Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

Evaluating Spatial Normalization Methods for the Human Brain

Veronica S. Smith, *Member, IEEE*, Linda G. Shapiro, *Fellow, IEEE*, Donna Hanlon, Richard F. Martin, James F. Brinkley, Andrew V. Poliakov, George A. Ojemann and David P. Corina

***Abstract*—Cortical stimulation mapping (CSM) studies have shown cortical locations for language function are highly variable from one subject to the next. If individual variation can be normalized, patterns of language organization may emerge that were heretofore hidden. In order to uncover these patterns, computer-aided spatial normalization to a common atlas is required. Our goal was to determine a methodology by which spatial normalization methods could be evaluated and compared. We developed key metrics to measure accuracy of a surface-based (Caret) and volume-based (SPM2) method. We specified that the optimal method would i) minimize variation as measured by spread reduction between CSM language sites across subjects while also ii) preserving anatomical localization of all CSM sites. Eleven subject's structural MR image sets and corresponding CSM site coordinates were registered to the colin27 human brain atlas using each method. Local analysis showed that mapping error rates were highest in morphological regions with the greatest difference between source and target. Also, SPM2 mapped significantly less type 2 errors. Although our experiment did not show statistically significant global differences between the methods, our methodology provided valuable insights into the pros and cons of each method.**

I. Introduction

CORTICAL stimulation mapping (CSM) is a common technique used when conducting temporal lobe epileptic tumor resection in order to avoid areas on the cortical surface that are essential for language when resecting an epileptic focus. The technique involves bringing the patient to an awakened state after the craniotomy is completed and showing slides of line drawings of familiar objects like planes, boats, etc. The slides are projected at 4-second intervals, with the patient trained to name each one as it appears. A 1.5-10 mA current at 60 Hz is applied across 1-mm bipolar electrodes separated by 5 mm by a constant-current stimulator to selected cortical surface locations as the slide appears and continues until the appearance of the next slide. This electrical stimulation results in short-term reversible localized disruption of neural function. Three samples of stimulation effect are usually obtained for a given site. If stimulation results in a naming error two out of three times, then immediate feedback is provided identifying that site as essential for language function. These sites are avoided during resection [1].

CSM studies have shown that cortical locations for language function are highly variable from one subject to the next. If individual variation can be normalized, patterns of language organization may emerge that were heretofore hidden. In order to uncover such patterns, computer-aided spatial normalization to a common atlas is required.

Computer-aided spatial registration is a widely used solution for relating anatomy and functionality in neuroscience. We use the term 'registration' to mean determining the spatial alignment between images of the same or different subjects, acquired with the same or different modalities, and also the registration of images with a given coordinate system. The term 'normalization' is usually restricted to the intersubject situation and is the term we will use in this paper. Spatial normalization accuracy is critical to the success of quantitative analysis of the human cortex. Therefore, it is critical to select the optimal normalization method for any given research application. The problem is that there are many non-linear spatial normalization methods to choose from. Typically these methods fall into one of two categories.

Surface-based methods employ algorithms that make use of geometrical features in images such as points, lines and/or surfaces to determine the mapping of the positional normalization transformation. They identify feature correspondences between a pair of images that allow for the iterative calculation of the spatial transformation from one image to the other. Volume-based methods employ volumetric transformations involving intensity values in which the algorithm iteratively determines both the positional and intensity normalization transformations that optimize a voxel similarity measure [2].

The goal of our work is to determine a methodology by which normalization methods used for CSM studies could be evaluated and compared. In this paper, we describe the experimental procedures used during and after surgery, define our new spread-reduction and anatomical localization metrics, and test those metrics by comparing two normalization methods, the surface-based Caret [3] and the volume-based SPM2 [4], for mapping 2D and 3D patient brain data (source) to the colin27 human brain atlas (target).

II. Methodology

Typically there are 10-20 CSM sites per subject that are

Manuscript received May 2, 2005. This work was supported in part by the NIH Human Brain Project grant DC02310 and the NIMH/NSF/NCI/NLM/NASA Human Brain Project/Neuroinformatics grant R01 MH60974. V.S. Smith is with the University of Washington Structural Informatics Group (SIG), (e-mail: vss@u.washington.edu). L.G. Shapiro is with the University of Washington Computer Science and Engineering Department (email: shapiro@cs.washington.edu). D. Hanlon is with the Washington University Van Essen Lab in the Department of Anatomy and Neurobiology (e-mail: donna@brainvis.wustl.edu). R.F. Martin, J.F. Brinkley and A.V. Poliakov are with the University of Washington SIG in the Department of Biological Structure. Dr. G.A. Ojemann is a University of Washington attending neurosurgeon. D.P. Corina is with the University of Washington Department of Psychology.

identified with sterile numbered tags on the cortex and the locations recorded with a digital intraoperative photo. Following surgery, a visual comparison approach is used to transpose the location of each cortical site to a 3D volume reconstruction of the cortical surface. Using the blood vessels and anatomical structure in the rendering as landmarks, the neuroanatomist expert drags and drops the number that corresponds to the CSM site in the intraoperative photo onto the 3D rendered image. Once the site number has been dropped, a 'pick' operation is performed in order to determine the closest surface facet to the site. The site is assigned a 3D coordinate in MR 'magnet space' in which the center of the MR magnet is the origin. This data is stored in the CSM database. Based on repeatability studies, any given mapping will typically fall within a distance of 5.1 mm of the true site location. Since the CSM site locations mapped during surgery are accurate to 1 cm, the accuracy achieved using the visual comparison approach was deemed satisfactory [5].

CSM site magnet space coordinates could not be used to compare site locations across subjects prior to normalization because each subject's volume had its own magnet center, and in some cases the chin was rotated up or down or slightly to the side. Thus, we needed to shift the individual volume images into a common grid (i.e. standard voxel size, origin and orientation) to create pre-normalized site coordinates that could be used to measure distances between CSM sites across subjects. To achieve this, we aligned the anterior commissure (AC) and posterior commissure (PC) using the AFNI software package. This process resampled each individual volume to cubic 1 mm voxels and applied a rigid registration to align the volumes to a common origin: the intersection of the superior edge and posterior margin of the AC. AFNI also rotated the volume so that its Y axis ran from the inferior edge of the PC to the AC origin. Then, the AFNI command line utility, Vecwarp, was used to apply the transform to the individual coordinates, resulting in pre-normalized AC-PC aligned coordinate files.

Additionally, the expert assigned an anatomical location to each CSM site based on a cortical parcellation system (CPS), designed as a scheme for examining the neural substrate through intelligent computer querying of the CSM database. This system divides the lateral surface of the cortex into 37 subdivisions, labeled using the Foundational Model of Anatomy (FMA) expansion of NeuroNames terminology and is shown in figure 1.

In this study we included 11 subjects who had undergone CSM. There were a total of 198 CSM site locations recorded for these subjects. Of these sites, we were especially interested in the 21 language sites (2 sites/subject on average) identified as a source of naming errors when electrically stimulated.

Previous studies [6]-[10] used a variety of metrics including the dispersion metric of selected landmarks, differential characteristics, tissue classification, spatial homogeneity of selected anatomical features, overlap percentage of restricted volume of interest, cross-comparison of 3D probability maps to evaluate and compare methods. However, these metrics were not applicable to the CSM data, which are not commonly collected. Therefore, we needed to design new metrics.

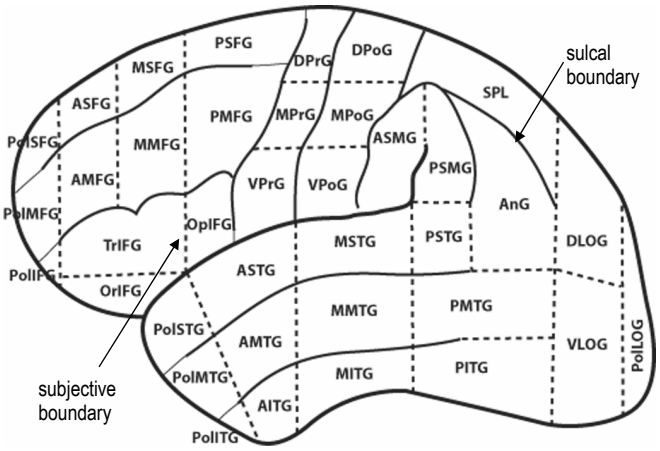

Fig. 1. Cortical Parcellation System for left lateral cortical surface

III. Language Site Spread Reduction Metric

We believe that a combination of anatomical and functional variation increases the distance between CSM language sites across patients. It follows, then, if variation is reduced, the distance between language sites across subjects will be reduced. Therefore, we expect this distance, which we will refer to as 'spread,' will be reduced after spatial normalization. The optimal method will maximize spread reduction.

In order to measure the spread prior to and after normalization, we first measured distances between language sites across subjects using descriptive statistics. Equations (1) and (2) show how we measured average distance across two subjects prior to and after normalization:

$$AvgPreD_{pp'} = \frac{1}{n*m}\sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{D}(x_{pi}, x_{p'j}) \quad (1)$$

$$AvgPostD_{pp'} = \frac{1}{n*m}\sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{D}(y_{pi}, y_{p'j}) \quad (2)$$

where n = number of language sites identified for subject p; m = number of language sites identified for subject p′; D = Euclidian distance between points; x = 2D or 3D site coordinates prior to normalization; y = 2D or 3D site coordinates after normalization. We expanded these calculations for each subject pair and summed the distances resulting in an overall average pre-normalized (APrD) and post-normalized distance (APoD).

With these distance measures, we could determine the spread reduction if the volume and surface areas of the source and target were the same. However, the mean surface and volume areas of the 11 subjects were less than the target's surface and volume areas. To accommodate for this difference we calculated an expected post-normalized distance (EPoD) in 2D and 3D space as follows:

$$EPoD_{2D} = \sqrt{CSA/ASA} * APrD_{2D} \quad (3)$$

$$EPoD_{3D} = \sqrt[3]{CVA/AVA} * APrD_{3D} \quad (4)$$

where APrD = average pre-normalized distance; CSA = colin27 surface area; CVA = colin27 volume area; ASA = subjects' average surface area and AVA = subjects' average volume area. We then calculated spread reduction (SR) as follows:

$$SR_{M,dim} = EPoD_{dim} - APoD_{M,dim} \quad (5)$$

where *M* = a given spatial normalization method and *dim* = a given spatial dimension (2D or 3D).

IV. Anatomical Localization Metric

Not only does an accurate spatial normalization method need to reduce spread, but it also must preserve anatomical relationships. To measure anatomical localization, we compared pre- and post-normalized site locations for all 198 CSM sites.

Following normalization, the expert viewed cortical flat maps including sulcal depth patterns and CSM sites via the Caret GUI. The SPM2 mapping was viewed on the left side of the screen and the Caret mapping on the right. The expert identified the location of each site as a CPS parcel. The post-normalized parcellation was recorded and compared to the pre-normalized parcellation. A correct mapping received a score of 1. Error types and scores were assigned as follows:

Type 1 Error: Site is located in an incorrect parcel across a subjective boundary and receives a score of -0.25.

Type 2 Error: Site is located in the sulcus adjacent to the correct parcel and receives a score of -0.5.

Type 3 Error: Site is located in the incorrect parcel across a sulcal boundary and receives a score of -1.

The possible score range was [-198,198].

With the two metrics, we were prepared to answer our key question: "What is the optimal spatial normalization method for registering two or more human brains such that both spread reduction and anatomical localization preservation are maximized, as measured by CSM?" Because the CSM data were collected on the cortical surface, we expected that surface-based normalization results would be superior to volume-based normalization results.

V. Experiments & Results

To test this hypothesis, we developed a 6-step evaluation protocol to identify the optimal spatial normalization method.

Step 1: We selected 11 whole brain structural MR image sets from a University of Washington Structural Informatics Group database of over 90 patients (CSM database).

Step 2: We selected the SureFit software package to create surface reconstructions of the fourth cortical layer of the left hemisphere of each subject's brain. Prior to launching the automated segmentation process, the MRI volume was resampled to 1 mm cubic voxels.

Step 3: To aid in visual assessment of anatomical localization preservation, we created cortical flat maps using Caret.

Step 4: To normalize the source (individual brain surface) to the target (colin27 atlas) using Caret, we first selected the Core6 landmarks required to constrain the registration. To do this, we followed the protocol as outlined in [14] to delineate the extent of each landmark border. We then ran the automatic normalization algorithm.

To normalize the source to the target using SPM2 we input the subject's MR image in the form of a MINC file with X increasing from patient left to right. No flipping was done during normalization. With the exception of using the template bounding box and cubic 1 mm voxel dimensions, the default spatial normalization settings were used. The selected template image was a T1 MINC average volume of the MNI152 average brain atlas.

Step 5: The deformation file from each method was applied to the individual coordinate file in magnet space coordinates, resulting in a post-normalized coordinate file registered to the MNI152 coordinate space. In Caret, a spherical registration algorithm used landmark borders to create a deformation file. The SPM2 algorithm spatially normalized the individual volume image to the avg152T1 MINC file to create a deformation file.

Step 6: The analysis of spread reduction and anatomical localization preservation revealed that there was not a statistically significant difference between the two methods globally. As outlined in table 1, Caret reduced the spread between language sites by 5.1 mm more than SPM2 in 2D space. In 3D space, Caret reduced the spread by 1.9 mm more than SPM2. A power t test calculation estimate showed that the number of subjects required to achieve statistical significance of $p < 0.05$ and 80% power was 55 for 2D analysis and 120 or more subjects for 3D analysis.

The overall anatomical localization accuracy revealed that Caret mapping accuracy was 1.6% better than SPM2, not significantly different. Qualitative analysis of the error types provides more insight into two spatial normalization problems. Most notably, the CPS parcels with the highest rate of errors were the regions with important morphometric differences between source and target. Also, a paired t test showed a statistically significant difference ($p < 0.01$) in the type 2 errors mapped by both methods.

TABLE 1
SUMMARY OF RESULTS

	2D Spread Reduction	3D Spread Reduction	Localization Accuracy Rate
Caret	9.9 mm	3.2 mm	79.6%
SPM2	4.8 mm	1.3 mm	78.0%

Error Rate Analysis by CPS Parcel

The average error rate in the middle part of the superior temporal gyrus (MSTG), as measured by averaging the sum of the Caret and SPM2 error rates, was 54.2%. Other parcels with 7 or more CSM sites having an error rate of 50% or more were the posterior part of the supramarginal gyrus (PSMG) and the ventral part of the precentral gyrus (VPrG). The common mapping errors support what visual inspection of the structural surfaces of both the source and target hemispheres revealed: locations of structural vagaries in both the colin27 and in the subjects' average surface

reconstruction were where mapping error rates were 50% or greater.

The colin27 atlas structural regions were observed by a neuroanatomist to be atypical in the ventral portion of the precentral gyrus (VPrG), supramarginal gyrus (SMG) and terminal-ascending segment of the lateral fissure. These uncommon localized folding patterns help explain the average error rates of 50% or more in the VPrG and PSMG.

Our analysis comparing a digital atlas of 12 normal subjects in [11] to 10 of the 11 epileptic subjects revealed that epileptic subjects have a broader superior temporal gyrus (STG) than do the normal subjects. Analysis of a sulcal depth difference flat map revealed that the greatest difference between epileptic and normal subjects' left hemispheres is in the middle part of the superior temporal gyrus (MSTG) on the CPS scheme.

We hypothesize that areas of important variability between source and target are a key cause of at least 20% of the total anatomical localization errors. A possible solution to this problem would be to create an atlas that averaged the sulcal shapes of epilepsy subjects, presumably resulting in more accurate spatial normalization.

TABLE 2
SUMMARY OF MAPPING ERRORS

	Type 1 Error	Type 2 Error	Type 3 Error	Total Error
Caret	18	18	27	63
SPM2	24	1	37	62

Analysis of Type 2 Errors

A paired t test of type 2 errors did reveal a statistically significant difference ($p < 0.01$) between the methods. SPM2 mappings resulted in only one type 2 error compared to 18 type 2 errors mapped using Caret (table 2). We believe that this difference is attributable to the underlying differences in normalization approaches used by the different methods. The SPM2 algorithm maximizes a voxel intensity match between source and target. As a result, a volume-based normalization will rarely result in a re-alignment of a gyral location (i.e. CSM sites are always on the gyrus) to a sulcal location, where the voxel intensity is markedly less than that found in a gyrus. Caret, however, maximizes alignment of a set of landmarks based on cortical folding patterns without consideration for voxel intensity. If the selected landmarks vary enough between the source and target, then the normalized sulci and gyri may be deformed in ways that confound mapping of functional data to corresponding regions of the anatomical substrate.

In addition, SPM2's language site localization accuracy was better than Caret's. SPM2 incorrectly mapped 6 of the 21 language sites, while Caret incorrectly mapped 9 language sites. Again, the superior temporal gyrus (STG) was the most problematic region for both methods. With 70% of language sites being located on the STG and the type 2 error mapping problem discussed previously, Caret's lower language site accuracy rate is understandable.

VI. CONCLUSION

We have developed a new methodology for evaluating spatial normalization methods using the spread-reduction and anatomical localization metrics. We tested the methodology by comparing the surface-based Caret method to the volume-based SPM2. Although our experiment did not show statistically significant global differences between the methods, our methodology provided valuable insights into the pros and cons of each.

REFERENCES

[1] Ojemann G, Ojemann J, Lettich E *et al.* Cortical language localization in left, dominant hemisphere. J Neurosurg, 1989; 71:316-326.

[2] Hill DLG, Batchelor PG, Holden M, Hawks DJ. Medical Image Registration. Phys. Med. Biol. 2001; 46:R1-R45.

[3] Van Essen DC. Drury HA, Dickson J, Harwell J, *et al.* An Integrated Software Suite for Surface-based Analysis of Cerebral Cortex. JAMIA. 2001; 8:443-459.

[4] Ashburner J, Friston KJ. Voxel-based morphometry: the methods. Neuroimage. 2000; 11:805-821.

[5] Modayur B, Jakobovits R, Maravilla K, Ojemann G, Brinkley J. Evaluation of a visualization-based approach to functional brain mapping. Proc AMIA Annu Fall Symp. 1997; 429-433.

[6] Hellier P, Barillot C, Corouge I, *et al.* Retrospective evaluation of intersubject brain registration. IEEE Trans Med Imaging. 2003 Sep; 22(9): 1120-30.

[7] Crivello F, Schormann T, Tzourio-Mazoyer N, *et al.* Comparison of spatial normalization procedures and their impact on functional maps. Hum Brain Mapp. 2002 Aug; 16(4):228-50.

[8] Grachev ID, Berdichevsky D, Rauch SL, *et al.,* A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks. NeuroImage. 1999; 9:250-268.

[9] Minoshima S, Koeppe RA, Frey KA, Kuhl DE. Anatomic standardization: linear scaling and nonlinear warping of functional brain images. J Nucl Med. 1994; 35:1528-1537.

[10] Sugiura M, Kawashima R, Sadato N, *et al.* Anatomical validation of spatial normalization methods for PET. J Nucl Med. 1999; 40:317-322.

[11] Van Essen DC. A population-average, landmark- and surface-based atlas of human cerebral cortex. Draft. 2005.