each test t , the five following values: d_c , d_r , s_{STAPLE} , s_{VOTE} , and s_{SBA} . To obtain the continuous scoring functions $s_{\text{STAPLE}}(d_c, d_r)$, $s_{\text{VOTE}}(d_c, d_r)$, and $s_{\text{SBA}}(d_c, d_r)$, we finally fitted, for each method m , a surface $s_m(d_c, d_r)$ using locally weighted linear regression (MATLAB Curve Fitting Toolbox, MathWorks, Natick, MA, USA).

This procedure was performed for each of the 2D and 3D training data sets as well as the combination of both sets resulting in three versions of SVS: SVS-2D (trained with 2D data), SVS-3D (trained with 3D data), and SVS-2D&3D

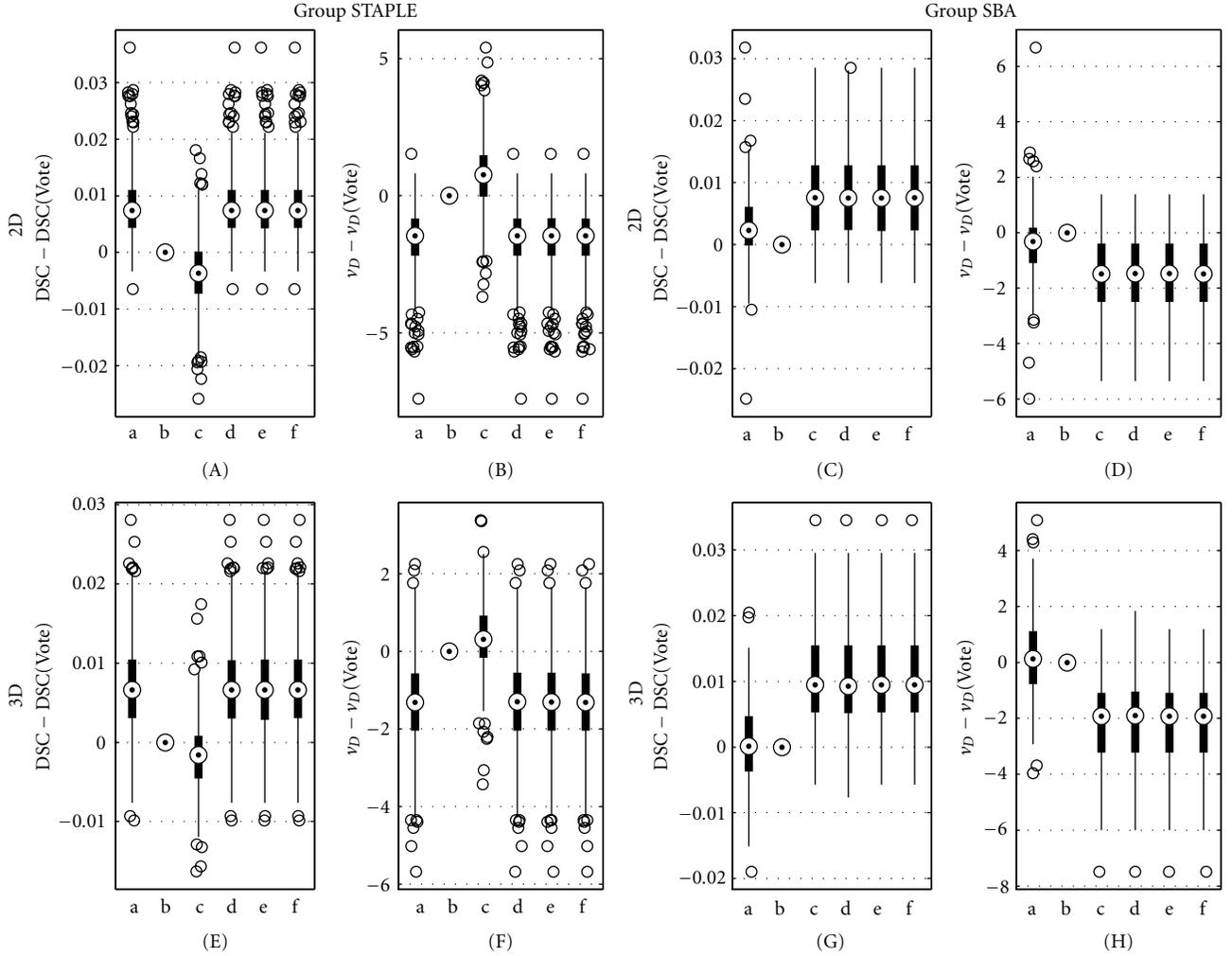


FIGURE 6: Boxplots of (A, C, E, and G) DSC and (B, D, F, and H) v_D , centered on Vote values, for each method for the (A)–(D) 2D and (E)–(H) 3D testing sets. (a) STAPLE, (b) Vote, (c) SBA, (d) SVS-2D, (e) SVS-3D, and (f) SVS-2D&3D. The centered DSC (v_D) corresponds to DSC (v_D) minus DSC (v_D) evaluated for Vote.

(trained with 2D and 3D data). We note that using this scheme, other label fusion methods could be incorporated in SVS, increasing only the number of scoring functions $s(d_c, d_r)$.

Figure 4 presents, for SVS-2D (a), SVS-3D (b), and SVS-2D&3D (c), the scoring surface functions in the space (d_c, d_r, s) as well as the selection regions in the space (d_c, d_r) , where each method gives the highest score. The latter images thus correspond to the top views of the firsts. We observe that the three versions of SVS give very similar delimitations between the methods. Interestingly, with SVS-2D&3D, the border between STAPLE and SBA is almost linear in the region of (d_c, d_r) covered by the label fusion tests.

2.4.4. SVS Selection. We can now describe the SVS method as follows.

- (1) Compute the dissimilarity coefficient d_c and the dissimilarity ratio d_r from the raters' input labels, as described in Section 2.4.2.

- (2) Find the score for each label fusion method using its corresponding scoring surface function.
- (3) Select the label fusion method corresponding to the highest score.

In case of two or more equal scores, which do not imply identical label fusions, a weighted vote “meta fusion” of the label fusion results, obtained with STAPLE, Vote, and SBA, is performed using the scores as weights. In practice, this situation is uncommon. We point out that, besides the SVS versions presented here, this “meta fusion” approach, i.e. performing a label fusion of STAPLE, Vote, and SBA, has also been tested (results not presented), using each of STAPLE, Vote, and SBA as “meta fusion” method with and without score weights for the two latter methods. However, no “meta fusion” outperformed the versions of SVS presented in this study.

We also point out that d_c and d_r depend on the decision matrix E only, i.e. the input labels. Effectively, this ensures

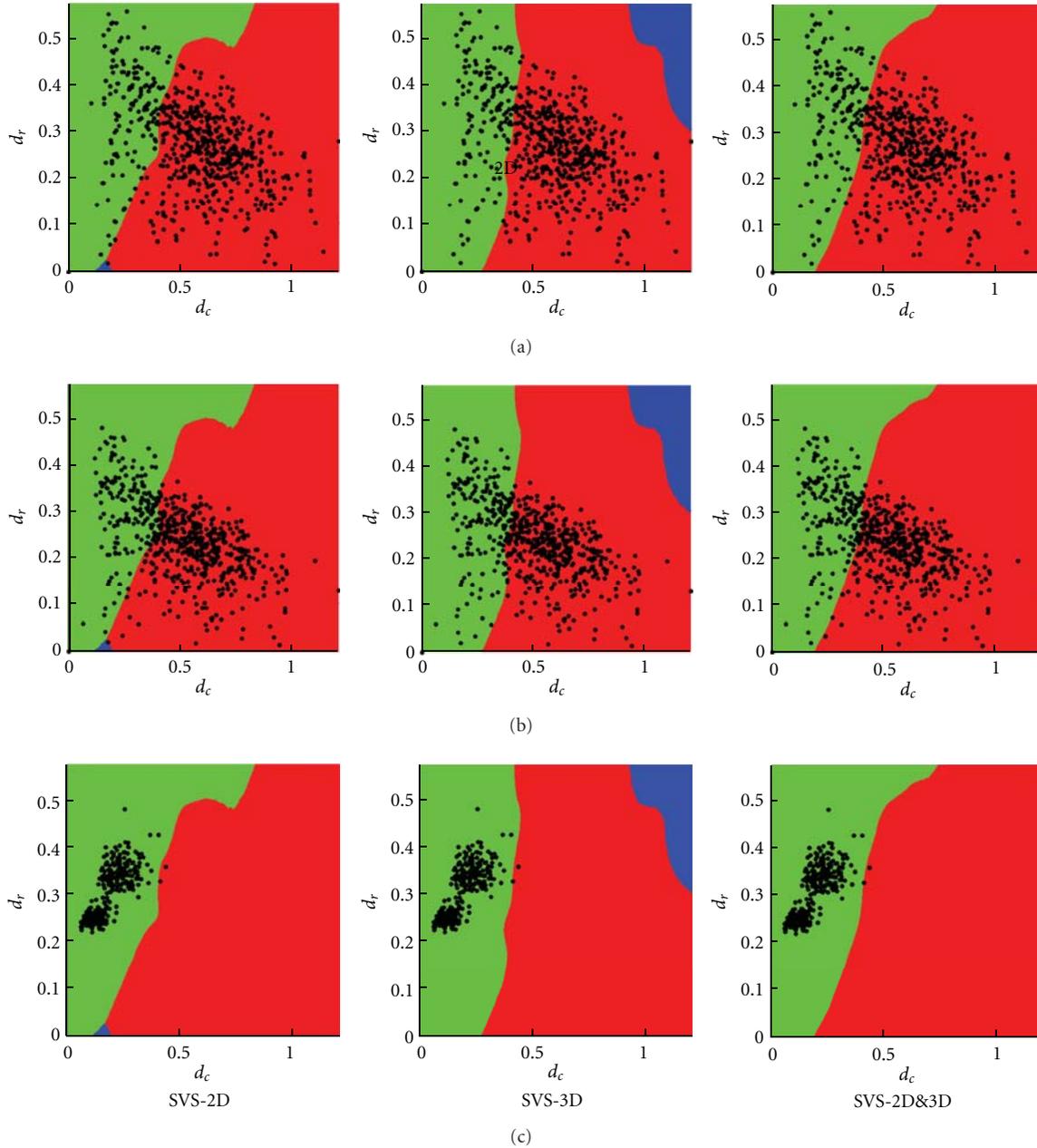


FIGURE 7: Scatter plots of the (d_c, d_r) pairs for all the tests of the (a) 2D and (b) 3D testing sets, and (c) real data set, overlaid on the SVS selection regions described in Figure 4 for (left) SVS-2D, (middle) SVS-3D, and (right) SVS-2D&3D. We note that the three versions of SVS selected SBA as the most appropriate method for nearly all tests of the real data set (c).

that there are no external parameters to the input data that may affect the sensitivity of the technique. Moreover, since d_c and d_r are normalized values, we believe that the technique should not be sensitive to the training data. In fact, we observe in Figure 4 that the different training sets gave similar regions.

2.5. Performance Measure. To measure the performance of the label fusion techniques, we computed v_D , as well as the Dice similarity coefficient (DSC), an established measure

widely reported in the field [1, 2, 9, 11, 12, 15], between each label fusion image and the ground truth. DSC is given by

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B|}, \quad (11)$$

where $|Z|$ is the area or volume of the segmented region Z .

To further characterize our testing sets and insure the deformation factor f_σ reflected its initial intent, we computed the DSC between each deformed image and its ground truth. Figures 5(a), 5(b), 5(d), and 5(e) show the relationship

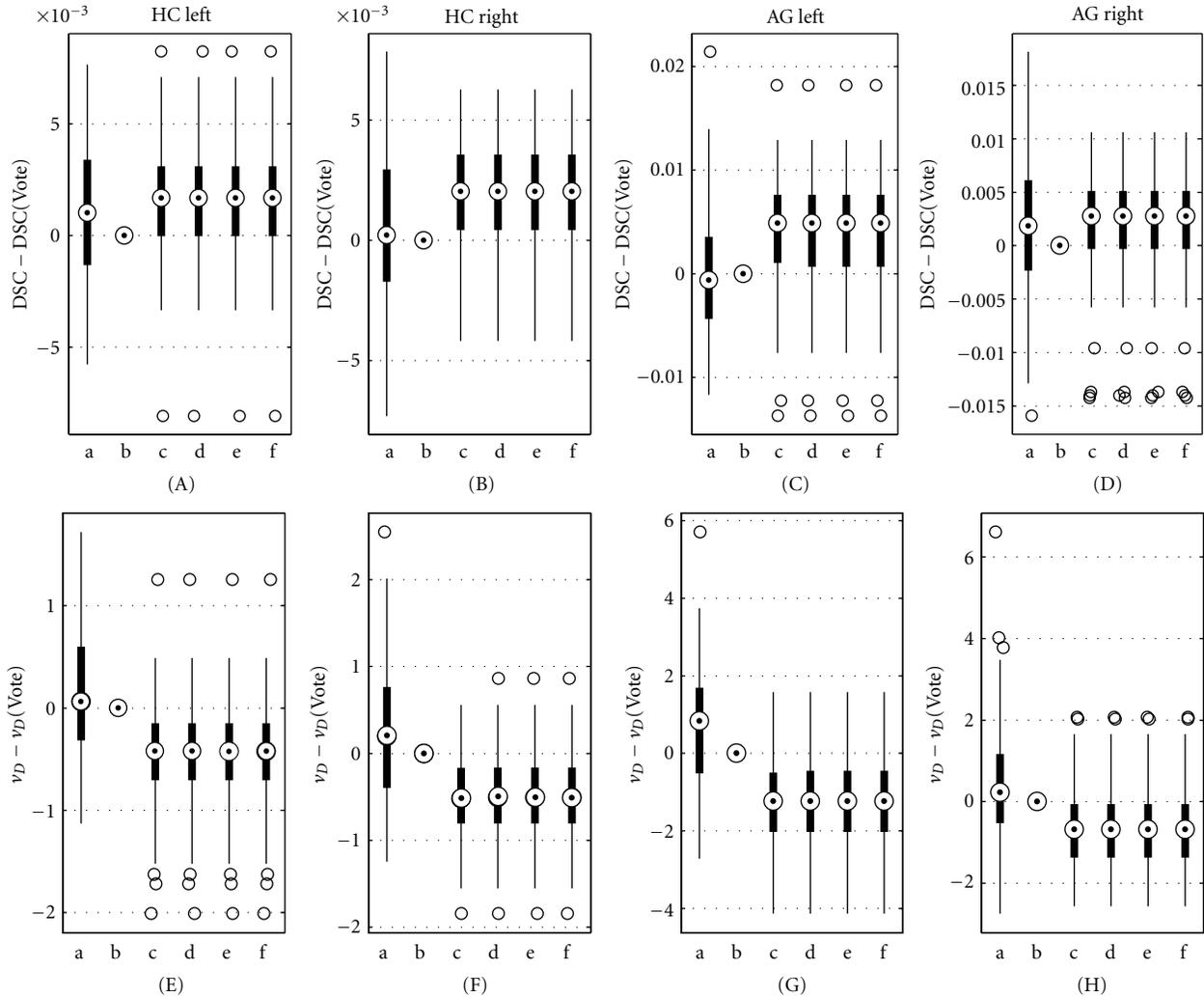


FIGURE 8: Boxplots of (A)–(D) DSC and (E)–(H) v_D , centered on Vote values, for (A, E) HC left, (B, F) HC right, (C, G) AG left, and (D, H) AG right. (a) STAPLE, (b) Vote, (c) SBA, (d) SVS-2D, (e) SVS-3D, and (f) SVS-2D&3D. The centered DSC (v_D) corresponds to DSC (v_D) minus DSC (v_D) evaluated for Vote.

between DSC, along with v_D , and the deformation factor f_σ for the 2D (a, b) and 3D (d, e) testing sets. Figures 5(c) and 5(f) show the quasilinear relationship between DSC and v_D .

3. Results

3.1. 2D Simulated Data. The three existing techniques (STAPLE, Vote, and SBA) as well as the three versions of SVS (SVS-2D, SVS-3D, and SVS-2D&3D) were used to perform the label fusion of the 10 images of each of the 625 tests of the 2D testing set. Figures 6(A)–6(D) show boxplots of DSC (A, C) and v_D (B, D), centered on the Vote values, obtained with the six fusion methods. To see the improvement brought by SVS (methods d, e, and f), the boxplots have been separated in two groups, *Group STAPLE* and *Group SBA*, determined by the selection performed by SVS-2D&3D on the testing set (see Figure 7(a), right). The data in *Group STAPLE* and

Group SBA are the tests for which SVS-2D&3D selected STAPLE and SBA, respectively. We see that the SVS boxplots, matching the selected method's, give in both groups higher DSC and lower v_D , while each of STAPLE (method a) and SBA (method c) is outperformed in its counterpart group. Regarding Vote (method b), it gives better performance than SBA in *Group STAPLE* but seems to be the worse method in *Group SBA*. We also see that the three versions of SVS are similar despite the different training sets. Figure 7(a) presents the distribution of the (d_c, d_r) pairs for all the tests of the testing set among the regions of SVS-2D, SVS-3D, and SVS-2D&3D.

3.2. 3D Simulated Data. The experiment described in the last section was also performed on the 3D testing set. DSC and v_D boxplots are presented in Figures 6(E, G) and 6(F, H), respectively. The results are very similar to the 2D testing set's; the

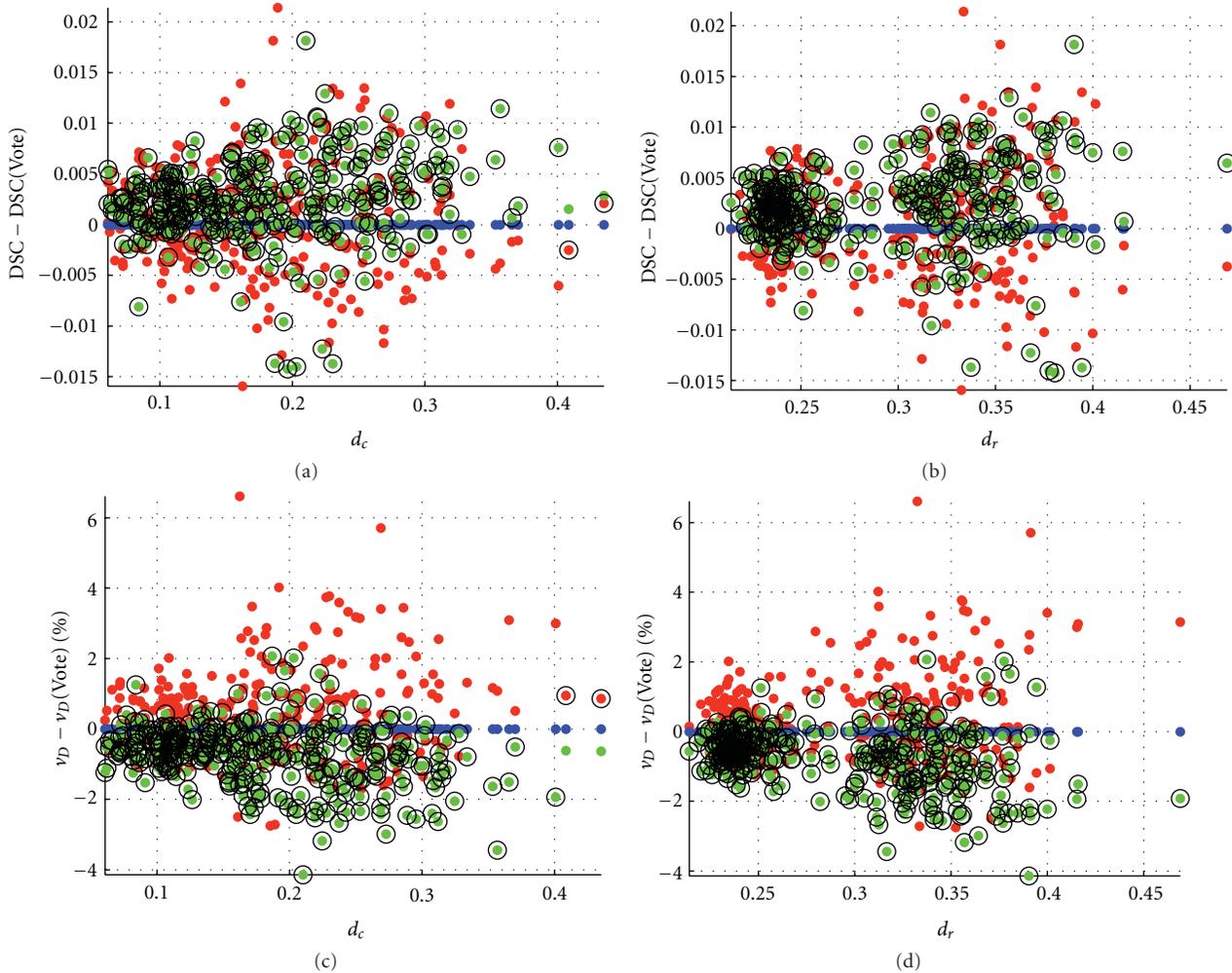


FIGURE 9: Scatter plots of (a, b) DSC and (c, d) v_D , centered on Vote values, as functions of (a, c) d_c and (b, d) d_r , for the 312 label fusion tests of the real data set using SVS-2D&3D. The results were nearly identical for SVS-2D and SVS-3D. The centered DSC (v_D) corresponds to DSC (v_D) minus DSC (v_D) evaluated for Vote. DSC and v_D are represented as dots for STAPLE (red), Vote (blue), and SBA (green), and as black circles for SVS-2D&3D.

three versions of SVS give in both groups higher DSC and lower v_D . Figure 7(b) shows the distribution of the (d_c , d_r) pairs among the selection regions.

3.3. Real Data. Figure 8 presents the DSC (A–D) and v_D (E–H) boxplots, respectively, obtained for each of HC and AG, left and right. Since the three versions of SVS (methods d, e, and f) selected SBA for nearly all label fusions, as shown in Figure 7(c), the boxplots are almost identical to SBA’s. We see that SBA/SVS overall gives the highest DSC and the lowest v_D . This is also shown in Figure 9, which presents scatter plots of DSC (a, b) and v_D (c, d), centered on the Vote values, as a function of d_c (a, c) and d_r (b, d) for all the 312 label fusion cases. SBA/SVS is overall superior to STAPLE and Vote, with DSC and v_D respectively above and below STAPLE and Vote means for the majority of the label fusion cases.

4. Discussion

4.1. Findings. We showed on a large set of different simulated data that the label fusion method giving the label closest

to the ground truth was not the same depending on the dissimilarity among the raters.

Regarding robustness, we showed that SVS outperformed any single method among STAPLE, Vote, and SBA, regardless of the training set. Applying SVS-2D (trained with 2D data) and SVS-3D (trained with 3D data) on 3D and 2D data, respectively, we still obtained better performance than STAPLE, Vote, and SBA. Effectively, the three versions of SVS showed similar results, explained by similar selection regions (Figure 4). This suggests that SVS is independent of the type of training set, 2D or 3D, and that the delimitations of the selecting regions with SVS-2D&3D could represent what we should really expect since there are more training tests.

We also demonstrated that with real data, Vote was not necessarily the method of choice; in our study, SBA was better than Vote and STAPLE. To our knowledge, SBA has not been widely used in the literature, and it might have been underestimated.

4.2. *Limitations.* The first and obvious limitation of the SVS technique is that it is upper-bound limited to the best technique (either STAPLE, Vote, or SBA) in each case.

Secondly, we used DSC and v_D in this study as the criteria for assessing the label fusion methods, the first being commonly used in the literature. However, we think that v_D gives a better indication of the difference between a rater image and the ground truth. This is demonstrated in Figure 8 for HC left, HC right, and AG right. For these regions, while STAPLE's DSC medians are higher (better) than Vote's, v_D medians are higher (worse), meaning that there are more false positives and/or negatives. Also, in Figures 5(c) and 5(f), we show that compared to a point with given v_D and DSC, a neighbor point with a higher v_D (more false positives and/or negatives) can still give a higher (better) or similar DSC, especially for high v_D . This difference between DSC and v_D might be explained by the fact that DSC normalizes by the mean area/volume of the label fusion and ground truth, while v_D normalizes by the area/volume of the ground truth only. Therefore, the denominator in v_D remains constant, while the denominator in DSC varies between label fusions. The comparison is thus not performed on the same basis. Although we could argue on which measure is the most appropriate, this questions the validity of DSC as a performance measure for label fusion if the ground truth is available. We thus keep in mind for future work that DSC is not necessarily the best criterion in this case and that v_D should be used instead.

Thirdly, we did not assess the influence of the number and the selection of input labels on the performance of the label fusion strategies. While these two aspects are important, as reported in some studies [2, 12], our objectives were primarily to characterize three existing label fusion strategies and to propose a selection method based on our observations. We will confront these aspects in future work.

5. Conclusion

We proposed a method that automatically selects the most appropriate label fusion method based on the dissimilarity of input labels. Overall, the SVS technique performed better with simulated data compared to either individual technique among STAPLE, Vote, and SBA. For real data, SVS selected SBA for almost all cases, which was overall superior to STAPLE and Vote.

Abbreviations

AD:	Alzheimer's disease
AG:	Amygdala
AU:	Arbitrary units
DSC:	Dice similarity coefficient
HC:	Hippocampus
MRI:	Magnetic resonance imaging
STAPLE:	Simultaneous Truth and Performance Level Estimation
SBA:	Shape-Based Averaging
SVS:	STAPLE-Vote-SBA

2D: Two dimensional
3D: Three dimensional.

Acknowledgments

The authors thank Dr. J. C. Pruessner and Dr. D. L. Collins (McGill University, Montréal, Canada), and the International Consortium for Brain Mapping for access to label and MRI data. This work was supported by an operating grant from the Ministère du Développement Économique, de l'Innovation, et de l'Exportation du Québec.

References

- [1] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.
- [2] D. L. Collins and J. C. Pruessner, "Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion," *NeuroImage*, vol. 52, no. 4, pp. 1355–1366, 2010.
- [3] J. R. Petrella, R. E. Coleman, and P. M. Doraiswamy, "Neuroimaging and early diagnosis of alzheimer disease: a look to the future," *Radiology*, vol. 226, no. 2, pp. 315–336, 2003.
- [4] K. Kantarci and C. R. Jack, "Neuroimaging in Alzheimer disease: an evidence-based review," *Neuroimaging Clinics of North America*, vol. 13, no. 2, pp. 197–209, 2003.
- [5] G. Chetelat and J. C. Baron, "Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging," *NeuroImage*, vol. 18, no. 2, pp. 525–541, 2003.
- [6] L. O. Wahlund, O. Almkvist, K. Blennow et al., "Evidence-based evaluation of magnetic resonance imaging as a diagnostic tool in dementia workup," *Topics in Magnetic Resonance Imaging*, vol. 16, no. 6, pp. 427–437, 2005.
- [7] M. Chupin, A. R. Mukuna-Bantumbakulu, D. Hasboun et al., "Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: method and validation on controls and patients with Alzheimer's disease," *NeuroImage*, vol. 34, no. 3, pp. 996–1019, 2007.
- [8] J. Barnes, R. G. Boyes, E. B. Lewis et al., "Automatic calculation of hippocampal atrophy rates using a hippocampal template and the boundary shift integral," *Neurobiology of Aging*, vol. 28, no. 11, pp. 1657–1663, 2007.
- [9] X. Artaechevarria, A. Muñoz-Barrutia, and C. Ortiz-de-Solórzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, Article ID 4785214, pp. 1266–1277, 2009.
- [10] M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, Article ID 5487420, pp. 1714–1729, 2010.
- [11] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.
- [12] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, 2009.

- [13] T. Rohlfing and C. R. Maurer Jr., "Multi-classifier framework for atlas-based image segmentation," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2070–2079, 2005.
- [14] S. K. Warfield, K. H. Zou, and M. Wells III, "Validation of image segmentation and expert quality with an expectation-maximization algorithm," in *Proceedings of the 5th International Conference on Medical Imaging Computing and Computer Assisted Interventions (MICCAI '02)*, pp. 298–306, Springer, Tokyo, Japan, September 2002.
- [15] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, 2004.
- [16] T. Rohlfing, D. B. Russakoff, and C. R. Maurer Jr., "Expectation maximization strategies for multi-atlas multi-label segmentation," *Inf Process Med Imaging*, vol. 18, pp. 210–221, 2003.
- [17] T. Rohlfing and C. R. Maurer Jr., "Shape-based averaging," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 153–161, 2007.
- [18] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: theory and rationale for its development," *NeuroImage*, vol. 2, no. 2, pp. 89–101, 1995.
- [19] J. C. Pruessner, D. L. Collins, M. Pruessner, and A. C. Evans, "Age and gender predict volume decline in the anterior and posterior hippocampus in early adulthood," *Journal of Neuroscience*, vol. 21, no. 1, pp. 194–200, 2001.
- [20] T. Rohlfing, D. B. Russakoff, and C. R. Maurer Jr., "Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 983–994, 2004.

Research Article

Modulating the Focus of Attention for Spoken Words at Encoding Affects Frontoparietal Activation for Incidental Verbal Memory

Thomas A. Christensen,¹ Kyle R. Almryde,¹ Lesley J. Fidler,¹ Julie L. Lockwood,¹
Sharon M. Antonucci,^{1,2} and Elena Plante¹

¹Department of Speech, Language & Hearing Sciences, The University of Arizona, Tucson, AZ 85721, USA

²Department of Communicative Sciences and Disorders, New York University, NY 10012, USA

Correspondence should be addressed to Thomas A. Christensen, tac1@email.arizona.edu

Received 15 April 2011; Revised 20 August 2011; Accepted 24 August 2011

Academic Editor: Ramin Parsey

Copyright © 2012 Thomas A. Christensen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Attention is crucial for encoding information into memory, and current dual-process models seek to explain the roles of attention in both recollection memory and incidental-perceptual memory processes. The present study combined an incidental memory paradigm with event-related functional MRI to examine the effect of attention at encoding on the subsequent neural activation associated with unintended perceptual memory for spoken words. At encoding, we systematically varied attention levels as listeners heard a list of single English nouns. We then presented these words again in the context of a recognition task and assessed the effect of modulating attention at encoding on the BOLD responses to words that were either attended strongly, weakly, or not heard previously. MRI revealed activity in right-lateralized inferior parietal and prefrontal regions, and positive BOLD signals varied with the relative level of attention present at encoding. Temporal analysis of hemodynamic responses further showed that the time course of BOLD activity was modulated differentially by unintentionally encoded words compared to novel items. Our findings largely support current models of memory consolidation and retrieval, but they also provide fresh evidence for hemispheric differences and functional subdivisions in right frontoparietal attention networks that help shape auditory episodic recall.

1. Introduction

Attention is known to alter neural processing at multiple levels of both the peripheral and central nervous systems, and both auditory and visual attention have been conceptualized as operating in both “top-down” and “bottom-up” modes [1–7]. Top-down mechanisms reflect goal-based control in order to direct attention to particular targets or to sustain attention over time. In contrast, bottom-up mechanisms have traditionally been defined by the phenomenon of reflexive attentional orienting, as when attention is drawn without intent by highly salient sensory stimuli such as a sudden loud noise or flash of light. Recently, however, some investigators have more broadly considered bottom-up effects as relevant for any incoming stimuli, with the relative saliency of the stimulus influencing whether it is ultimately encoded into memory [8]. Two recent theoretical models address the question of what roles stimulus saliency

might play in first successfully encoding information into memory and then later retrieving it. As discussed below, the “Embedded Processes” model and the “Attention-to-Memory” model, while similar, also highlight the potentially divergent roles that attention at encoding may play in later recall. Brain mapping studies based on these models have only started to identify the neural substrates that underlie these cognitive processes in different domains.

In real life situations, numerous bits of information co-occur in the environment and vie for attention. Given that memory is a limited capacity system [9–11], the information encoded into memory must be restricted to what is relevant. According to the “Embedded Processes” model, attention is central to this process. Information is first encoded into working memory through an active process known as “attentional scanning” [12]. This process first searches through a set of potential memory items, then selects the most salient

items and brings them into the “focus of attention” [12–14]. However, even though the attentional scanner is responsible for selecting the appropriate stimuli, its capacity is limited, and it therefore can maintain focus on only small amounts of information at a time. This also places limits on the capacity of working memory. Cowan and colleagues have speculated that regions in and around the temporal-parietal junction are key to focusing attention [14]. More recently, others have found that the majority of functional neuroimaging data on working memory is consistent with the tenets of this model [15], with ventral posterior (inferior parietal cortex and intraparietal sulcus) regions associated with attentional focus and working memory maintenance, and lateral prefrontal regions involved in the executive control of attention and further manipulation of information, once it is in working memory [16, 17]. If this interpretation is correct, then manipulations of attention at encoding should preferentially affect activation in parietal regions.

Another recent paradigm, the “Attention-to-Memory” model [8, 18–20], shares functional features with the previous model, but it also emphasizes a distinction between the roles of ventral and dorsal parietal cortex in memory retrieval. Successful recognition memory for intentionally studied items is predicted by increased activation of *dorsal* posterior parietal cortex at stimulus encoding, whereas activation in *ventral* posterior parietal cortex at encoding predicts success in incidental memory and perceptual priming tests [21, 22]. This latter observation suggests a role for this ventral region not only in the bottom-up encoding of highly salient or unexpected sensory stimuli, but also in the successful retrieval of unintentionally encoded memories [21]. According to these prevailing theories, retrieval of incidental or perceptual memory representations requires a bottom-up shift in selective attention, whereas retrieval of intentional memories requires a top-down mechanism that reflects the individual’s conscious intent to focus attention on incoming signals [21, 22]. Importantly, however, the vast majority of studies exploring the intersection of attention and episodic memory—and thus the theories on which these studies are based—are heavily grounded in the use of visual paradigms [23–30]. Even studies that focus specifically on verbal encoding have mostly employed reading rather than listening tasks [31–37]. It is therefore necessary to examine the effects of attention on item retrieval in listening tasks in order to establish whether the dorsal/ventral distinction observed in parietal cortex is specific to visual stimuli, or instead reflects a more generalized organization in this area of the brain.

To examine these ideas in greater depth, it is necessary to learn more about the role of attention in establishing a selective focus on task-relevant sensory stimuli. Moreover, in order to understand how limited attentional resources are allocated, it is also important to study the encoding of items both *within* the focus of attention as well as those *peripheral* to it. Our goal in this study was to devise a method to modulate attention toward spoken words during a simple listening task. Based on earlier studies [38], we predicted that words not intentionally within the focus of attention would be encoded into memory to some degree. During the

encoding phase of our study [39], participants initially heard a list of words belonging to two semantic classes (animals or foods), and these words were presented by two different talkers (female or male). Participants were given explicit instructions to remember only the animal words presented in the female voice (the targets). Importantly, this instruction resulted in a focus on the female voice, which had the consequence of placing the nontarget food words presented in that voice also within the focus of attention. We predicted that these words would also carry higher salience at encoding than words heard in the male voice, creating what we refer to as the *High-Attention* condition in this study. Following this rationale, the food words presented in the male voice would not be fully in the primary focus of attention (the *Low-Attention* condition). We contrasted these two encoding conditions with responses to novel food words that were not previously heard at any time during the experiment. We reasoned that processing of these *NEW* words should reflect only the role of attention in initial word encoding, but not a later role in retrieval. Using this paradigm coupled with functional magnetic resonance imaging (fMRI), we were able to disambiguate the neural activation patterns associated with attention at first encoding from those reflecting the effects of attention at encoding on subsequent retrieval.

2. Materials and Methods

2.1. Participants. We recruited 14 native English-speaking adults (9 women; ages 18–49; mean 24 years) living in the Tucson area. Exclusion criteria were a history of speech, language, or other neurological disorders, and all volunteers reported good general health with no contraindications for MRI scanning. The study was approved by the University of Arizona Institutional Review board, and written informed consent was obtained from all participants. For one participant, behavioral response data are not available due to computer error.

2.2. Auditory Stimuli. Word stimuli were single, concrete nouns (one to three syllables) that were either names of animals or foods. Some words were recorded in an unfamiliar female voice, and some in an unfamiliar male voice. Stimulus durations ranged from 204–903 ms (mean = 532 ms) and were presented in an event-related format with an average interstimulus interval (ISI) of 871 ± 157 ms. ISI was jittered and “null” events were included to facilitate estimation of the hemodynamic responses for deconvolution analysis (see below). Ordering of words belonging to the different stimulus categories described below was pseudorandom, and participants listened to them through MRI-compatible stereo headphones (Resonance Technology Inc., Northridge, Calif, USA).

2.3. Procedure. The experimental design is shown in Figure 1. A prescan practice phase was followed by an encoding phase and a test phase (both with scanning). These were then followed by a surprise postscan memory test that specifically measured the listener’s capacity to remember any words that were encoded during the earlier phases of

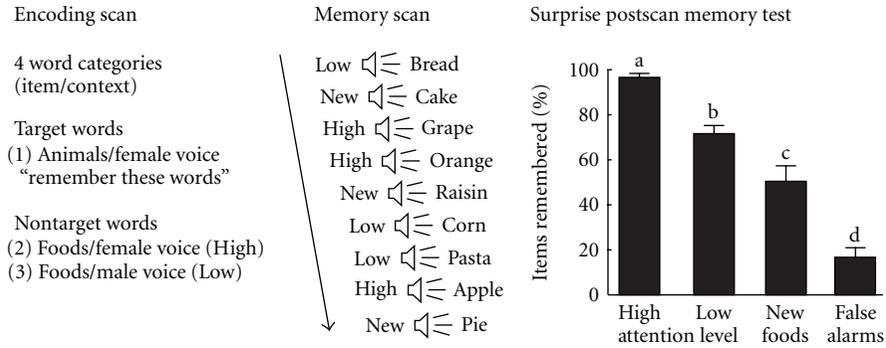


FIGURE 1: Experimental design for studying incidental encoding and memory. During the *Encoding scan* [39], listeners were presented with a list of single nouns (animals or foods). Half of the words were recorded in a female voice, and half in a male voice, thus yielding multiple combinations of item and voice context. Throughout the experiment, listeners' attention was focused on only one subset of words: animal words presented in the female voice (*Target words*). This was accomplished first by giving them explicit instructions to select and remember these words during the *Encoding scan*, and later, by instructing them to select these items by responding "yes" during the *Memory scan*. However, as predicted, some of the *Nontarget* words were also encoded unintentionally, with food words presented in the female voice carrying greater salience (the *High-Attention* condition) than food words presented in the male voice (the *Low-Attention* condition). During the *Memory scan*, participants were presented once again with both *High-* and *Low-Attention* food words, along with novel food words (*NEW*) as a control category representing the baseline attention condition. Following the *Memory scan*, a *Surprise memory test* was used to confirm the extent to which *High-Attention*, *Low-Attention*, and *NEW* food words were successfully encoded. The chart at right shows the group-averaged memory scores ($N = 13$; means \pm SEM) for these three word categories, along with ten foils to measure *false alarm* rate. Bars capped by different letters are significantly different at $P < 0.001$ (two-tailed ANOVA).

the experiment, whether or not the listener was asked to remember them. The results of the test phase and scan, along with results from the postscan behavioral test, are the subject of this report. Results obtained during the encoding phase have been published previously [39].

2.3.1. Encoding Phase and Scan. Participants listened to the animal and food words that were presented in either the male or female voice during a prescan practice period and during the encoding phase. Using a block design, the encoding phase consisted of two scans that were counterbalanced for order across participants. In one scan, approximately one-third of the total words used for this study were presented binaurally, with the male and female voices presented sequentially. In a second scan, two-thirds of the words were presented dichotically, with the female voice heard in one ear and the male voice heard in the other (order counterbalanced across participants). The ratio of targets (animal word + female voice) to nontargets was 1 : 3 in both tasks. Although all participants were explicitly instructed to focus on and remember the animal words presented in the female voice, these were not actually of interest for the purposes of this study. Rather, this instruction was used deliberately to assure that the primary focus of attention was on the female voice. As a result, food words presented in the female voice were associated with a higher level of attention (defined as the *High-Attention* condition) and food words presented in the male voice were associated with a lower level of attention (the *Low-Attention* condition).

2.3.2. Test Phase and Scan. An incidental memory scan immediately followed the encoding scan. During this scan, participants heard words presented by a second female

speaker whose voice had not been heard earlier. This precluded the possibility that the voice in the memory task could provide a cue to the encoding context. Participants were asked to respond "yes" via button press if the word presented was one of their target words (animal words originally spoken in the female voice) and "no" to all other words. Therefore, both the decision ("not a target") and response demands (selecting the "no" button) associated with all words analyzed in this study were identical. Participants were instructed to respond quickly but accurately. During this task, participants mostly heard *nontarget* words. A combination of *High-Attention* and *Low-Attention* food words (58 trials each) as well as 58 *NEW* food words that had not been presented previously were pseudorandomized and presented in a fixed order to all participants. To lessen fatigue, participants were tested over the course of two separate scans (each lasting ~ 9 min with a brief rest in between), during which a combined total of 232 words (randomized with 58 nulls) were presented. Data from the two scans were then concatenated for further analysis.

2.3.3. Surprise Postscan Memory Test. In order to assure that there was an effect of the attentional manipulation at encoding, we administered a surprise memory test within 15 minutes of the final scan. Participants were presented verbally with a list of 40 food items and asked to recall whether they had heard each word at any time during the experiment. The list consisted of 10 *High-Attention*, 10 *Low-Attention*, and 10 *NEW* food words that were all heard during the study phase and scan. A fourth category, 10 food items not presented in any of the scans, was added to measure false-alarm rate. These 40 words were presented verbally in a fixed pseudorandom order by the experimenter and

the participants verbally responded “yes” or “no” to indicate whether they remembered having heard these words at any time during the experiment.

2.4. Whole-Brain fMRI Imaging. Scans were acquired with a 3.0T GE Signa VH/i scanner (General Electric Medical Systems, Madison, Wis, USA) equipped with a quad-head RF coil. First, T1-weighted, fast-spin echo (FSE) axial images covering the entire brain were acquired in 26 slices with an in-plane resolution of $3.44 \times 3.44 \times 5$ mm. Next, two functional T2*-weighted scans were acquired using a spiral in/out pulse sequence (TR = 2.3 s, TE = 30 ms, flip angle = 90° , 26 slices at 5 mm with no gap, matrix = 64×64 , FOV = 22 cm^2) [40]. Finally, high-resolution spoiled gradient-echo (SPGR) images were obtained in the sagittal plane (TR = 30, TE = min, flip angle = 30° , 124 slices at 1.5 mm, matrix = 256×256 , FOV = 25 cm^2) and aligned with the FSE images for improved regional localization and coregistration of functional data across participants after transforming the images into Talairach space [41]. Due to scanner-related problems, we were unable to use the SPGR images from three participants, and instead relied on the first set of axial FSE anatomical images for alignment with the functional scans.

2.4.1. Data Analysis. Structural and functional brain images were analyzed for each participant individually with AFNI [42] followed by a group analysis. Blood-oxygenation-level-dependent (BOLD) contrast images were coregistered with anatomical data after preprocessing using standard procedures for slice-time correction, removing linear signal drift and correcting for head motion. All volumes were realigned to the base volume and spatially smoothed using a 6 mm Gaussian kernel. Data were then normalized to a scale of 0–100%, and functional images were coregistered to the structural data followed by transformation into standard Talairach space. The first four volumes in each run were also discarded to allow for T1 equilibration, and the two runs were then concatenated for further analysis. A general linear model using a gamma-spline hemodynamic response function was used to estimate magnitude parameters for events of interest for each stimulus condition in each individual. In order to capture the amplitude of the BOLD activity over time, eight separate activity models were developed. The first was time locked to the stimulus onset, and each of the seven subsequent models was offset by one-TR (2.3 s) increments from stimulus onset. Stimulus functions were then convolved with the fMRI time-series data from each individual. Parameter estimates for the resulting regressors for each condition were calculated using the least-squares fit of the models to the time-series data. Finally, a group analysis was performed with repeated-measures ANOVA (treating individuals as a random effect) to help confirm key cortical regions that showed differential neural activity for each of the four conditions. Monte Carlo simulation (8 mm FWHM blur; 1000 iterations) was used to correct the group data for multiple comparisons. Voxel-wise (uncorrected) threshold was $P = 0.005$, and minimum corrected cluster volumes in original space were 32 contiguous voxels at $P < 0.05$.

2.4.2. Data-Driven Cluster Analysis. To identify regions of interest (ROI) associated with positive BOLD signal across the *High-Attention*, *Low-Attention*, and *NEW* categories, we first combined the three group-averaged datasets. This allowed us to develop ROI that were associated with activation in any of the three conditions of interest. We were particularly interested in areas in and around the dorsolateral prefrontal cortex and the temporoparietal junction due to their proposed involvement in executive control as well as phonological short-term storage and language processing [12, 43, 44]. However, since the proposed involvement of these regions is based almost entirely on visual rather than auditory paradigms, and there is growing evidence for bilateral involvement in language processing [45, 46], we examined the following ROI in both hemispheres. Magnitude estimates and peak Talairach coordinates are listed in Table 1.

Masks corresponding to distinct anatomical regions of interest were developed, and ROI were defined based on the threshold-corrected regions of significant activation. Brain regions showing significant positive BOLD activation are listed in Table 1 and described as follows (approximate Brodmann areas in parentheses): *Frontal Lobe* (dorsal to ventral): medial frontal gyrus (6/32), middle frontal gyrus (9/46), inferior frontal gyrus (47/pars orbitalis), and anterior insular cortex (13). *Parietal Lobe* (inferior parietal lobule): angular gyrus (39) and supramarginal gyrus (40). We looked for additional ROI along the anterior-posterior axis of the superior temporal lobe due to its routine involvement in speech and language tasks [47–51], but found only one with significant activation: superior temporal gyrus including and extending around the transverse temporal gyrus (BA41/42). Sub-cortical ROI included basal ganglia (caudate body) and thalamic nuclei (anterior). Finally, several ROI were localized to the anterior and posterior lobes of the cerebellum.

3. Results and Discussion

3.1. Surprise Postscan Memory Test. The surprise memory test was intended to provide proof of concept that the attentional manipulation at encoding actually produced an effect. Specifically, *High-Attention* words should be better remembered than *Low-Attention* words, and each of these should be better attended than *NEW* words. Conversely, participants should not indicate recognition for words introduced as foils that were not presented earlier at any time in the experiment.

Figure 1 shows the result of the group memory analysis (full variance model: $F_{3,48} = 17.44$, $N = 13$ listeners; $P < 0.001$, two-tailed). As predicted, the strongest memory scores were for responses to the *High-Attention* words. These words showed the best hit rate at $93.8 \pm 2.4\%$ (mean \pm SEM), compared to $70.8 \pm 4.6\%$ for *Low-Attention* words, and $51.0 \pm 6.5\%$ for *NEW* words (all means significantly different based on post-hoc, two-tailed ANOVA; see Figure 1). Additionally, words introduced as foils during the postscan memory test were correctly rejected $86.2 \pm 3.3\%$ of the time, indicating a low false alarm rate.

TABLE 1: Regions selected for ROI analysis.

	Left hemisphere		Right hemisphere	
	x, y, z	T_{\max}	x, y, z	T_{\max}
Condition 1: High Attention				
<i>Frontal lobe</i>				
Medial frontal gyrus (6/32)	-2, 3, 48	8.97	1, 12, 45	7.31
Middle frontal gyrus (9/46)	-52, 22, 28	3.26	47, 18, 37	3.17
Inferior frontal gyrus (47)	-52, 17, -3	3.72	46, 16, -4	4.39
Insula (13)	-36, 21, 5	4.92	36, 21, 6	6.50
<i>Parietal lobe</i>				
Supramarginal gyrus (40)	-35, -62, 39 ⁽¹⁾	3.76	39, -55, 46 ⁽²⁾	3.39
<i>Superior temporal lobe</i>				
Heschl's gyrus (41/42)	-50, -18, 10	5.54	53, -17, 10	9.09
<i>Caudate body</i>	-14, -6, 19	4.32	14, -6, 19	3.77
<i>Cerebellum—anterior lobe</i>	-32, -50, -27	3.84	26, -53, -26	5.59
<i>Cerebellum—posterior lobe</i>	-7, -69, -23	2.81	6, -69, -23	3.04
<i>Thalamus—anterior</i>	-7, -13, 1	5.62	1, -9, 14	4.19
Condition 2: Low Attention				
<i>Frontal lobe</i>				
Medial frontal gyrus (6/32)	-2, 3, 47	7.05	1, 12, 45	5.99
Middle frontal gyrus (9/46)	-50, 17, 27	3.78	53, 16, 28	2.95
Inferior frontal gyrus (47)	-50, 24, -1	3.98	49, 26, -2	3.11
Insula (13)	-34, 23, 4	4.38	40, 16, 0	4.46
<i>Superior temporal lobe</i>				
Heschl's gyrus (41/42)	-57, -21, 13	3.60	51, -18, 10	8.45
<i>Cerebellum—anterior lobe</i>	-36, -47, -27	5.89	31, -50, -27	3.69
<i>Cerebellum—posterior lobe</i>	-1, -70, -21	8.57	7, -70, -23	2.88
<i>Thalamus—anterior</i>	-8, -9, 9	4.82	5, -11, 12	5.21
Condition 3: NEW words				
<i>Frontal lobe</i>				
Medial frontal gyrus (6/32)	-1, 9, 46	5.94	1, 12, 45	7.31
Middle frontal gyrus (9)	-52, 19, 27	3.26	57, 22, 27	3.17
Inferior frontal gyrus (45)	-59, 22, 7	3.48	49, 26, 0	3.18
Inferior frontal gyrus (47)	—, —, —	—	51, 22, 1	2.89
Insula (13)	-42, 15, 0	3.19	42, 14, -1	6.50
<i>Parietal lobe</i>				
Angular gyrus/middle temp. gyr. (39)	-41, -63, 24	3.76	46, -65, 21	3.39
<i>Superior temporal lobe</i>				
Heschl's gyrus (41/42)	-58, -18, 11	5.54	52, -14, 10	9.09
<i>Cerebellum—anterior lobe</i>	-36, -54, -28	3.21	38, -54, -27	3.00
<i>Cerebellum—posterior lobe</i>	-1, -69, -22	2.83	5, -72, -22	2.30
<i>Thalamus—anterior</i>	-8, -8, 12	4.70	3, -8, 11	3.25

Group-averaged data showing peak intensity values (T_{\max}) and coordinates (mm in Talairach space) for activated clusters (closest Brodmann areas in parentheses) in each of the three conditions compared to the resting state. Corrected activation threshold = $P < 0.05$, $N = 14$. ⁽¹⁾Activity spread into intraparietal sulcus in 58% of listeners. ⁽²⁾Activation in SMG, but peak was more localized to angular gyrus in 29%, and intraparietal sulcus in 21% of listeners.

3.2. fMRI Results. A temporal analysis of the BOLD responses for each stimulus condition was performed by modeling the fMRI time series over a range of time lags, as shown in Figure 2. Although this analysis lacks the temporal precision of electrophysiological methods [1], it can provide valuable information about the relative timing of neural events associated with each test condition [31]. As stated

above, eight separate hemodynamic response models were constructed for each stimulus type, and the peak of each model was time-shifted by one TR in order to capture peak responses across an 8-TR (18.4 s) time window. We then calculated the BOLD responses from the peak activation at each time lag, as shown in Figure 2. As expected, strong BOLD activity was observed in the superior temporal lobe

in the vicinity of primary auditory cortex, and this activity did not vary with the three stimulus conditions (Table 1). In addition to the expected early activity in left superior temporal gyrus, BOLD responses associated with all three stimulus categories were, on average, 20% greater in right superior temporal gyrus, which is consistent with the other right-lateralized activation patterns discussed below.

3.2.1. Segregation of Function in Dorsal and Ventral Prefrontal Cortex. As shown in Figure 2(a), identification of words in all three categories was associated with activity in left middle frontal gyrus (a portion of dorsolateral prefrontal cortex—DLPFC) and this activity was closely mirrored in anterior insular cortex (Figure 2(b)). Furthermore, the time course of the BOLD response in these two left-hemisphere regions was similar for all three word categories regardless of the initial encoding condition. In contrast, right DLPFC showed a distinctly different response pattern, with *High-Attention* words yielding significantly greater responses relative to *Low-Attention* words, and the latter yielding significantly greater responses relative to *NEW* words. This pattern indicates a strong differential effect of attention at encoding in right DLPFC (BA9/46) compared to left DLPFC (Figure 2(a)) and bilateral insula (Figure 2(b)). Importantly, this result is consistent with earlier studies using written words that linked activity in this region of right DLPFC with post-retrieval processing and/or monitoring functions [37, 52–54]. If right DLPFC is indeed involved selectively in post-retrieval processing, we would also expect, in accord with the “Attention-to-Memory” model, that words strongly attended at encoding might evoke a greater memory response than words only weakly attended at encoding. This graded pattern of activity was, in fact, observed for the *High-* and *Low-Attention* words, as shown in Figure 2(a). Conversely, this region would not be expected to activate in response to *NEW* words because these words were not yet encoded into memory. In accord with this hypothesis, only negligible positive BOLD activity was observed in right DLPFC in response to *NEW* words (Figure 2(a), right column).

As shown in Figure 2(c), another bilateral set of clusters was located in the inferior frontal gyrus (IFG). This area is believed to be an important rostral component of the so-called “ventral frontoparietal stream” that is frequently observed in studies involving the detection of novel or low-frequency events, particularly when they are unexpected [3]. Although this pathway has been observed in numerous visuospatial attention studies, our findings using auditory language stimuli are also consistent with this anatomical framework. Figure 2 shows how the three stimulus conditions in our study differentially activated the ventral frontoparietal stream. The focal point for these clusters in the frontal lobes was found in pars orbitalis (BA47), and in accord with the ventral frontoparietal model, activity was also lateralized to the right hemisphere (Figure 2(c), right column). Both the *High-* and *Low-Attention* conditions were associated with comparable activation in left and right IFG. However, in left IFG, a BOLD response on a similar time course was absent for the *NEW* words, indicating that the early response in IFG to previously encoded words may

specifically reflect memory for these previously presented items. Despite the absence of an early BOLD response in left IFG for *NEW* words, these items were instead associated with a late-onset response in the right hemisphere (Figure 2(c), right column). In other words, responses to the *NEW* words differed from the previously attended words both in terms of hemispheric lateralization and in the late time course of activation in IFG (BA47), suggesting that this late activation more likely reflects the initial encoding of *NEW* words into memory.

3.2.2. Segregation of Function in Ventral Posterior Cortex. With respect to language processing, converging evidence suggests a key role for posterior parietal and temporal regions in verbal memory and attention [12, 31, 37, 44]. One memory model predicts that items correctly identified as previously encountered (old) will trigger increased activity in left intraparietal cortex, relative to missed old and correctly rejected novel items [31]. However, as noted above, selective attention is another cognitive function that is commonly associated with posterior parietal cortex, and it remains unclear how the influence of attention may affect the neural networks underlying recollection and familiarity, especially in the auditory domain.

It has been proposed that inferior parietal lobule, a region within ventral posterior parietal cortex, may serve a specialized function in the expression of attention, as proposed in both the Embedded-Processes model [14, 15] and the Attention-to-Memory (AtoM) model [8, 18–20]. These models allowed us to generate informed predictions about the neural substrates underlying a possible attention-dependent memory effect in our study. If ventral posterior parietal cortex serves specifically as the “focus of attention” as in the Embedded-Processes model, we would expect activation in this region to be differentiated according to the level of attention directed toward each stimulus at encoding, as shown in Figure 1. Another prediction is that any region responsible for the *reactivation* of focused attention prior to memory retrieval should display activity earlier in the hemodynamic response than regions responsible for post-stimulus processing steps, such as semantic analysis.

As shown in Figure 2(d) and Table 1, the principal locus of activation in posterior parietal cortex was found bilaterally in the supramarginal gyrus (SMG). This activity, furthermore, was associated specifically with listening to and identifying the *High-Attention* words, but not the *Low-Attention* words. This is consistent with the notion that activation in and around the SMG is related to attentional scanning that placed our *High-Attention* words into the “focus of attention” according to the Embedded-Processes model. To examine the spatial location of the group activation in greater detail, the activation was overlaid onto the individual anatomical scans of each participant. This participant-by-participant anatomical localization revealed that the peak activation was located in the left SMG for 86% (12 of 14 individuals), and in the right SMG for 64% of the participants. As shown in Table 2, activation in the left hemisphere was completely isolated to the SMG in 36% of listeners, whereas in 50%, the activity also spread

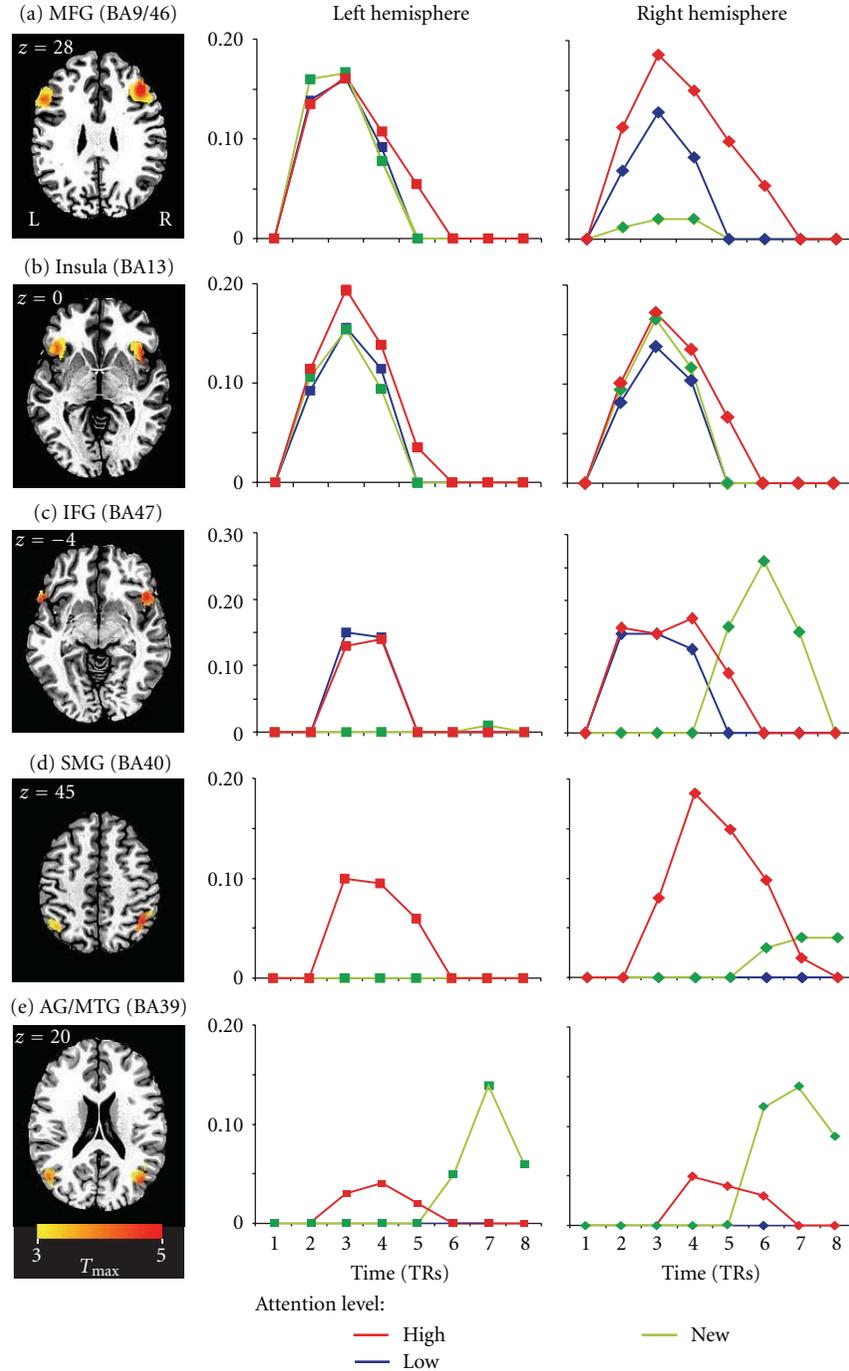


FIGURE 2: Brain regions showing significant group activation ($N = 14$) associated with recognizing previously-encoded words (*High Attention, Low Attention*), and *NEW* words relative to the resting state. At left, axial slices show patterns of BOLD activation in each bilateral ROI overlaid onto canonical anatomical images (z -depth in mm). (a) Middle Frontal Gyrus, MFG; (b) Anterior Insula; (c) Inferior Frontal Gyrus, IFG; (d) Supramarginal Gyrus, SMG; (e) Angular Gyrus/Middle Temporal Gyrus, AG/MTG. To the right of the activation maps in each row are plots of % BOLD signal change associated with memories of items in each word category, plotted over time (8 TRs). Each value represents peak activity extracted from each ROI at each time point, then averaged over the entire group. Only activity peaks that survived correction for multiple comparisons are shown. No standard errors were above 0.01%, therefore error bars were omitted for clarity.

from the SMG into the adjoining intraparietal sulcus (IPS). In contrast, the primary focus of activation in the right hemisphere was more variable, localizing to the SMG for 50% of the participants, to the angular gyrus (AG) for 29%

of participants, and to the IPS for 21% of participants. This greater variation likely reflects the greater variability of the anatomical landmarks within the posterior right hemisphere for our participants.

TABLE 2: Regions of primary focus and secondary spread for parietal ROIs associated with the *High-Attention* condition in each of the 14 subjects.

	Left hemisphere		Right hemisphere	
	Primary focus	Secondary spread	Primary focus	Secondary spread
S1	SMG	IPS	SMG	AG
S2	SMG	—	AG	SMG/IPS
S3	SMG	IPS	SMG	IPS
S4	SPL	IPS	AG	SMG/IPS
S5	SMG	IPS	SMG	IPS
S6	SMG	IPS	SMG	AG
S7	SMG	IPS	SMG	IPS
S8	SMG	IPS	SMG	IPS
S9	IPS	SMG	IPS	SMG
S10	SMG	—	IPS	SMG
S11	SMG	—	SMG	—
S12	SMG	—	IPS	—
S13	SMG	—	AG	SMG/IPS
S14	SMG	IPS	AG	SMG/IPS

AG: angular gyrus (BA39); IPS: intraparietal sulcus (BA40/7); SMG: supramarginal gyrus (BA40); SPL: superior parietal lobule (BA7).

TABLE 3: Regions of primary focus and secondary spread for parietal ROIs associated with the *NEW* condition in each of the 14 subjects.

	Left hemisphere		Right hemisphere	
	Primary focus	Secondary spread	Primary focus	Secondary spread
S1	AG	pMTG	pMTG	—
S2	SMG	AG	pMTG	—
S3	AG	pMTG	pMTG	—
S4	AG	pMTG	pMTG	—
S5	AG	pMTG	pMTG	AG
S6	AG	pMTG	pMTG	—
S7	SMG	AG	AG	pMTG
S8	AG	pMTG	AG	pMTG
S9	AG	pMTG	pMTG	—
S10	SMG	AG	AG	—
S11	AG	pMTG	pMTG	—
S12	SMG	AG	AG	—
S13	SMG	AG	AG	pMTG
S14	AG	pMTG	pMTG	—

AG: angular gyrus (BA39); pMTG: posterior middle temporal gyrus (BA39); SMG: supramarginal gyrus (BA40).

In contrast to these spatial activation patterns for previously encoded *High-Attention* words, the principal locus of activation in posterior parietal cortex for *NEW* words was associated with delayed activity in BA39; specifically AG and the adjacent portion of posterior middle temporal gyrus (pMTG) in both hemispheres (Figure 2(e); Table 3). In the left parietal lobe, the activity was centered within the AG for 64% of the participants (9 of 14 individuals) with additional activity that spreads around the peak activation into pMTG. In the right parietal lobe, the primary focus was centered more often in pMTG (64%) than in AG (36%) (Table 3). Thus the BOLD activation associated with *NEW* words was distinctly more ventral and posterior than the parietal

activation pattern for previously encoded *High-Attention* words. Moreover, the time course of *NEW* word-related activity in right posterior parietal cortex was similar to that observed in right IFG for *NEW* words (compare Figures 2(c) and 2(e)). This finding suggests that the timing of activation in right BA39 may be an important determinant in how different stimulus contexts are represented in this region. The timing difference may reflect two distinct functional roles for this component of the frontoparietal attention network in processing previously encoded (*High-Attention*) words and newly encoded (*NEW*) words. One possibility is that the delayed response in BA39 may reflect the increased time required to access semantic content associated with

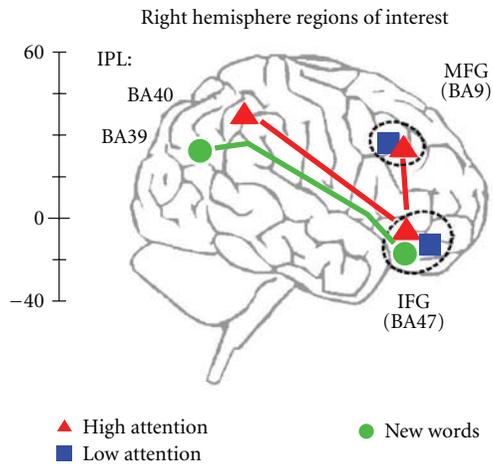


FIGURE 3: Neuroanatomical model of attentive listening networks in the right hemisphere. Both frontoparietal pathways belong to the proposed ventral, “bottom-up”, attention stream (see text). The top stream that processes *High-Attention* words (red pathway) connects BA40 in inferior parietal lobule (IPL) and BA47 in inferior frontal gyrus (IFG). The delayed-activation stream for novel stimuli (green pathway) shows the relationship between BA39 in IPL and BA47 in IFG. The red stream for processing *High-attention* words has another frontal component in right middle frontal gyrus (MFG) that appears to play a role in later postretrieval processing (see text). Responses to the *Low-attention* words were not observed in parietal cortex, but were present in BA9 and BA47, as well as in bilateral anterior insula (BA13) and dorsal anterior cingulate/supplementary motor area (BA6) (Table 1).

these *NEW* words [55], or alternatively, activation may be associated more directly with processing novel events [3].

Based on these results, we propose a neuroanatomical framework that involves two subnetworks in the ventral frontoparietal attention stream, as shown in Figure 3. The relatively early onset activation for previously encoded words in BA40 and BA47 suggests an early attentional role for this subnetwork (Figure 3, *red pathway*), while the delayed time course of *NEW* word responses in BA39 and BA47 could reflect the activity of a separate ventral pathway associated principally with bottom-up processing of novel word stimuli (*green pathway*). Our results are therefore consistent with previous findings that ventral parietal cortex plays a pivotal role in language-based tasks [37, 55–57]. Inferior parietal activity has been found more reliably in studies of working memory than in those of explicit recall [15, 58–60], and our results are also consistent with these findings. Moreover, during the encoding phase of our study (Figure 1), right BA40 was activated at the time our spoken stimulus words were first encoded [39]. The results of the memory phase in the present study show that right BA40 was reactivated at retrieval, a finding strongly in line with the transfer-appropriate processing model [10, 25, 29–31, 61–63]. We propose that early selective activation of right BA40 serves to initiate incidental memory for the studied words by bringing them back into the “focus of attention” that was established earlier during encoding [13, 64–66].

4. Conclusion

In summary, our data support a functional framework in which the brain regions that are engaged in identifying words are also sensitive to the level of attention at the time the words were initially encoded. A structural model generated from our findings is illustrated in Figure 3. Our data are consistent with a “focus of attention” centered in right SMG (BA40), but we also identified a frontal region in right DLPFC (BA9/46) that is consistent with an “Attention-to-Memory” function in postretrieval processing. Other frontal regions (left DLPFC) were insensitive to the attentional manipulation at encoding, responding similarly to both the two studied word categories as well as to the newly encoded (*NEW*) words. The pattern in left DLPFC is therefore consistent with other cognitive processes such as executive control of attention that would not be expected to vary across our word categories. A key result was that once a spoken word was unintentionally encoded, subsequent retrieval of that word varied as a function of the initial level of attention directed toward the word at the time of encoding. This is consistent with the notion that attention is critical not only for efficient encoding of words into memory, but it also helps to preserve salient information that is required for the successful retrieval of that information from memory.

Acknowledgments

This research was supported by a grant from the NIH/ National Institute on Deafness and Other Communication Disorders (K01 DC008812) to Thomas A. Christensen.

References

- [1] M. H. Giard, A. Fort, Y. Mouchetant-Rostaing, and J. Pernier, “Neurophysiological mechanisms of auditory selective attention in humans,” *Frontiers in Bioscience*, vol. 5, pp. D84–D94, 2000.
- [2] C. D. Frith and K. J. Friston, “The role of the thalamus in “top down” modulation of attention to sound,” *NeuroImage*, vol. 4, no. 3, pp. 210–215, 1996.
- [3] M. Corbetta and G. L. Shulman, “Control of goal-directed and stimulus-driven attention in the brain,” *Nature Reviews Neuroscience*, vol. 3, no. 3, pp. 201–215, 2002.
- [4] M. I. Posner and M. K. Rothbart, “Research on attention networks as a model for the integration of psychological science,” *Annual Review of Psychology*, vol. 58, pp. 1–23, 2007.
- [5] A. Messinger, M. A. Lebedev, J. D. Kralik, and S. P. Wise, “Multitasking of attention and memory functions in the primate prefrontal cortex,” *Journal of Neuroscience*, vol. 29, no. 17, pp. 5640–5653, 2009.
- [6] T. A. Polk, R. M. Drake, J. J. Jonides, M. R. Smith, and E. E. Smith, “Attention enhances the neural processing of relevant features and suppresses the processing of irrelevant features in humans: a functional magnetic resonance imaging study of the stroop task,” *Journal of Neuroscience*, vol. 28, no. 51, pp. 13786–13792, 2008.
- [7] W. A. Freiwald, “Attention to objects made of features,” *Trends in Cognitive Sciences*, vol. 11, no. 11, pp. 453–454, 2007.

- [8] R. Cabeza, E. Ciaramelli, I. R. Olson, and M. Moscovitch, "The parietal cortex and episodic memory: an attentional account," *Nature Reviews Neuroscience*, vol. 9, no. 8, pp. 613–625, 2008.
- [9] A. Baddeley, "The magic number and the episodic buffer," *Behavioral and Brain Sciences*, vol. 24, no. 1, pp. 117–118, 2001.
- [10] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual Review of Neuroscience*, vol. 18, pp. 193–222, 1995.
- [11] N. Lavie, "Distracted and confused?: selective attention under load," *Trends in Cognitive Sciences*, vol. 9, no. 2, pp. 75–82, 2005.
- [12] N. Cowan, W. Lichty, and T. R. Grove, "Properties of memory for unattended spoken syllables," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 16, no. 2, pp. 258–269, 1990.
- [13] N. Cowan, "Visual and auditory working memory capacity," *Trends in Cognitive Sciences*, vol. 2, no. 3, pp. 77–78, 1998.
- [14] N. Cowan, "An embedded-processes model of working memory," in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, A. Miyake, Ed., pp. 62–101, Cambridge University Press, New York, NY, USA, 1999.
- [15] J. M. Chein, S. M. Ravizza, and J. A. Fiez, "Using neuroimaging to evaluate models of working memory and their implications for language processing," *Journal of Neurolinguistics*, vol. 16, no. 4-5, pp. 315–339, 2003.
- [16] D. LaBerge, "Sustained attention and apical dendrite activity in recurrent circuits," *Brain Research Reviews*, vol. 50, no. 1, pp. 86–99, 2005.
- [17] N. Cowan, D. Li, A. Moffitt et al., "A neural region of abstract working memory," *Journal of Cognitive Neuroscience*, vol. 23, no. 10, pp. 2852–2863, 2011.
- [18] R. Cabeza, "Role of parietal regions in episodic memory retrieval: the dual attentional processes hypothesis," *Neuropsychologia*, vol. 46, no. 7, pp. 1813–1827, 2008.
- [19] E. Ciaramelli, C. L. Grady, and M. Moscovitch, "Top-down and bottom-up attention to memory: a hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval," *Neuropsychologia*, vol. 46, no. 7, pp. 1828–1851, 2008.
- [20] E. Ciaramelli, C. Grady, B. Levine, J. Ween, and M. Moscovitch, "Top-down and bottom-up attention to memory are dissociated in posterior parietal cortex: neuroimaging and neuropsychological evidence," *Journal of Neuroscience*, vol. 30, no. 14, pp. 4943–4956, 2010.
- [21] M. Wimber, H. J. Heinze, and A. Richardson-Klavehn, "Distinct frontoparietal networks set the stage for later perceptual identification priming and episodic recognition memory," *Journal of Neuroscience*, vol. 30, no. 40, pp. 13272–13280, 2010.
- [22] M. R. Uncapher and A. D. Wagner, "Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory," *Neurobiology of Learning and Memory*, vol. 91, no. 2, pp. 139–154, 2009.
- [23] A. D. Wagner, B. J. Shannon, I. Kahn, and R. L. Buckner, "Parietal lobe contributions to episodic memory retrieval," *Trends in Cognitive Sciences*, vol. 9, no. 9, pp. 445–453, 2005.
- [24] K. L. Vilberg and M. D. Rugg, "Memory retrieval and the parietal cortex: a review of evidence from a dual-process perspective," *Neuropsychologia*, vol. 46, no. 7, pp. 1787–1799, 2008.
- [25] R. Cabeza and L. Nyberg, "Imaging cognition II: an empirical review of 275 PET and fMRI studies," *Journal of Cognitive Neuroscience*, vol. 12, no. 1, pp. 1–47, 2000.
- [26] K. A. Paller and A. D. Wagner, "Observing the transformation of experience into memory," *Trends in Cognitive Sciences*, vol. 6, no. 2, pp. 93–102, 2002.
- [27] D. Gaffan, "Widespread cortical networks underlie memory and attention," *Science*, vol. 309, no. 5744, pp. 2172–2173, 2005.
- [28] R. Henson, "A mini-review of fMRI studies of human medial temporal lobe activity associated with recognition memory," *Quarterly Journal of Experimental Psychology B*, vol. 58, no. 3-4, pp. 340–360, 2005.
- [29] E. I. Skinner and M. A. Fernandes, "Neural correlates of recollection and familiarity: a review of neuroimaging and patient data," *Neuropsychologia*, vol. 45, no. 10, pp. 2163–2179, 2007.
- [30] E. I. Skinner, M. A. Fernandes, and C. L. Grady, "Memory networks supporting effort retrieval success retrieval under conditions of full and divided attention," *Experimental Psychology*, vol. 56, no. 6, pp. 386–396, 2009.
- [31] M. E. Wheeler and R. L. Buckner, "Functional-anatomic correlates of remembering and knowing," *NeuroImage*, vol. 21, no. 4, pp. 1337–1349, 2004.
- [32] B. A. Shaywitz, S. E. Shaywitz, K. R. Pugh et al., "The functional neural architecture of components of attention in language-processing tasks," *NeuroImage*, vol. 13, no. 4, pp. 601–612, 2001.
- [33] A. L. W. Bokde, M. A. Tagamets, R. B. Friedman, and B. Horwitz, "Functional interactions of the inferior frontal cortex during the processing of words and word-like stimuli," *Neuron*, vol. 30, no. 2, pp. 609–617, 2001.
- [34] C. J. Fiebach, J. Rissman, and M. D'Esposito, "Modulation of inferotemporal cortex activation during verbal working memory maintenance," *Neuron*, vol. 51, no. 2, pp. 251–261, 2006.
- [35] J. A. Fiez and S. E. Petersen, "Neuroimaging studies of word reading," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 3, pp. 914–921, 1998.
- [36] A. N. Herbster, M. A. Mintun, R. D. Nebes, and J. T. Becker, "Regional cerebral blood flow during word and nonword reading," *Human Brain Mapping*, vol. 5, no. 2, pp. 84–92, 1997.
- [37] K. B. McDermott, S. E. Petersen, J. M. Watson, and J. G. Ojemann, "A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging," *Neuropsychologia*, vol. 41, no. 3, pp. 293–303, 2003.
- [38] K. Hugdahl, R. Westerhausen, K. Alho, S. Medvedev, M. Laine, and H. Hämäläinen, "Attention and cognitive control: unfolding the dichotic listening story: cognition and neurosciences," *Scandinavian Journal of Psychology*, vol. 50, no. 1, pp. 11–22, 2009.
- [39] T. A. Christensen, S. M. Antonucci, J. L. Lockwood, M. Kittleston, and E. Plante, "Cortical and subcortical contributions to the attentive processing of speech," *Neuroreport*, vol. 19, no. 11, pp. 1101–1105, 2008.
- [40] G. H. Glover and C. S. Law, "Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts," *Magnetic Resonance in Medicine*, vol. 46, no. 3, pp. 515–522, 2001.
- [41] J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain*, Thieme, New York, NY, USA, 1988.
- [42] R. W. Cox, "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages," *Computers and Biomedical Research*, vol. 29, no. 3, pp. 162–173, 1996.
- [43] A. Baddeley, "The fractionation of working memory," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 24, pp. 13468–13472, 1996.
- [44] C. Jacquemot and S. K. Scott, "What is the relationship between phonological short-term memory and speech processing?" *Trends in Cognitive Sciences*, vol. 10, no. 11, pp. 480–486, 2006.

- [45] G. Hickok and D. Poeppel, "Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language," *Cognition*, vol. 92, no. 1-2, pp. 67-99, 2004.
- [46] G. Hickok and D. Poeppel, "The cortical organization of speech processing," *Nature Reviews Neuroscience*, vol. 8, no. 5, pp. 393-402, 2007.
- [47] P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike, "Voice-selective areas in human auditory cortex," *Nature*, vol. 403, no. 6767, pp. 309-312, 2000.
- [48] J. Crinion, R. Turner, A. Grogan et al., "Language control in the bilingual brain," *Science*, vol. 312, no. 5779, pp. 1537-1540, 2006.
- [49] S. K. Scott, C. C. Blank, S. Rosen, and R. J. S. Wise, "Identification of a pathway for intelligible speech in the left temporal lobe," *Brain*, vol. 123, no. 12, pp. 2400-2406, 2000.
- [50] S. K. Scott, "Auditory processing—speech, space and auditory objects," *Current Opinion in Neurobiology*, vol. 15, no. 2, pp. 197-201, 2005.
- [51] A. A. Zekveld, D. J. Heslenfeld, J. M. Festen, and R. Schoonhoven, "Top-down and bottom-up processes in speech comprehension," *NeuroImage*, vol. 32, no. 4, pp. 1826-1836, 2006.
- [52] M. D. Rugg, R. N. A. Henson, and W. G. K. Robb, "Neural correlates of retrieval processing in the prefrontal cortex during recognition and exclusion tasks," *Neuropsychologia*, vol. 41, no. 1, pp. 40-52, 2003.
- [53] P. C. Fletcher, T. Shallice, C. D. Frith, R. S. J. Frackowiak, and R. J. Dolan, "The functional roles of prefrontal cortex in episodic memory: II. Retrieval," *Brain*, vol. 121, no. 7, pp. 1249-1256, 1998.
- [54] K. B. McDermott, T. C. Jones, S. E. Petersen, S. K. Lageman, and H. L. Roediger, "Retrieval success is accompanied by enhanced activation in anterior prefrontal cortex during recognition memory: an event-related fMRI study," *Journal of Cognitive Neuroscience*, vol. 12, no. 6, pp. 965-976, 2000.
- [55] M. L. Seghier, E. Fagan, and C. J. Price, "Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network," *Journal of Neuroscience*, vol. 30, no. 50, pp. 16809-16817, 2010.
- [56] J. T. Becker, D. K. MacAndrew, and J. A. Fiez, "A comment on the functional localization of the phonological storage subsystem of working memory," *Brain and Cognition*, vol. 41, no. 1, pp. 27-38, 1999.
- [57] S. M. Ravizza, M. R. Delgado, J. M. Chein, J. T. Becker, and J. A. Fiez, "Functional dissociations within the inferior parietal cortex in verbal working memory," *NeuroImage*, vol. 22, no. 2, pp. 562-573, 2004.
- [58] V. J. Schmithorst, S. K. Holland, and E. Plante, "Diffusion tensor imaging reveals white matter microstructure correlations with auditory processing ability," *Ear and Hearing*, vol. 32, pp. 156-167, 2011.
- [59] M. E. Wheeler and R. L. Buckner, "Functional dissociation among components of remembering: control, perceived oldness, and content," *Journal of Neuroscience*, vol. 23, no. 9, pp. 3869-3880, 2003.
- [60] J. A. Fiez, E. A. Raife, D. A. Balota, J. P. Schwarz, M. E. Raichle, and S. E. Petersen, "A positron emission tomography study of the short-term maintenance of verbal information," *Journal of Neuroscience*, vol. 16, no. 2, pp. 808-822, 1996.
- [61] T. Shallice, P. Fletcher, C. D. Frith, P. Grasby, R. S. J. Frackowiak, and R. J. Dolan, "Brain regions associated with acquisition and retrieval of verbal episodic memory," *Nature*, vol. 368, no. 6472, pp. 633-635, 1994.
- [62] J. Jonides, E. H. Schumacher, E. E. Smith et al., "The role of parietal cortex in verbal working memory," *Journal of Neuroscience*, vol. 18, no. 13, pp. 5026-5034, 1998.
- [63] R. J. Zatorre, M. Bouffard, P. Ahad, and P. Belin, "Where is 'where' in the human auditory cortex?" *Nature Neuroscience*, vol. 5, no. 9, pp. 905-909, 2002.
- [64] J. D. Johnson, S. G. R. McDuff, M. D. Rugg, and K. A. Norman, "Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis," *Neuron*, vol. 63, no. 5, pp. 697-708, 2009.
- [65] N. Cowan and N. L. Wood, "Constraints on awareness, attention, processing, and memory: some recent investigations with ignored speech," *Consciousness and Cognition*, vol. 6, no. 2-3, pp. 182-203, 1997.
- [66] N. Cowan, "Activation, attention, and short-term memory," *Memory and Cognition*, vol. 21, no. 2, pp. 162-167, 1993.

Review Article

MRI Texture Analysis in Multiple Sclerosis

Yunyan Zhang^{1,2,3}

¹ Department of Radiology, University of Calgary, Calgary, AB, Canada T2N 1N4

² Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada T2N 1N4

³ Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada T2N 1N4

Correspondence should be addressed to Yunyan Zhang, yunyzhan@ucalgary.ca

Received 16 March 2011; Accepted 6 September 2011

Academic Editor: Eric C. Bourekas

Copyright © 2012 Yunyan Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiple sclerosis (MS) is a complicated disease characterized by heterogeneous pathology that varies across individuals. Accurate identification and quantification of pathological changes may facilitate a better understanding of disease pathogenesis and progression and help identify novel therapies for MS patients. Texture analysis evaluates interpixel relationships that generate characteristic organizational patterns in an image, many of which are beyond the ability of visual perception. Given its promise detecting subtle structural alterations texture analysis may be an attractive means to evaluate disease activity and evolution. It may also become a new tool to assess therapeutic efficacy if technique issues are resolved and pathological correlates are further confirmed. This paper describes the concept, strategies, and considerations of MRI texture analysis; summarizes applications of texture analysis in MS as a measure of tissue integrity and its clinical relevance; then discusses potentially future directions of texture analysis in MS.

1. Introduction

Multiple Sclerosis (MS) is characterized by heterogeneous histopathology including inflammatory infiltrates, demyelination, remyelination, axonal damage, and gliosis [1]. Consequences of irreversible structural injury eventually lead to progressive physical disability and functional impairment [2]. T2 lesion number and volume are commonly used to evaluate disease activity and burden [3], which however are pathologically nonspecific and correlate only moderately with clinical outcomes. Accurate identification and quantification of pathological changes may facilitate a better understanding of disease pathogenesis and progression and help identify novel therapies for MS patients.

Structural abnormalities that appear regular may be extracted by visual inspection while complex patterns of pathology that are commonly encountered in medical images are difficult to interpret and require the employment of advanced analysis techniques [4]. As an emerging quantitative approach, texture analysis demonstrates promise to detect subtle structural alterations that are not perceivable on conventional magnetic resonance imaging (MRI). This paper describes the concept, strategies, and considerations of MRI

texture analysis; summarizes the potential of texture analysis as a measure of tissue structural property and the clinical relevance; then discusses possible future directions of MRI texture analysis in MS.

2. The Concept

Texture analysis is an image postprocessing approach that extracts quantitative information from a digital image based on mathematical analysis: it can be applied to any image and is used in fields as diverse as medicine and geology [5–8] (Figure 1). A two-dimensional (2D) MR image is a digitized picture of elements (pixels), characterized by spatial location and gray-level intensities. MRI texture analysis evaluates the organizational pattern of image pixels that is unique to the underlying substrates in a tissue. Texture features are, in fact, mathematical parameters that describe the distribution of gray-level intensities to reflect the structural regularity of the imaged tissue [9]. Consequently the structural property of a tissue determines the nature of its texture, which is inevitably affected by histopathological development. Intuitively, texture can be described as smooth or rough, regular or irregular, fine or coarse.

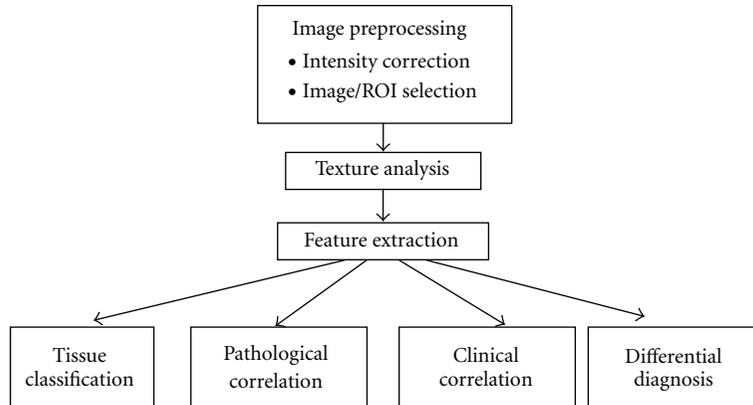


FIGURE 1: Overview of general texture analysis pipelines in the MRI of MS.

3. Approaches

Numerous strategies have been proposed for the examination of image texture. Depending on the particular method used to assess interpixel relationships, a variety of dedicated texture features can be derived. Both statistical and spatial frequency-based approaches are used preferably in neurological imaging applications including MS, which are described below (Table 1). Note that the mathematical basis, calculation procedure, and feature extraction strategies of these texture analysis methods have been discussed comprehensively in previous publications and are not discussed in detail in this paper.

3.1. Statistical Approach. This approach attempts to characterize image texture using statistical parameters. First-order statistics (i.e., mean and variance of gray level, mean and variance of gradient) provides a global assessment of pixel distribution and is relatively intuitive and self-explaining. Second- or higher-order statistics characterizes local texture properties in an image which is based primarily on the co-occurrence matrix (GLCM) [7] and run-length matrix (RLM) [10].

The GLCM method has been investigated heavily since its first introduction by Haralick et al. in 1973 [7] and has demonstrated considerable promise in MRI texture analysis. In brief, the GLCM method computes the joint probability of two pixels which have particular (cooccurring) gray-level values, with a distance d ($d = 1, \dots, n$ pixels in a dimension) apart along a given direction (0, 45, 90, and 135 degrees) (Figure 2). Fourteen parameters can be derived from each GLCM, which collectively reflect the homogeneity, local contrast, correlation, and complexity of an image.

The RLM method [9, 10] quantifies image texture in a similar manner as the GLCM approach. Basically, a gray level run dictates the number of times two or more pixels having the same value in a preset direction, and the RLM is the matrix of run-length frequency occurring in an image in each (generally 4) direction considered. Features derived from RLM represent fine (long runs) or coarse texture (short runs) of an image.

3.2. Spatial Frequency-Based Approach. Pixel patterns that create a unique texture in an image also generate a unique frequency distribution at the spatial scale of the pattern. Specifically, fast changing gray level values represent high frequency content. Conversely, low frequencies relate to slow changing or no change gray-level values [9]. Ideally, the frequency content held in an image can be calculated using the Fourier transform (FT). However, the FT cannot isolate frequency profiles specific to individual spatial locations [11, 12]. Moreover, the determination of slow or fast gray-value changes also depends on the scale that is utilized to examine the image. For instance, an image area would demonstrate little variations if analyzed by a large scale (or from a far distance) whereas an area would yield detailed information if identified by a small scale (or from a close distance).

A few advanced transforms have been developed trying to characterize localized frequencies specific to each pixel location. One of the versatile approaches is the wavelet transform (WT) which characterizes multiscale frequency content (i.e., wavelet coefficients) at each spatial location of an image [13]. This flexibility allows the WT to decompose image texture scale specifically which is ideal for a tissue of heterogeneous pathology. Nonetheless the WT is computationally expensive and lack of intuitive [8]; it has yet no direct applications in MS. The Stockwell transform (ST) is a recent advancement in spatial frequency analysis [11, 14]. While analog to the WT for its multiresolution ability, the ST is a Fourier-based analysis that provides unique frequency spectra (Figure 3) at all pixel locations of an image. Moreover, rotationally invariant information can be obtained by eliminating the angular variance of image texture using the polar form of ST (PST) [15], which is a gifted advantage as medical images are prone to movement artifacts (Figure 3).

4. Considerations

While texture analysis is potentially a new tool to identify subtle structural changes in a tissue, there are concerns re In theory, differences in acquisition parameters, imaging sequences, and the homogeneity of radio-frequency (RF) field cause alterations in pixel arrangements and therefore

TABLE 1: Overview of common texture analysis approaches in MS.

	Assessment	Utility
Statistical approach		
First-order	Global assessment of pixel distribution	Self-explanatory yet lack of detail
Second-order		
Gray-level cooccurrence matrix (GLCM)	Joint probability of two pixels having cooccurring gray level values at a given distance and direction	Multiple properties of a texture (coarseness, correlation, contrast), less sensitive to larger scales
Run length matrix (RLM)	The number of times two or more pixels having the same value in a preset direction	Several properties of a texture (coarseness), less sensitive to larger scales
Spectral approach		
Fourier transform	Entire frequency profile, using sinusoid basis functions	Useful for signals without temporal changes
Wavelet transform	Scale-based frequency content, using a deformable localizing “mother” wavelet as basis function	Multiscale analysis; less intuitive and can be computation-expensive
Stockwell transform	Scale-based frequency content, using fast Fourier transform and a flexible Gaussian localizing window	Fourier-based multiscale frequency content; computation time varies by image size and algorithm

MRI texture. This potential limit is drawing much attention and has partially confounded the clinical use of texture analysis. On the other hand, few recent studies argued that MRI texture analysis is relatively tolerant to the imaging variables and support further investigation [16–18].

Mayerhoefer et al. [16] assessed the impact of variance in acquisition parameters (number of acquisitions, repetition time, echo time, sampling bandwidth, and spatial resolution) to texture indices and pattern discrimination. Texture features were increasingly sensitive to the variation of acquisition parameters in images with increasingly spatial resolution. However, the variation in imaging parameters had little impact on texture measurements if the image had sufficient spatial resolution. Meanwhile, GLCM-derived texture parameters outperformed the other statistical and WT ones. In another study of similar purpose, Harrison et al. [17] evaluated whether texture analysis was sensitive to image acquisition and processing protocols by assessing differences between three types of imaging sequences, two anatomical levels of interest, three sequential slices, and two methods of delineation of regions of interest (ROIs). A total of 280 statistical and WT-based parameters were extracted from lesions, WM regions further away from lesions (WM), WM regions adjacent to lesions (NAWM), cerebrospinal fluid (CSF), and basal ganglia of 38 patients with either MS or clinically isolated syndrome (CIS). The authors showed that MRI texture analysis provided an excellent distinction between tissues of interest (96–100% accuracy). There was no significant difference in the results of texture-based classification between imaging sequences, anatomical levels, or between temporal imaging slices within tissue. Moreover, Savio et al. [18] have studied the effect of slice thickness to texture analysis of MS lesions and the NAWM. The signal intensity of three 1-mm consecutive slices was averaged to simulate a 3-mm slice that is commonly utilized in clinical imaging. There were moderate differences in the distribution

of texture values between 1-mm and 3-mm slices, which nonetheless did not compromise the classification results (lesion versus NAWM) even using different slice thickness between training and test datasets.

The robustness of texture analysis seems to be indicated by these investigations. However, the impact of imaging factors was assessed based on the accuracy of texture-associated classifiers. While it is a popular approach in computer-assisted diagnosis, the classification step may have complicated the evaluation process and added difficulty for interpretation. Direct comparison of texture characteristics may be helpful in the future to clarify this issue. Furthermore, it has not been fully investigated whether texture measures are portable between imaging centers; some researchers [19, 20] suggest that the use of test objects at different MRI scanners may be a practical means for feature normalization.

5. Applications

5.1. Characterization of Structure Properties. While pathologically different, the activity of many MS lesions cannot be distinguished on conventional MRI based on their appearances. Using a texture-based segmentation approach, Yu et al. [21] examined the feasibility of texture analysis to differentiate activity levels of 32 lesions from 8 relapsing-remitting (RR) MS. Forty-two features were extracted from first- and higher-order (GLCM and RLM) statistics from each lesion. By referencing the status of gadolinium (Gd)-enhancement, texture analysis based on T2-weighted MRI (T2W) allowed an accurate classification of both active (88%) and nonactive lesions (96%). As an exception to most texture analysis studies, features from RLM demonstrated better discrimination potential than that from GLCM. In addition, new evidence showed that a texture-based pattern recognition system was able to differentiate MS lesions from other neurological abnormalities on MRI [22].

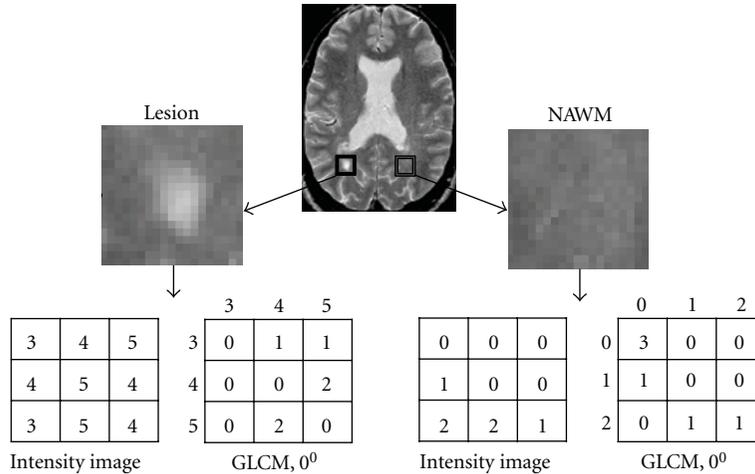


FIGURE 2: Schematic demonstration of texture analysis in the T2-weighted MR image from an MS patient (top) using a statistical method (gray-level cooccurrence matrix, GLCM). Areas of a focal lesion (box in the right) and the contralateral NAWM are delineated respectively, where texture analysis is conducted. Shown are different GLCMs computed at 0 degree based on the small sample subregions from the lesion (bottom right) and the NAWM of the patient.

Despite the dominance of multifocal pathology, diffuse abnormalities are observed in the normal appearing white matter (NAWM) and gray matter (NAGM) of MS patients [23]. Recently, J. Zhang et al. [24] assessed the texture property of MS lesions, NAWM, and normal WM from 16 patients and 16 controls and compared the segmentation ability of joint statistical and spectral features (>200) with that of GLCM features alone. The classification accuracy based on combined sets of texture parameters was superior (100%) compared to GLCM features only between MS lesions and NAWM or normal WM; however, the classification power was compromised (58.33%) in an attempt to differentiate NAWM from normal WM. The disadvantage of statistical texture analysis to detect NAWM abnormality was also implicated in a previous study [25], which compared the texture of frontal NAWM in 41 MS patients with the texture of normal WM at the same region of the brain in 10 controls. No texture difference was identified, which was presumably due to limited resolution of MR images.

Great potential of texture analysis has also been demonstrated using spectral features derived from the PST. In a longitudinal study, Y. Zhang et al. [26] analyzed the texture of new enhancing lesions in comparison with chronic lesions and the NAWM on T2W MRI from 10 RRMS patients. The texture in contrast-enhancing lesions was significantly coarser than that in chronic lesions or the NAWM on the same imaging slice; it improved gradually thereafter only in the acute lesions. The texture in chronic lesions remained relatively stable over time, as did the texture in the NAWM. This study also showed that the recovery of acute lesions seemed to be associated with the degree of coarse texture during enhancement. Using a similar method, Y. Zhang et al. [27] investigated whether texture was different between new acute T1 hypointense lesions (acute black holes, ABHs) that persist and those that resolve over time. The ABHs were classified as transient (tABHs), suggesting

repair [28], or persistent (pABHs), reflecting destruction [29] based on their eventual T1 MRI appearance (isointense or hypointense) 5–8 months later. The tABHs demonstrated significantly finer texture than the pABHs when first seen on T1W MRI. These intriguing data suggest that inflammatory demyelination generates a heterogeneous signal manifesting as coarse texture, whereas organized tissue generates a homogeneous signal manifesting as fine texture (the texture of repairing tissue or the NAWM).

5.2. Clinical Relevance. The demonstration of clinical relevance is important for a potential imaging measure as part of the validation process. Mathias et al. [30] performed texture analysis on T2W MRI of the spinal cord from 40 MS patients (10 each of RRMS, primary progressive MS, secondary progressive MS, and benign MS) and 10 controls. Eight texture features (4 first-order and 4 GLCM statistics) were extracted from the entire cross-sectional area of the segmented spinal cord. Mean texture features were significantly different between all patients and normal controls and between normal controls and patient subgroups except for benign MS. In particular, while not all significant, MRI texture was generally coarser (i.e., increased entropy and decreased angular second moment) in MS patients than in controls, which was identified before detectable spinal cord atrophy. Moreover, two texture features (mean gradient and mean gray level) correlated significantly with disability as assessed by the Expanded Disability Status Scale (EDSS) in MS subjects. The authors suggested that texture analysis might be a potential tool to monitor changes associated with patient disability; however, the reproducibility and sensitivity of texture measurements must be improved. Similarly, by analyzing the GLCM texture of magnetization transfer ratio (MTR, a potential measure of tissue injury) maps of the brain from 23 controls, 38 patients with CIS and 35 patients with MS, Tozer et al. [31] found that

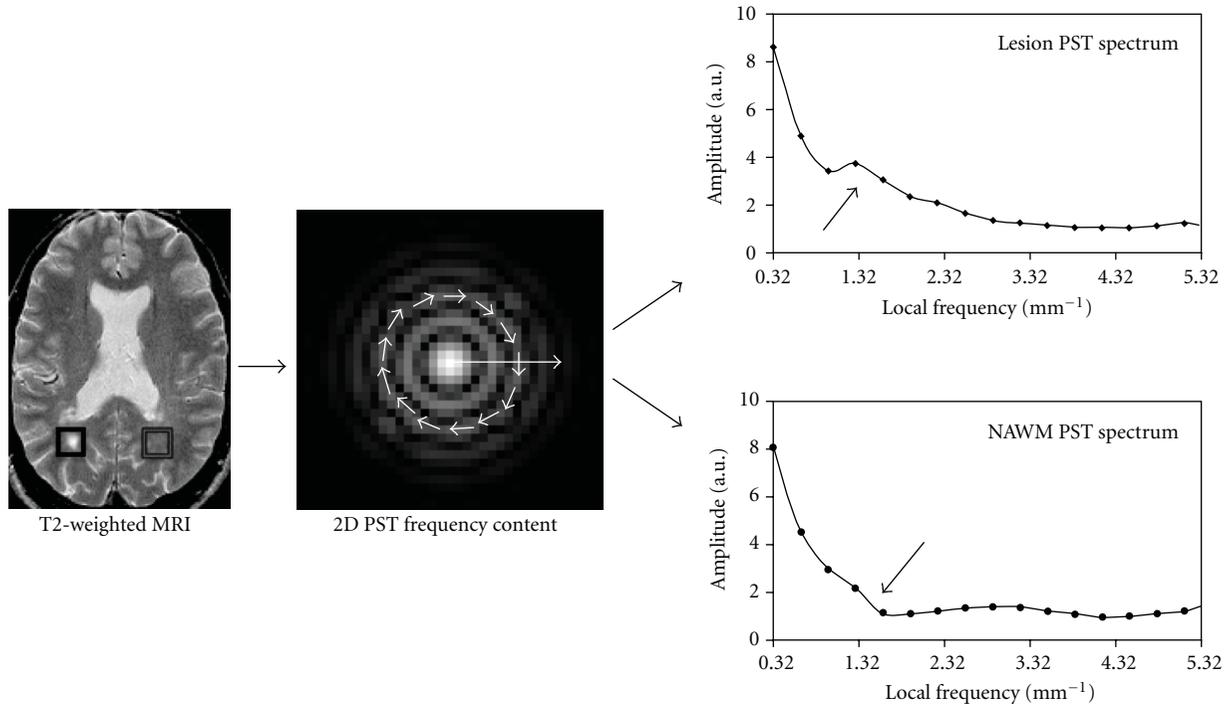


FIGURE 3: Texture analysis of the same image demonstrated in Figure 2 using a spectral method (polar Stockwell transform, PST). The PST spectra were first calculated as radius and orientation in polar coordinate (middle figure with arrows in circle); then a 1-dimensional spectrum (texture curve) was obtained by integrating frequencies along the radial direction for each frequency and pixel. Plots in the right side of the figure demonstrate the average texture curve of central 5×5 pixels in the lesion (right box) and the contralateral NAWM of the patient. Note that the amplitude of the low-frequency range appears greater (coarser) in the lesion than in the NAWM (arrows).

the texture in MS differed from CIS and controls but did not differ between the later two groups. Furthermore, GM MTR texture correlated with neurological disability and WM MTR texture was associated with cognitive measures in MS. In the future, more studies of texture analysis for its capacity to characterize GM pathology would be desirable as conventional MRI is suboptimal to define GM plaques, which however demonstrated considerable pathological and clinical relevance [32, 33].

In a more recent study, Loizou et al. [34] assessed the texture of MS lesions and the NAWM from 38 patients with CIS and normal WM from 20 volunteers and examined the relationship between texture severity of MS lesions and disease progression over approximately 2 years. T2 MRI texture was analyzed using an amplitude modulation-frequency modulation (AM-FM) method, a technique with similarities to the multiscale spectral analysis. Consistent with other published reports, there was significant texture difference between lesions, NAWM, and normal WM. More interestingly, lesions of coarser texture at baseline associated with more severe disability (EDSS > 2) accumulated over 2 years while lesions of finer texture at baseline were found in patients who developed less disability (EDSS ≤ 2). The AM-FM classifier based on medium frequency instantaneous amplitude (relatively coarse texture) provided the best segmentation results to these lesions (area under the ROC curve = 0.76). The AM-FM features were proposed as

potential measures of lesion load and disability progression in MS patients.

6. Future Directions

6.1. Pathological Correlates. It is hypothesized that pathological processes induce ultrastructural changes on the nanometer to micrometer scale, which manifest as pixel pattern changes on the millimeter scale of MR images. While it is not validated in human subjects, pathological correlates of texture analysis has been demonstrated in animal models of MS. Y. Zhang et al. [35] analyzed the texture of MS-like lesions on 7T T2W MRI in the spinal cord of mice with experimental allergic encephalomyelitis (EAE, a model of MS) using the PST. Increased PST texture was observed in EAE lesions compared with the control tissue, which corresponded with inflammation and demyelination. Moreover, texture analysis was evidenced to detect remyelination in a cuprizone mouse model of demyelination [36]. In that study, each mouse was fed either with cuprizone to induce demyelination or with normal diet as controls for 8 weeks. Texture analysis was performed before (day 0) and 13, 29, 32, 41, and 56 days on treatment. Yu et al. [36] showed that texture discrimination functions classified myelinated (day 0) and demyelinated (day 56) brain regions with near-perfect accuracy (95% and 98%, resp.). Furthermore, one of the texture parameters from the RLM, the horizontal gray-level nonuniformity

(HGLNU), varied in concordance with changes in the myelination status of four brain regions: the HGLNU improved when reported remyelination occurred. While this study was limited by correlating indirectly with histological data inferred from literature, texture analysis appears to be a potential measure of myelin integrity. Nonetheless, further validation in MS subjects is necessary for a better understanding and interpretation of texture measurements in a clinical context. As a common validating system for MRI techniques, postmortem samples could be a great candidate; in that way, pathological specificity of texture analysis can be also investigated.

6.2. Evaluating Treatment Impact. Given its potential as a sensitive measure of tissue integrity and the clinical relevance, MRI texture analysis may be used to evaluate treatment impact in both overt MS lesions and in the disease prone NAWM. In this way, an integrated picture of therapeutic efficacy can be obtained beyond the detectability of lesion number and volume. In a feasibility study of 5 RRMS patients with active lesions, Y. Zhang et al. [37] investigated the texture of acute and chronic MS lesions and the NAWM before and after minocycline treatment using the GLCM method. Compared with inactive lesions and the NAWM, active lesions showed the greatest texture abnormality and exhibited the largest texture changes over 6 months, although statistical analysis was not significant. Recently, the promise of MRI texture analysis to identify treatment effect has been demonstrated in patients with non-Hodgkin lymphoma; [38] however, this is subject to further investigation in MS subjects.

7. Conclusions

MRI texture analysis demonstrates great potential in the study of MS. As an image processing strategy, texture analysis shows promise to extract clinically meaningful information from routine conventional MRI. Given its clinical relevance, texture analysis may be an attractive means to characterize disease activity and progression. However, the robustness of texture analysis to imaging protocols must be clarified before a significant clinical effect can be established. Furthermore, despite the promising outcomes in animal models, pathological correlates of texture analysis are subject to confirmation in MS subjects, whereby the role of texture analysis can be further tested as a potential tool of evaluating treatment benefits for MS patients.

References

- [1] A. Compston and A. Coles, "Multiple sclerosis," *The Lancet*, vol. 372, no. 9648, pp. 1502–1517, 2008.
- [2] C. Bjartmar and B. D. Trapp, "Axonal degeneration and progressive neurologic disability in multiple sclerosis," *Neurotoxicity Research*, vol. 5, no. 1-2, pp. 157–164, 2003.
- [3] D. K. Li and D. W. Paty, "Magnetic resonance imaging results of the PRISMS trial: a randomized, double-blind, placebo-controlled study of interferon- β 1a in relapsing-remitting multiple sclerosis. Prevention of relapses and disability by interferon- β 1a subcutaneously in multiple sclerosis," *Annals of Neurology*, vol. 46, no. 2, pp. 197–206, 1999.
- [4] G. D. Tourassi, "Journey toward computer-aided diagnosis: role of image texture analysis," *Radiology*, vol. 213, no. 2, pp. 317–320, 1999.
- [5] E. M. Darling and R. D. Joseph, "Pattern recognition from satellite altitudes," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, pp. 38–47, 1968.
- [6] E. L. Hall, R. P. Kruger, S. J. Dwyer, D. L. Hall, R. W. McLaren, and G. S. Lodwick, "A survey of preprocessing and feature extraction techniques for radiographic images," *IEEE Transactions on Computers*, vol. 20, no. 9, pp. 1032–1044, 1971.
- [7] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.
- [8] A. Kassner and R. E. Thornhill, "Texture analysis: a review of neurologic MR imaging applications," *American Journal of Neuroradiology*, vol. 31, no. 5, pp. 809–816, 2010.
- [9] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images," *Clinical Radiology*, vol. 59, no. 12, pp. 1061–1069, 2004.
- [10] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 172–179, 1975.
- [11] H. Zhu, Y. Zhang, X. Wei, L. M. Metz, and J. R. Mitchell, "MR multi-spectral texture analysis using space-frequency information," in *Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS '04)*, pp. 173–179, June 2004.
- [12] R. A. Brown and R. Frayne, "A comparison of texture quantification techniques based on the Fourier and S transforms," *Medical Physics*, vol. 35, no. 11, pp. 4998–5008, 2008.
- [13] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [14] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: the S transform," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 998–1001, 1996.
- [15] Y. Zhou, *Boundary detection in petrographic images and applications of S-transform space-wavenumber analysis to image processing for texture definition*, Ph.D. thesis, The University of Western Ontario, 2002.
- [16] M. E. Mayerhoefer, P. Szomolanyi, D. Jirak, A. Materka, and S. Trattnig, "Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study," *Medical Physics*, vol. 36, no. 4, pp. 1236–1243, 2009.
- [17] L. C. Harrison, M. Raunio, K. K. Holli et al., "MRI texture analysis in multiple sclerosis: toward a clinical analysis protocol," *Academic Radiology*, vol. 17, no. 6, pp. 696–707, 2010.
- [18] S. J. Savio, L. C. Harrison, T. Luukkaala et al., "Effect of slice thickness on brain magnetic resonance image texture analysis," *BioMedical Engineering Online*, vol. 9, article 60, 2010.
- [19] S. Herlidou-Même, J. M. Constans, B. Carsin et al., "MRI texture analysis on texture test objects, normal brain and intracranial tumors," *Magnetic Resonance Imaging*, vol. 21, no. 9, pp. 989–993, 2003.
- [20] R. A. Lerski, L. R. Schad, R. Luypaert et al., "Multicentre magnetic resonance texture analysis trial using reticulated foam test objects," *Magnetic Resonance Imaging*, vol. 17, no. 7, pp. 1025–1031, 1999.

- [21] O. Yu, Y. Mauss, G. Zollner, I. J. Namer, and J. Chambron, "Distinct patterns of active and non-active plaques using texture analysis on brain NMR images in multiple sclerosis patients: preliminary results," *Magnetic Resonance Imaging*, vol. 17, no. 9, pp. 1261–1267, 1999.
- [22] P. Theocharakis, D. Glotsos, I. Kalatzis et al., "Pattern recognition system for the discrimination of multiple sclerosis from cerebral microangiopathy lesions based on texture analysis of magnetic resonance images," *Magnetic Resonance Imaging*, vol. 27, no. 3, pp. 417–422, 2009.
- [23] D. H. Miller, A. J. Thompson, and M. Filippi, "Magnetic resonance studies of abnormalities in the normal appearing white matter and grey matter in multiple sclerosis," *Journal of Neurology*, vol. 250, no. 12, pp. 1407–1419, 2003.
- [24] J. Zhang, L. Tong, L. Wang, and N. Li, "Texture analysis of multiple sclerosis: a comparative study," *Magnetic Resonance Imaging*, vol. 26, no. 8, pp. 1160–1166, 2008.
- [25] C. Gasperini, M. A. Horsfield, J. W. Thorped et al., "Macroscopic and microscopic assessments of disease burden by MRI in Multiple Sclerosis: relationship to clinical parameters," *Journal of Magnetic Resonance Imaging*, vol. 6, no. 4, pp. 580–584, 1996.
- [26] Y. Zhang, H. Zhu, J. R. Mitchell, F. Costello, and L. M. Metz, "T2 MRI texture analysis is a sensitive measure of tissue injury and recovery resulting from acute inflammatory lesions in multiple sclerosis," *NeuroImage*, vol. 47, no. 1, pp. 107–111, 2009.
- [27] Y. Zhang, A. Traboulsee, Y. Zhao, L. M. Metz, and D. K. Li, "Texture analysis differentiates persistent and transient T1 black holes at acute onset in multiple sclerosis: a preliminary study," *Multiple Sclerosis*, vol. 17, no. 5, pp. 532–540, 2011.
- [28] A. Bitsch, T. Kuhlmann, C. Stadelmann, H. Lassmann, C. Lucchinetti, and W. Brück, "A longitudinal MRI study of histopathologically defined hypointense multiple sclerosis lesions," *Annals of Neurology*, vol. 49, no. 6, pp. 793–796, 2001.
- [29] J. H. Van Waesberghe, W. Kamphorst, C. J. de Groot et al., "Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability," *Annals of Neurology*, vol. 46, no. 5, pp. 747–754, 1999.
- [30] J. M. Mathias, P. S. Tofts, and N. A. Losseff, "Texture analysis of spinal cord pathology in multiple sclerosis," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 929–935, 1999.
- [31] D. J. Tozer, G. Marongiu, J. K. Swanton, A. J. Thompson, and D. H. Miller, "Texture analysis of magnetization transfer maps from patients with clinically isolated syndrome and multiple sclerosis," *Journal of Magnetic Resonance Imaging*, vol. 30, no. 3, pp. 506–513, 2009.
- [32] G. R. Campbell, I. Ziabreva, A. K. Reeve et al., "Mitochondrial DNA deletions and neurodegeneration in multiple sclerosis," *Annals of Neurology*, vol. 69, no. 3, pp. 481–492, 2011.
- [33] M. Calabrese, F. Rinaldi, P. Grossi, and P. Gallo, "Cortical pathology and cognitive impairment in multiple sclerosis," *Expert Review of Neurotherapeutics*, vol. 11, no. 3, pp. 425–432, 2011.
- [34] C. P. Loizou, V. Murray, M. S. Pattichis, I. Seimenis, M. Pantziaris, and C. S. Pattichis, "Multiscale amplitude-modulation frequency-modulation (AM-FM) texture analysis of multiple sclerosis in brain MRI images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 119–129, 2010.
- [35] Y. Zhang, J. Wells, R. Buist, J. Peeling, V. W. Yong, and J. R. Mitchell, "A novel MRI texture analysis of demyelination and inflammation in relapsing-remitting experimental allergic encephalomyelitis," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9, no. 1, pp. 760–767, 2006.
- [36] O. Yu, J. Steibel, Y. Mauss et al., "Remyelination assessment by MRI texture analysis in a cuprizone mouse model," *Magnetic Resonance Imaging*, vol. 22, no. 8, pp. 1139–1144, 2004.
- [37] Y. Zhang, H. Zhu, R. Ferrari et al., "Texture analysis of MR images of minocycline treated MS patients," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9, pp. 786–793, 2003.
- [38] L. C. Harrison, T. Luukkaala, H. Pertovaara et al., "Non-Hodgkin lymphoma response evaluation with MRI texture classification," *Journal of Experimental and Clinical Cancer Research*, vol. 28, no. 1, article 87, 2009.

Research Article

Landmark Optimization Using Local Curvature for Point-Based Nonlinear Rodent Brain Image Registration

Yutong Liu,¹ Balasrinivasa R. Sajja,¹ Mariano G. Uberti,¹
Howard E. Gendelman,² Tammy Kielian,³ and Michael D. Boska¹

¹Department of Radiology, University of Nebraska Medical Center, 981045 Nebraska Medical Center, Omaha, NE 68198, USA

²Department of Pharmacology and Experimental Neuroscience and Center for Neurodegenerative Disorders, University of Nebraska Medical Center, Omaha, NE 68198, USA

³Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE 68198, USA

Correspondence should be addressed to Yutong Liu, yutongliu@unmc.edu

Received 25 January 2011; Revised 20 July 2011; Accepted 22 July 2011

Academic Editor: Ramin Parsey

Copyright © 2012 Yutong Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. To develop a technique to automate landmark selection for point-based interpolating transformations for nonlinear medical image registration. **Materials and Methods.** Interpolating transformations were calculated from homologous point landmarks on the source (image to be transformed) and target (reference image). Point landmarks are placed at regular intervals on contours of anatomical features, and their positions are optimized along the contour surface by a function composed of curvature similarity and displacements of the homologous landmarks. The method was evaluated in two cases ($n = 5$ each). In one, MRI was registered to histological sections; in the second, geometric distortions in EPI MRI were corrected. Normalized mutual information and target registration error were calculated to compare the registration accuracy of the automatically and manually generated landmarks. **Results.** Statistical analyses demonstrated significant improvement ($P < 0.05$) in registration accuracy by landmark optimization in most data sets and trends towards improvement ($P < 0.1$) in others as compared to manual landmark selection.

1. Introduction

In rodent brains images may become distorted due to instrument imperfections or, in the case of histology, tissue processing. For example, magnetic field inhomogeneity causes geometric distortions in echo planar imaging (EPI) MRI; mechanical forces acting on a harvested brain during slicing may cause tissue tearing. And chemical preparation for histological analysis may cause deformation of tissue, evident in histological micrographs. Correcting (or “rectifying,” “warping”) the distorted images is required to adequately represent rodent brains (in the case of EPI distortion in MRI), or to compare different acquisition modalities (e.g., comparison of in vivo images and histological micrographs). Normally, affine (linear + translation) transformation cannot reconcile severe distortions making nonlinear transformation necessary.

Among the nonlinear transformation techniques, point-based interpolating transformation techniques are widely

employed because they are easy to implement and flexible for different applications [1–3]. A typical point-based interpolating approach is comprised of three steps: (1) placing homologous point landmarks on the source image (image to be transformed) and the target image (image used as the reference), respectively, (2) computing the interpolating transformation (e.g., polynomial splines, B splines and thin-plate splines [4, 5]) between the source and target images, and (3) aligning the landmarks exactly and mapping other parts of image using the computed transformation.

Accurate interpolating transformation requires an exact match of homologous landmarks. Manual identification of landmark points is time consuming and prone to intra- and interobserver variations. A number of investigators have attempted to automate the landmark definition process by exploiting the geometry of anatomical or biological structures. Typical geometrical features include line intersections [6, 7], local curvature maxima [8–10], and centroid of closed

TABLE 1: Pseudocode of the landmark generation and optimization technique.

(1) Identify homologous contours on the source and target images either by manually drawing or using image processing techniques.
(2) Split the homologous contours into C pairs of corresponding source and target curves.
(3) For $c = 1$ to C
For $i = 3$ to n
Add the i th landmark on the middle of the longest segment of the c th source curve and the homologous landmark on the corresponding target segment.
Optimize landmarks by minimizing the cost function in (1).
End
(4) End

boundary region [11]. Employing the anatomical and biological features has greatly simplified the landmark generation. However, some biomedical images do not contain well-distributed features to generate reliable landmarks for accurate registration. More importantly, noise, artifacts, and other factors can cause errors when identifying the features using either automated or manual selection. A few studies have taken landmark location errors into account. Rohr et al. developed a method to relax the exact landmark matching (i.e., allowing the algorithm to relocate the landmarks) using thin-plate splines [12] by minimizing the bending energy functional [13, 14]. This method can cope with isotropic as well as anisotropic landmark errors. Bookstein [15, 16] used a linear regression model and a technique called “curve décolletage” to relax the interpolation condition. Image properties such as edges of objects [13] have been used to relax the exact landmark matching. However, the effect of the landmark relaxation on the registration accuracy was unknown in prior studies.

In the current investigation, we developed a technique to automate and optimize the landmark generation using the local curvature on anatomical contours and validated the technique. The technique presented here is for two dimensional (2-D) brain image nonlinear registration. Although the rodent brain is a three dimensional (3-D) object, some brain imaging studies are essentially carried out on 2-D planes such as in histological microscopy, and in 2-D MRI studies. More importantly, nonlinear distortion mostly occurs on 2-D as well in these imaging studies. For example, tissue tearing, shearing, shrinkage, and enlargement, during sectioning and section handling, and eddy current in 2-D EPI MRI acquisition, mostly cause 2-D in-plane nonlinear distortions. To guarantee the correspondence of the two brain images to be registered, some prior process 3-D registration may be necessary. Detailed description of the process follows.

2. Materials and Methods

2.1. Algorithm Development. The strategy of this method is first to generate contours on corresponding anatomical features on the source and target images and then generate landmarks on homologous anatomical contours. The landmarks are then relaxed from the original locations and allowed to slide along the contours to achieve optimal matching.

The relocation of the landmarks is governed by a cost function constituted by the local curvatures of landmarks and their displacements. The procedure is described using the pseudocode in Table 1.

The homologous contours on which the landmarks are generated can be manually drawn on the images or identified as the borders of objects using border detection methods, or first identified automatically using border detection methods and then manually modified to correct errors resulting from noise and artifacts. Depending on the image properties, appropriate border detection methods such as dynamic programming or active contour models [17] can be used. The homologous contours are manually split into several corresponding source and target curves that are relatively regular in shape (c.f. Figures 2(a), 2(b), 3(a), and 3(b)). The splitting usually improves the computational stability and efficiency of the nonlinear transformation in our experience (we used the thin-plate splines for the transformation calculations).

The next step is landmark generation and optimization on each curve. An example of this step is illustrated in Figure 1. The first two landmarks and their homologues are fixed at the two ends of the corresponding source and target curves (Figure 1(a)). The remaining landmarks (from the 3rd to the n th; n is the user-preset number of landmarks on the curve) are generated and optimized in an iterative fashion as shown in Table 1. The landmarks split the source and target curves into homologous curve segments (e.g., $i - 1$ landmarks split the curve into $i - 2$ segments). The i th landmark is placed in the middle of the longest segment of the source curve (can be on the target curve depending on the user; the source curve was used in this study), and its homologue is placed on the corresponding target curve segment (Figure 1(b)). Before the $(i + 1)$ th landmark is added, the landmarks are relaxed from their initial locations and slid on the curve segments to match each other (this procedure is called landmark optimization, Figures 1(b) and 1(c)) by minimizing the cost function

$$M = \left| \frac{\kappa_i^S}{\max(\kappa^S)} - \frac{\kappa_i^T}{\max(\kappa^T)} \right| + \lambda \cdot \left(\frac{\Delta_{i-1}^S}{l^S} + \frac{\Delta_{i-1}^T}{l^T} \right), \quad (1)$$

where S and T indicate the source and target, respectively, Δ_{i-1} is the displacement of landmarks on the $(i - 1)$ th curve

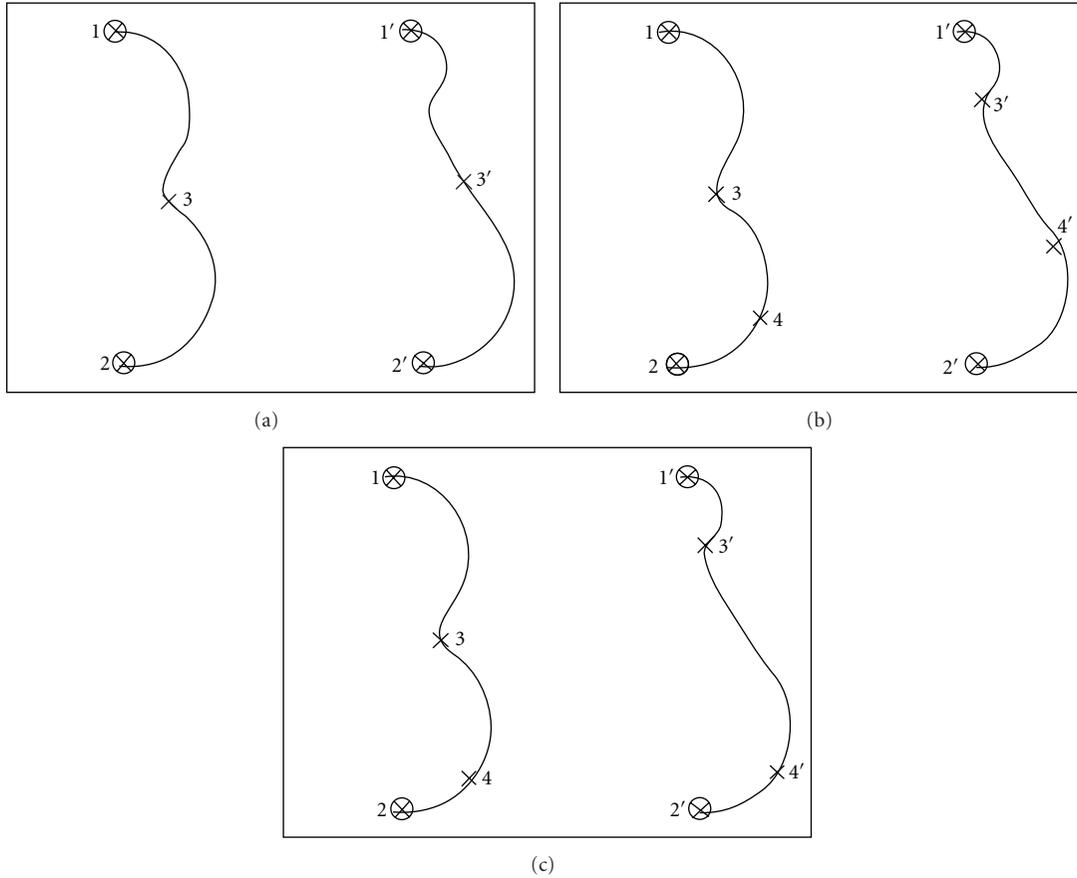


FIGURE 1: An example of the landmark generation and optimization. The left and right columns show the source and target curves, respectively. (a) The first two landmarks (1 and 2, circled X) and their homologues (1' and 2', circled X) are fixed at the two ends of the corresponding source and target curves. The 3th landmark (3) is placed in the middle of the longest segment of the source curve, and its homologue (3') is placed on the target curve. (b) The landmarks (3 and 3') are then relaxed from their initial locations and slid to match each other; and the 4th landmark (4) is placed in the middle of the longest segment of the source curve, and its homologue (4') is placed on the target curve. (c) The 4th landmarks (4 and 4') are then relaxed from their initial locations to match each other.

segment, l is the length of the curve segment, and κ is the local curvature defined by

$$\kappa = \frac{\|\dot{\gamma} \times \ddot{\gamma}\|}{\|\dot{\gamma}\|^3}, \quad (2)$$

where γ is a curve, “ $\|\cdot\|$ ” represents the Euclidean distance, and “ $\dot{\cdot}$ ” and “ $\ddot{\cdot}$ ” are the first and second derivatives, respectively. The first term of the cost function is the difference between the local curvatures (normalized by the maximum curvatures of the curve segments) of a source landmark and the homologous target landmark. The second term is the displacements of the landmarks from their initial locations normalized by segment length. The cost function is a combination of the curvature similarity of the source and target landmarks and the displacements of the landmarks. The landmarks are relaxed and moved to match their curvatures (the first term in (1)) to minimize the cost function, and their displacements are weighed by the second term. The weighting parameter $\lambda > 0$ regulates the sliding distances of the landmarks along the contours. By increasing λ , landmark displacement is restricted. Equation (1) is minimized using

the Nelder-Mead algorithm [18]. When the landmarks have been optimized, any point-based registration method can be used to warp the source image and register the source and target images. We utilized the thin-plate splines [12] to generate the warping field for registering the target images to the source in this work.

2.2. Algorithm Evaluation. The landmark generation and optimization technique was evaluated using two types of mouse brain imaging studies. In the first study, mouse brain MRI was nonlinearly registered with histological sections. The second study examined the ability of the algorithm to correct geometric distortion of EPI—a fast MRI acquisition technique. The convergence criterion of the cost function minimization was set as either landmark displacement of less than 10^{-4} pixels in subsequent iterations or a maximum of 500 iterations, and the weighting factor λ was set to between 0.2 and 0.3 in both of these studies. The thin-plate splines, which have been extensively used in medical image registration [7, 19–30], were used for the nonlinear transformations. The accuracy of registration between the transformed source image (MRI in the first study, and EPI

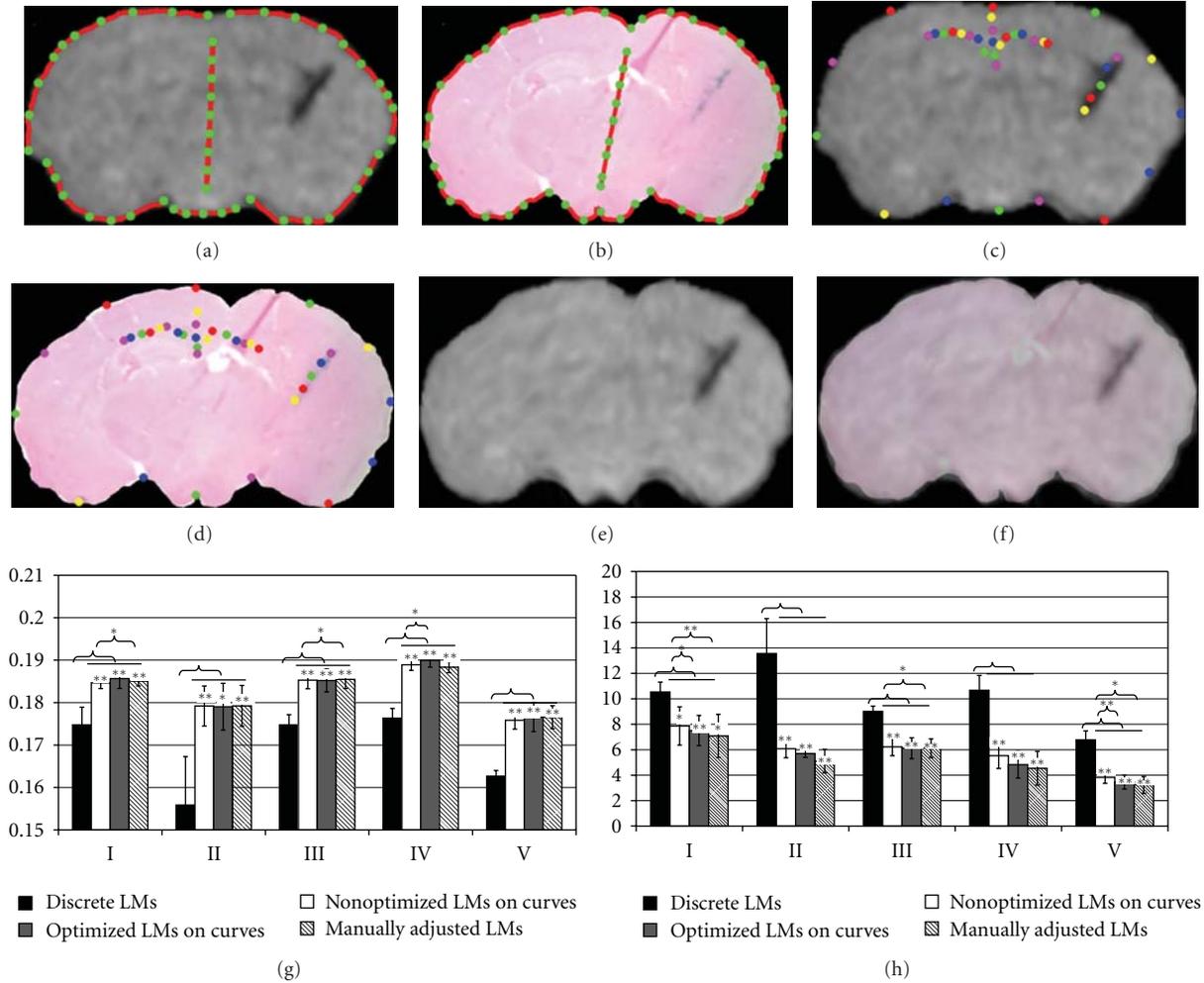


FIGURE 2: The registration of MRI and histological slices using optimized landmarks on curves (a) and (b), manually selected discrete landmarks (c) and (d). The colors of homologous landmarks are matched in (c) and (d). The MRI slice transformed using the thin-plate splines with optimized landmarks in (a) and (b) is shown in (e), and overlaid on the histological section (f). The average NMI and TRE of transformed MRI and histological slices in five image pairs (I–V) were shown in (g) and (h), respectively. The NMI and TRE were calculated on the MRI slices transformed using manually selected “discrete landmarks (LMs)” (black columns), landmarks generated on curves but not optimized (“Nonoptimized LMs on curves”) (white columns), curve landmarks optimized using (1) (“Optimized LMs on curves”) (gray columns), and curve landmarks manually refined (“Manually adjusted LMs”) (stripped columns) *: $P < 0.1$; **: $P < 0.05$.

in the second study) and the reference image was measured by two methods using the normalized mutual information (NMI) and target registration error (TRE), respectively. NMI is defined as

$$\text{NMI} = \frac{H(S_t) + H(T)}{H(S_t, T)}, \quad (3)$$

where $H(S_t)$ and $H(T)$ are the marginal entropies, respectively, of the transformed source image S_t and T , and $H(S_t, T)$ denotes their joint entropy. TRE [3] is the distance between a set of target points (\mathbf{P}_T) and warped corresponding source points (\mathbf{P}_S). In general,

$$\text{TRE} = T(\mathbf{P}_S) - \mathbf{P}_T, \quad (4)$$

where T is the transformation, which is the thin-plate splines in this study. Therefore TRE is an array of vectors. In this

study, only the average magnitude of TRE is of interest and thus reported. The size of the point set was dependent on the nature of the images, and at least four points were identified on each pair of images.

2.2.1. Registration of MRI to Histological Sections. Mouse brains were harvested after 3D T2*-weighted MRI, fixed and embedded in paraffin for histological sectioning. The blockface of the embedded brain was photographed during sectioning (blockface imaging). Individual blockface images were stacked to reconstruct the 3D brain volume. Brain slices were stained with Prussian blue and hematoxylin. The MRI volume was linearly registered to the blockface volume, and then computationally resliced in the coronal plane to match the corresponding histological sections [31, 32]. The MRI slices and histological sections were then nonlinearly registered by three experienced technicians using the thin-plate

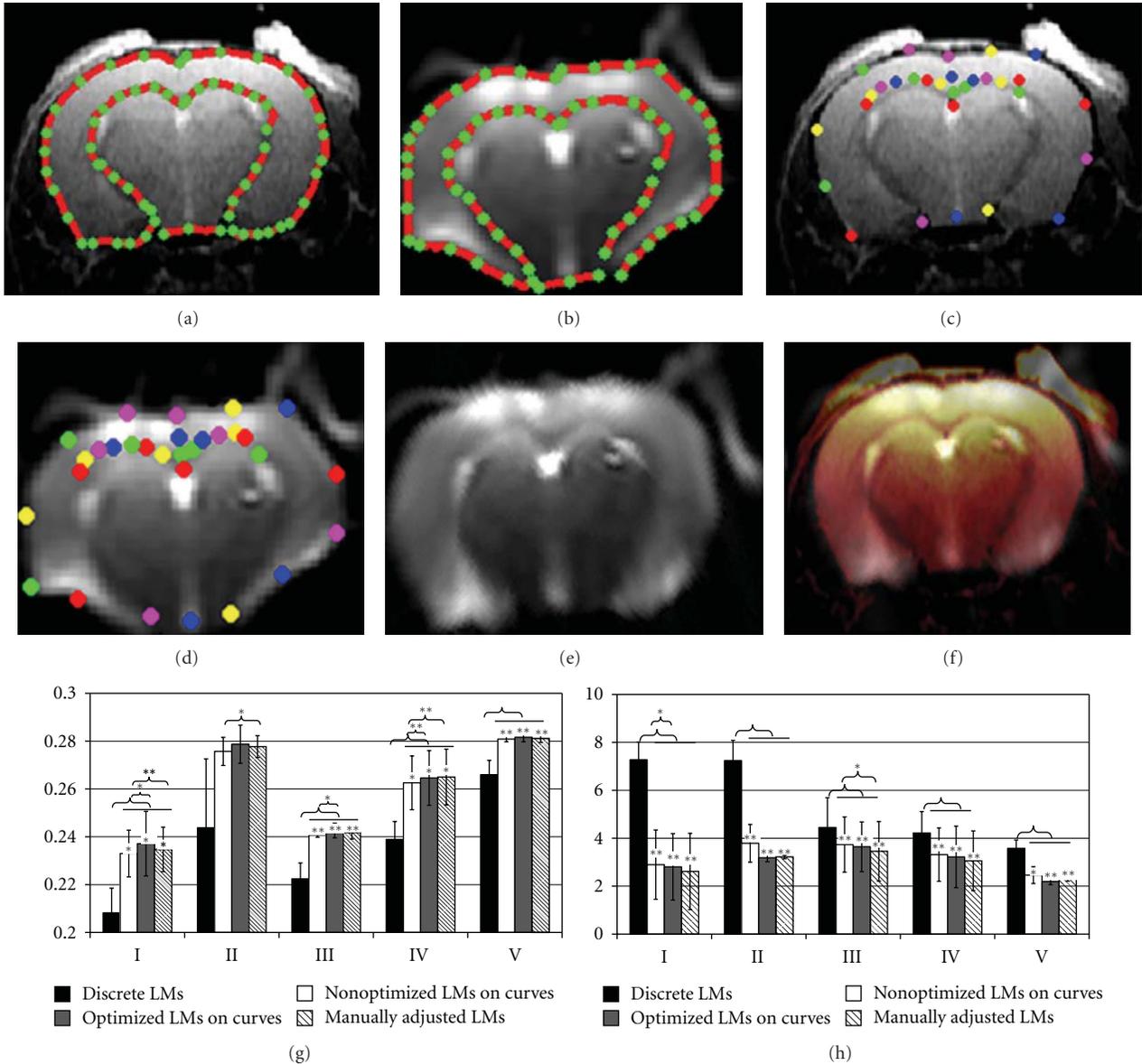


FIGURE 3: The correction of the geometrical distortion on a B0 image (a) in EPI DTI employing the T2-wt image (b) as the reference by optimized landmarks on curves (a) and (b), and manually selected discrete landmarks (c) and (d). The colors of homologous landmarks are matched in (c) and (d). The B0 image transformed using the thin-plate splines with optimized landmarks in (a) and (b) is shown in (e), and overlaid on the T2-wt image plotted in pseudocolor (f). The average NMI and TRE of corrected B0 and T2-wt images in five image pairs (I–V) were shown in (g) and (h), respectively. NMI and TRE were calculated on the B0 images transformed using manually selected “discrete landmarks (LMs)” (black columns), landmarks generated on curves but not optimized (“Nonoptimized LMs on curves”) (white columns), curve landmarks optimized using (1) (“Optimized LMs on curves”) (gray columns), and curve landmarks manually refined (“Manually adjusted LMs”) (stripped columns) *: $P < 0.1$; **: $P < 0.05$.

splines with the landmarks generated and optimized by the presented technique. The technicians were instructed to draw the source and target contours and generate curves according to their knowledge of anatomy. The automatic landmark optimization technique was evaluated by comparing two registration results: the first was from landmarks generated on the user defined contours but adjusted manually by the technicians, and the other registration was conducted using discrete landmarks manually selected by the technicians.

2.2.2. Correction of EPI Distortion. Five mice were scanned with T₂-weighted (T₂-wt) spin-echo imaging and diffusion tensor imaging (DTI) with EPI acquisition. The MRI slices were prescribed at the same anatomical locations in these scans. In the DTI EPI scan, a baseline image without diffusion weighting (B0) was also acquired. The B0 images were used as the source images, and the T2-wt images were used as the target images for registration. Since the B0 image and all diffusion-weighted images undergo same geometric distortion, the transformation procured from the T2-wt/B0

registration is applied to correct the diffusion-weighted images. The B0 images were registered to the T2-wt images by the same three technicians in the MRI/histology registration using manually selected landmarks, landmarks manually adjusted on anatomical contours, and automatically generated and optimized on contours, respectively.

3. Results

3.1. Registration of MRI to Histological Sections. Figures 2(a) and 2(b) demonstrate a pair of MRI and histological slices before coregistration. The curves drawn by a technician and landmarks generated and optimized on the curves are also shown on these figures. The transformed MRI and its overlay on the histological section are shown on Figures 2(e) and 2(f). Figures 2(c) and 2(d) show the discrete manually selected landmarks by a technician. The landmarks that were manually adjusted on curves are not shown.

Registration results with landmarks generated using different methods in five pairs of MRI and histological slices (from Pair I to V) were compared. The mean and standard error of the NMI and TRE of each pair of images by the three technicians is presented in Figures 2(g) and 2(h), respectively. NMI and TRE are shown from the manually selected discrete landmarks (“discrete landmarks (LMs)” in Figures 2(g) and 2(h)), landmarks generated on anatomical curves and optimized using (1) (“Optimized LMs on curves” in Figures 2(g) and 2(h)). The NMI and TRE of landmarks on curves but not optimized (“Nonoptimized LMs on curves”) and of landmarks manually adjusted on curves (“Manually adjusted LMs”) are also presented to investigate the improvement in registration accuracy resulting from landmark optimization. The number of optimized, manually adjusted and nonoptimized landmarks, was the same on each curve. The nonoptimized landmarks were obtained by simply generating the landmarks on the curves without performing optimization using (1), keeping each landmark at its initial position.

As demonstrated in Figure 2(g) (NMI results) and Figure 2(h) (TRE results), generating landmarks on anatomical curves (no matter the landmarks were optimized or not) either significantly improved registration accuracy ($P < 0.05$, paired t -test) or showed a trend towards improvement ($0.05 \leq P < 0.10$) in all image pairs compared to manually selecting discrete landmarks. The automatic landmark optimization using (1) resulted in a trend towards registration accuracy improvement ($0.05 \leq P < 0.10$) in one image pair by NMI calculations (Figure 2(g)) compared to nonoptimized landmarks, and a trend towards improvement ($0.05 \leq P < 0.10$) in two image pairs by TRE calculations (Figure 2(h)). Compared to nonoptimized landmarks, manually adjusted landmarks showed a trend toward improvement in registration accuracy ($0.05 \leq P < 0.10$) on two image pairs by NMI (Figure 2(g)). The TRE results showed a significant accuracy improvement ($P < 0.05$) or a trend towards improvement ($0.05 \leq P < 0.10$) on three image pairs by manually adjusting the landmarks (Figure 2(h)). No difference was found between automatically optimized and manually adjusted landmarks.

3.2. Correction of EPI Distortion. A typical pair of B0 and T2-wt images is shown in Figures 3(a) and 3(b). The geometrical distortion is evident on the B0 by visual comparison to the T2-wt images. Figures 3(a) and 3(b) show the anatomical curves and optimized landmarks on a pair of T2-wt/B0 images, and the discrete landmarks manually selected by the same technician are shown in Figures 3(c) and 3(d). The result of distortion correction using the optimized landmarks is shown in Figure 3(e). For an improved visualization, the corrected B0 (in grayscale) is overlaid on the T2-wt image plotted in pseudocolor in Figure 3(f).

The same statistical analysis was performed in the EPI distortion correction as for the MRI/histology registration, allowing a direct comparison of the results (Figures 3(g) and 3(h)). Generating landmarks on anatomical curves (irrespective of whether or not the landmarks were optimized) either significantly improved registration accuracy ($P < 0.05$) or showed a trend towards improvement ($0.05 \leq P < 0.10$) in all image pairs compared to manually selecting discrete landmarks except in one image pair (in Figure 3(g), by NMI calculations). The automatic landmark optimization using (1) resulted in significant accuracy improvement ($P < 0.05$) in one image pair and a trend towards registration accuracy improvement ($0.05 \leq P < 0.10$) in two image pairs by NMI calculations (Figure 3(g)) compared to nonoptimized landmarks, and a trend towards improvement ($0.05 \leq P < 0.10$) in one image pair by TRE calculations (Figure 3(h)). Compared to nonoptimized landmarks, manually adjusted landmarks showed a significant improvement ($P < 0.05$) in registration accuracy on two image pairs by NMI, and a trend toward improvement ($0.05 \leq P < 0.10$) on one image pair (Figure 3(g)). The TRE results showed a trend towards improvement ($0.05 \leq P < 0.10$) on one image pair by manually adjusting the landmarks (Figure 3(h)). No difference was found between automatically optimized and manually adjusted landmarks.

4. Discussion

In this study, we developed a landmark generation and optimization technique for point-based nonlinear image registration methods. This technique extracts landmarks from the anatomical contours and optimizes the landmark positions by minimizing the cost function constituted by the displacements and the local curvatures of the landmarks. The technique was evaluated in two distinct applications: the registration of MRI and histological slices and distortion correction of EPI MRI images. Statistical analyses have shown that the automation of landmark selection resulted in significant accuracy improvement in image registration compared to manually selected landmarks. Although in most experiments the improvement in NMI and TRE resulting from the landmark optimization was not statistically significant compared to the results using nonoptimized landmarks, the trends towards improvement in registration accuracy was demonstrated in several experiments. Manually adjusting the landmarks on curves could improve registration accuracy on several experiments and show a trend towards improvement in some other experiments. We found no difference in regis-

tration accuracy between landmark automatic optimization and manual adjustment. However, automated landmark selection provides increased efficiency by minimizing the required user intervention.

We used two methods including NMI to validate our technique. NMI has been intensively used as an image similarity measure. It is not a monotonic function of the image similarity and, thus, may be trapped at local minima when used as a driving force for image registration. In this study, the image pairs to compare have all been already registered; thus, NMI was only calculated on a small interval on which it is reasonable to think that NMI is monotonic. TRE was also calculated for the registration evaluation in this study. TRE is a measure of the registration accuracy of a set of points on the images. The points for TRE calculation are usually identified on anatomical features; thereby, the evaluation may be more meaningful than NMI with regard to the anatomical accuracy of registration.

The anatomical contours were manually drawn in this study. Border detection methods can be used to automatically identify the contours. It is reasonable to think that automatically generating contours can minimize inter- and intra-user variance. But on the other hand, some automatic border detection techniques are more susceptible to noise and artifacts compared to manual delineation. We are currently investigating the registration accuracy using different border detection methods.

Not only can this method be used in imaging studies similar to those presented here, but also can be used for multiple brain registration, or similarly, for registration to an atlas with minimum modification. Before using this method, a 3-D affine transformation is likely necessary to first align the brain volumes together, or to the atlas, and then each brain volume needs to be resliced to match individual brain slices, or to the atlas slices.

Overall, this method results in improved registration accuracy and efficiency. However, this technique still requires user intervention and thus suffers inter- and intra-investigator inconsistencies. We are currently improving this technique by including the image intensity and more geometrical information in addition to displacement and curvature to fully automate the landmark generation process. In this study, the number of landmarks either manually selected or automatically generated on curves was determined by the technicians according to their experience. It is desirable to precalculate the necessary landmark number for different landmark generation methods to improve registration accuracy. Investigations using previously published methods to accomplish this automation are underway [33].

Acknowledgments

This work has been supported by NIH Grants K25MH-089851, P20 RR021937, P01 NS043985, P30 MH062261, and P01 DA028555. The authors acknowledge Erin McIntyre, Melissa Mellon, and Lindsay Rice for technical assistance.

References

- [1] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, vol. 46, no. 3, pp. R1–R45, 2001.
- [2] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [3] J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer Jr., "Image registration," in *Handbook of Medical Imaging*, M. Sonka and J. M. Fitzpatrick, Eds., vol. 2, pp. 447–513, SPIE Press, Bellingham, Wash, USA, 2004.
- [4] W. R. Crum, L. D. Griffin, D. L. G. Hill, and D. J. Hawkes, "Zen and the art of medical image registration: correspondence, homology, and quality," *NeuroImage*, vol. 20, no. 3, pp. 1425–1437, 2003.
- [5] Y. Liu, H. D'Arceuil, J. He et al., "A nonlinear mesh-warping technique for correcting brain deformation after stroke," *Magnetic Resonance Imaging*, vol. 24, no. 8, pp. 1069–1075, 2006.
- [6] G. C. Stockman, S. Kopstein, and S. Benett, "Matching images to models for registration and object detection via clustering (RST)," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 4, no. 3, pp. 229–241, 1982.
- [7] M. A. Jacobs, J. P. Windham, H. Soltanian-Zadeh, D. J. Peck, and R. A. Knight, "Registration and warping of magnetic resonance images to histological sections," *Medical Physics*, vol. 26, no. 8, pp. 1568–1578, 1999.
- [8] L. N. Kanal, B. A. Lambird, D. Lavine, and G. C. Stockman, "Digital registration of images from similar and dissimilar sensors," in *Proceedings of the International Conference on Cybernetics and Society*, vol. 4, pp. 347–351, 1981.
- [9] P. Putjarupong, C. Pintavirooj, W. Withayachumnankul, and M. Sangworasil, "Image registration exploiting five-point coplanar perspective invariant and maximum-curvature point," in *Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG '04)*, vol. 12, pp. 341–348, Plzen, Czech Republic, 2004.
- [10] D. C. Van Essen, H. A. Drury, S. Joshi, and M. I. Miller, "Functional and structural mapping of human cerebral cortex: solutions are in the surfaces," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 3, pp. 788–795, 1998.
- [11] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognition*, vol. 19, no. 6, pp. 459–466, 1986.
- [12] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
- [13] K. Rohr, H. S. Stiehl, R. Sprengel et al., "Nonrigid registration of medical images based on anatomical point landmarks and approximating thin-plate splines," in *Proceedings of the Aachenener Workshop*, pp. 41–46, Aachen, Germany, 1996.
- [14] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Transactions on Medical Imaging*, vol. 20, no. 6, pp. 526–534, 2001.
- [15] F. L. Bookstein, "Four metrics for image variation," *Progress in Clinical and Biological Research*, vol. 363, pp. 227–240, 1991.
- [16] F. L. Bookstein, "Landmark methods for forms without landmarks: morphometrics of group differences in outline shape," *Medical Image Analysis*, vol. 1, no. 3, pp. 225–243, 1997.

- [17] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, PWS Publishing Company, Boston, Mass, USA, 1998.
- [18] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [19] M. H. Davis, A. Khotanzad, D. P. Flamig, and S. E. Harms, "A physics-based coordinate transformation for 3-D image matching," *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, pp. 317–328, 1997.
- [20] B. Kim, J. L. Boes, K. A. Frey, and C. R. Meyer, "Mutual information for automated unwarping of rat brain autoradiographs," *NeuroImage*, vol. 5, no. 1, pp. 31–40, 1997.
- [21] C. R. Meyer, J. L. Boes, B. Kim et al., "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations," *Medical Image Analysis*, vol. 1, no. 3, pp. 195–206, 1997.
- [22] M. S. Buchsbaum, J. H. Fallon, T. C. Wei et al., "A method of basal forebrain anatomical standardization for functional image analysis," *Psychiatry Research*, vol. 84, no. 2-3, pp. 113–125, 1998.
- [23] J. R. DeQuardo, M. S. Keshavan, F. L. Bookstein et al., "Landmark-based morphometric analysis of first-episode schizophrenia," *Biological Psychiatry*, vol. 45, no. 10, pp. 1321–1328, 1999.
- [24] J. E. Downhill Jr., M. S. Buchsbaum, T. Wei et al., "Shape and size of the corpus callosum in schizophrenia and schizotypal personality disorder," *Schizophrenia Research*, vol. 42, no. 3, pp. 193–208, 2000.
- [25] K. Rohr, *Landmark-Based Image Analysis: Using Geometric and Intensity Models*, vol. 21 of *Computational Imaging and Vision Series*, Kluwer Academic Publishers, Dodrecht, The Netherlands, 2001.
- [26] M. I. Miller, A. Trouvé, and L. Younes, "On the metrics and Euler-Lagrange equations of computational anatomy," *Annual Review of Biomedical Engineering*, vol. 4, pp. 375–405, 2002.
- [27] I. Corouge, P. Hellier, B. Gibaud, and C. Barillot, "Interindividual functional mapping: a nonlinear local approach," *NeuroImage*, vol. 19, no. 4, pp. 1337–1348, 2003.
- [28] F. Darvas, J. J. Ermer, J. C. Mosher, and R. M. Leahy, "Generic head models for atlas-based EEG source analysis," *Human Brain Mapping*, vol. 27, no. 2, pp. 129–143, 2006.
- [29] G. Li, S. Nikolova, and R. Bartha, "Registration of in vivo magnetic resonance T1-weighted brain images to triphenyltetrazolium chloride stained sections in small animals," *Journal of Neuroscience Methods*, vol. 156, no. 1-2, pp. 368–375, 2006.
- [30] C. R. Meyer, B. A. Moffat, K. K. Kuszpit et al., "A methodology for registration of a histological slide and in vivo MRI volume based on optimizing mutual information," *Molecular Imaging*, vol. 5, no. 1, pp. 16–23, 2006.
- [31] Y. Liu, M. G. Uberti, H. Dou et al., "Ingress of blood-borne macrophages across the blood-brain barrier in murine HIV-1 encephalitis," *Journal of Neuroimmunology*, vol. 200, no. 1-2, pp. 41–52, 2008.
- [32] M. Uberti, Y. Liu, H. Dou, R. L. Mosley, H. E. Gendelman, and M. Boska, "Registration of in vivo MR to histology of rodent brains using blockface imaging," in *Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 7262 of *Proceedings of the SPIE*, Lake Buena Vista, Fla, USA, February 2009.
- [33] A. Joshi, D. Pantazis, H. Damasio, D. Shattuck, Q. Li, and R. Leahy, "Optimization of landmark selection for cortical surface registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 699–706, Miami, Fla, USA, June 2009.