Evaluating the Validity of Volume-Based and Surface-Based Brain Image Registration for Developmental Cognitive Neuroscience Studies in Children 4-to-11 Years of Age

Satrajit S. Ghosh[1,*], **Sita Kakunoori**[3], **Jean Augustinack**[3], **Alfonso Nieto-Castanon**[2], **Ioulia Kovelman**[2,4], **Nadine Gaab**[5], **Joanna A. Christodoulou**[2], **Christina Triantafyllou**[8], **John D. E. Gabrieli**[2,7], and **Bruce Fischl**[3,6,7]

[1] Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA

[2] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA

[3] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA

[4] University of Michigan, Ann Arbor, MI, USA

[5] Childrens Hospital, Boston, MA, USA

[6] Computer Science and AI Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

[7] Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA

[8] Athinoula A. Martinos Brain Imaging Center at the McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

Understanding the neurophysiology of human cognitive development relies on methods that enable accurate comparison of structural and functional neuroimaging data across brains from people of different ages. A fundamental question is whether the substantial brain growth and related changes in brain morphology that occur in early childhood permit valid comparisons of brain structure and function across ages. Here we investigated whether valid comparisons can be made in children from ages 4–11, and whether there are differences in the use of volume-based versus surface-based registration approaches for aligning structural landmarks across these ages. Regions corresponding to the calcarine sulcus, central sulcus, and Sylvian fissure in both the hemispheres were manually labeled on T1-weighted structural magnetic resonance images from 31 children ranging in age from 4.2 to 11.2 years old. Quantitative measures of shape similarity and volumetric-overlap of these manually labeled regions were calculated when brains were aligned using a 12-parameter affine transform, SPM's nonlinear normalization, a diffeomorphic registration (ANTS), and FreeSurfer's surface-based registration. Registration error for normalization into a common reference framework across participants in this age range was lower than commonly used functional imaging resolutions. Surface-based registration provided

---

[*]Corresponding author satra@mit.edu.

significantly better alignment of cortical landmarks than volume-based registration. In addition, registering children's brains to a common space does not result in an age-associated bias between older and younger children, making it feasible to accurately compare structural properties and patterns of brain activation in children from ages 4–11.

Introduction

Advances in structural and functional neuroimaging allow for unprecedented opportunities to discover how the development of the child's brain supports the growth of the child's mind, both in typical development and in developmental disorders. Measurement of structural and functional brain development from childhood to adulthood necessitates quantitative comparisons of structure across ages, either to examine structural changes or to localize functional changes. Such quantitative comparisons are typically achieved by registering or mapping individual brains into a common or normalized stereotactic space (Miller et al., 1993; exceptions are functional or structural regions-of-interest approaches). A potential hazard in developmental studies is that the many changes in brain morphology that occur in development may confound such normalization across ages in both cross-sectional and longitudinal neuroimaging studies. Here, we examined two important issues regarding such normalization in developmental cognitive neuroscience. First, we assessed whether normalization is valid from ages 4 to 11, a period of major brain and mental growth. Normalization appears to be valid in 7-to-8 year olds (Burgund et al., 2002), but there is no evidence in younger ages. With increasingly early identification and treatment of developmental disorders or risks for developmental disorders such as attention-deficit/ hyperactivity disorder (ADHD), dyslexia, and bipolar disorder, it is important to know if brains of children ages 4–6 can be studied developmentally with standard spatial normalization procedures. Second, we examined whether a surface-based approach for normalization offers superior accuracy for cortical regions relative to more commonly used volume-based approaches. Such quantification of the accuracy of normalization approaches, as investigated in this study, is essential for precise characterization of developmental changes in structure and function.

The first decade of life represents a period of extensive anatomical changes and functional maturation of the human brain. Most major fissures or sulci are visible on the surface of the brain at the time of birth (Cowan, 1979). However, the brain continues to expand in volume and morphological changes persist for years after birth (see Toga et al., 2006, for a review). Cellular (e.g., generation and pruning of neurons and synapses, myelination) and macroanatomical (e.g., cortical thinning, white matter expansion, changes in sulci and gyri) changes are closely associated with cognitive development. Magnetic resonance imaging (MRI) provides a window into these developmental changes, but is currently limited to detecting changes mainly at a macroscopic level, such as positions of sulci and gyri, cortical thickness, connectivity, and curvature. Prior studies (Rademacher et al., 1993; Hinds et al., 2008; Fischl et al., 2008) have reported that sulci and gyri correspond well to cytoarchitectonic features in several regions of the brain. Therefore much can be learned even from such coarse information. However, in order to compare these features across children of different ages, brain images from different participants are registered or spatially normalized into a common coordinate system.

Registration or spatial normalization is the process of transforming brains from different participants into a common reference frame. Registering brains helps in comparing: (i) structural and functional properties across the participants within a study; and (ii) results from different brain imaging studies. Currently, in most fMRI group analyses, volume-based registration is used to transform brain-imaging data into canonical spaces (e.g., Talairach

space - Talairach and Tournoux, 1988; MNI space – Evans et al., 1992). Several algorithms exist for performing volume-based registration (reviewed in Ardekani et al., 2005; Gholipour et al., 2007; Klein et al., 2009). These approaches typically employ a transformation that matches the overall extents of the brains to one another or to an average brain template (e.g., MNI152, MNI305 spaces; Evans et al., 1993). In general, they use intensity differences to guide registration. However, such approaches tend to ignore the topological properties and geometric features (e.g., sulci and gyri) of the cortex. As a result, these normalization procedures that are meant to align or register anatomical regions across participants leave a large amount of residual inter-subject anatomical variability (Amunts et al., 1999; Nieto-Castanon et al., 2003; Hinds et al., 2008) and therefore blur individual anatomical distinctions.

Surface-based algorithms (Fischl et al., 1999; Davatzikos et al., 1996; Drury et al., 1996, Thompson and Toga, 1996, Cointepas et al., 2001, Tosun and Prince, 2008) were developed to improve the accuracy of cortical registration and thereby reduce inter-subject variability. These approaches account for morphological and topological properties of the human brain. When performing registrations, surface-based approaches treat the cerebral cortex as a sheet and seek to find an alignment that matches sulcal and gyral patterns typically quantified using some type of curvature of the cortex. Surface-based registration has been shown to map cytoarchitectonic borders more accurately between brains than affine volume-based registration (Fischl et al., 2008). Hinds et al. (2008) reported significant reduction in prediction error of locating V1 using an atlas constructed via surfaced-based registration over a nonlinear volume-based approach (Hömke, 2006; Schormann and Zilles, 1998). Using this surface-based atlas Hinds et al. (2009) demonstrated that predicted- and histologically defined structural boundaries of primary visual cortex align well with functionally defined boundaries.

In this study, we chose four registration algorithms: three volume-based and one surface-based. Two volume-based algorithms were chosen on the basis of being among the most commonly used methods in published literature. These were SPM 5 nonlinear normalization (Ashburner et al., 1999) and a 12-parameter affine transform (e.g., similar to FLIRT - Jenkinson et al., 2002). In addition, we chose ANTS (Avants et al., 2006), a diffeomorphic-registration algorithm, which consistently ranked highest among nonlinear volume-registration algorithms evaluated in Klein et al. (2009) and which showed no significant difference in registration accuracy compared to FreeSurfer in a study comprising labeled data in adults (Klein et al., 2010). For surface-based registration, we used FreeSurfer, a fully automated, freely available morphological analysis software package that does not require manually created landmarks to perform registration. Currently, other surface registration methods require manually assigned landmarks (e.g., Caret - Van Essen et al., 2001; curve-LDDMM - Qui and Miller, 2007) or are unable to apply nonlinear transforms to arbitrary labels (e.g., BrainVisa; Cointepas et al., 2001). They were not included in this study.

Evaluating the accuracy of registration algorithms on brain images typically requires comparison of the automatic registration to an objective criterion based on individual anatomy. Prior studies have used consistent, manual labeling of cortical landmarks or features such as gyri and sulci as such a criterion to compare the accuracy of volume-based (Klein et al., 2009) and surface-based (Fischl et al., 2004; Desikan et al., 2006) registration approaches. However, all the underlying brain images used in these studies were from adult participants. In contrast, the current study aimed to evaluate the accuracy of volume- and surface-based registration in aligning macroanatomically defined brain regions across a set of pediatric brain images of varying ages.

Only one previous study (Burgund et al., 2002) examined the accuracy of registering anatomical landmarks from a pediatric population. The study investigated the feasibility of using a common volume-based stereotactic coordinate system for comparing functional studies involving adults and children between 7 and 8 years of age. In that study, a 12-parameter affine transform was used to normalize the structural MR images of 20 children and 20 adults to a 12-subject average that was conformed to Talairach space (Talairach and Tournoux, 1988). The investigators manually traced points along 10 different sulci identifiable in specific planar sections on each of these normalized volumes, as well as points along the outer boundaries of the brain in the three cardinal orientations (axial, sagittal and coronal). They observed that the location and variability of these manually traced positions after normalization was fairly consistent across the age groups. Furthermore, using computer simulations of fMRI data with 5mm resolution, they demonstrated that the observed variability did not generate any significant spurious effects. Based on these observations, they concluded that: (i) stereotactic normalization does not significantly distort brain morphology between adults and children; and (ii) the small distortions observed do not limit the ability to compare functional activation between adults and children in such a space. Furthermore, they indicated that "more work comparing younger children's brains (below 7 years) to adult brains is needed before similar stereotactic approaches should be applied to that group."

In this study, we extend the work described above by: i) using more anatomically precise delineations of boundaries of cortical regions based on surface geometry as opposed to picking points on a particular (anatomically arbitrary) imaging plane; ii) investigating a younger and larger age-range of children (4.2–11.1 years of age); and iii) using FreeSurfer's surface-based registration approach in addition to two commonly used volume-based registrations (a linear 12-parameter affine transform and a nonlinear normalization approach from SPM 5) and a diffeomorphic volume-registration method (ANTS; Avants et al., 2006). The critical questions were whether it is valid to compare structural and functional brain images in child development (ages 4 to 11), and whether there are advantages for any particular kind of cortical image registration in this age range.

Materials and Methods

Participants

Participants were 31 right-handed children between the ages of 4.2 years and 11.2 years (mean = 7.33 years; std = 1.96 years). The children were selected from a larger sample recruited for a cross-sectional investigation of reading development. Children who were diagnosed with developmental dyslexia or were on psychiatric medication were excluded from this study. Written informed consent was obtained from the parents of the children and the Institutional Review Board at the Massachusetts Institute of Technology approved procedures.

MR protocol

T1-weighted structural scans of the children's brains were acquired using an MPRAGE sequence on a Siemens Magnetom Trio 3T scanner (16 subjects: TR: 2350 ms, TE: 3.45 ms, TI: 1100ms, Flip angle: 7 deg, Duration: 4:35 minutes, mean age: 8.79 yrs; 15 subjects: TR: 2000 ms, TE: 3.39 ms, TI: 900ms, Flip angle: 9 deg, Duration: 3:38 minutes, mean age: 5.77yrs; GRAPPA: ×2). The acceleration and the different scanning sequences were necessary to minimize scan time and therefore reduce the possibility of head movement, especially in the younger children. The structural scans had voxel dimensions of $1.3 \times 1.0 \times 1.3$ mm with an FOV of 256 mm and a matrix size of 256×256.

FreeSurfer processing of MR images

Using the FreeSurfer software suite (http://surfer.nmr.mgh.harvard.edu), we processed the structural MR image from each participant. Briefly, this processing includes skull-stripping (Segonne et al., 2004), subcortical segmentation (Fischl et al., 2002; Fischl et al., 2004a), intensity normalization (Sled et al., 1998), surface generation (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000), topology correction (Fischl et al., 2001; Segonne et al., 2007), surface inflation (Fischl et al., 1999a), registration to a spherical atlas (Fischl et al., 1999b) and thickness calculation (Fischl and Dale, 2000). FreeSurfer morphometric procedures have been demonstrated to show good test-retest reliability across scanner manufacturers and across field strengths (Han et al., 2006). The automated processing resulted in a topologically correct cortical surface for each hemisphere for every participant. These surfaces were then registered to an atlas coordinate system containing statistics summarizing cortical geometry using a spherical morphing procedure designed to align cortical folding patterns (Fischl et al., 1999b). The outlines of the extracted surfaces were overlaid on the T1-weighted images and visually inspected for accuracy of the automatic surface-extraction process. The volumes corresponding to inaccurate surfaces were edited using FreeSurfer (edits: white matter - 16; control points – 3 of 31 participants), re-processed to generate the surfaces, and visually inspected for accuracy to ensure surface alignment with white matter boundary. If the surfaces were still inaccurate, the process was repeated (the process was repeated for 3 of the participants). Details of editing procedures may be found on the FreeSurfer Wiki (http://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/TroubleshootingData).

The editing was necessary to ensure that the surfaces were as accurate as possible, which in turn ensured that the labels were accurate. While the necessity of such editing is decided based on aims of an experiment, these are often done routinely to rectify surfaces that show significant deviation from visually observable gray-white and gray-csf boundaries. Such editing does pose a conundrum for evaluation studies such as this one. On one hand, there is a need to make the labels as accurate as possible. On the other hand, these are the same brains used for evaluating the registration. However, while editing is more critical for quantitative measurements and creating accurate labels, it is much less important for registration. Supplementary figure S2 shows the curvature map of the most edited subject, before and after the edits. The overall curvature map remains highly similar.

Labeling protocol

The boundaries of three prominent sulci, the central sulcus, calcarine sulcus and Sylvian fissure, were manually delineated on both left and right hemisphere cortical surface models for every participant (Figure 1 shows an example from the youngest and oldest participant). These sulci were chosen based on the observation that they are consistently and accurately identifiable (Van Essen, 2005) on most participants regardless of age. Sulci and gyri are the most prominent macroanatomical features of the cerebral cortex and are most accurately identified on surface-based representations and not on slices. This surface labeling was performed and assessed using information from <u>both</u> volumes and surfaces. In particular, the labeling was guided by known anatomical extents from volumes and surfaces and the curvature information from the surfaces. Furthermore, we converted the surface labels to volume labels and all numerical comparisons were done using volumes. As such, these volume labels served as volume landmarks although they were originally created on the surface. The manual labeling did not use any information from the surface-based registration. A neuroanatomist (JA) inspected the labels to ensure accuracy.

The following criteria were used to label central sulcus: 1) On the lateral surface of the brain, we localized the precentral and postcentral gyrus on either sides of the central sulcus

visual examination; 2) The central sulcus was extended inferiorly at its same angle along an imaginary line and the intersection point of this line with the Sylvian fissure was verified to be at approximately one third the distance from Sylvian fissure's anterior limit; and finally, 3) In a subset of cases, the central sulcus extended superiorly and the tip of the central sulcus notched paracentral gyrus on the medial side.

To delineate the calcarine sulcus, we used the medial view to locate the calcarine fissure's anterior-most and posterior-most points. For the anterior-most point, the posterior end of the hippocampal fissure was used as a guideline and the anterior point of the calcarine was marked at the same angle, making certain that the parieto-occipital fissure, which is merging into the calcarine was not included. We utilized the inflated, white matter and pial surfaces in the FreeSurfer 'tksurfer' surface rendering utility to evaluate the sulcus. Finally, in a subset of cases, we opened the label in the FreeSurfer volume viewer ('tkmedit') to determine and demarcate the calcarine sulcus from the parieto-occipital fissure.

For the Sylvian fissure, the major landmarks were temporal lobe inferiorly (i.e. Heschl's gyrus and posterior of the Sylvian fissure) and frontal lobe superiorly (i.e. Inferior frontal lobule). The temporal lobe defined the inferior limit of the Sylvian fissure and in surfaces the hippocampal fissure was immediately next to the Sylvian fissure. In most cases, we edited the anterior-most part of the Sylvian fissure because the sulci of the frontal opercula created a larger surface than what is the true anatomy of the Sylvian fissure alone.

Analysis

In order to quantify the accuracy of the different registration approaches, we first converted each of the surface labels to individual image volumes (2 hemispheres × 3 labels = 6 volumes) for every participant. The volume labels used to compute the distance measure (see below) corresponded to a sheet of cortex at the gray matter-white matter boundary. Transforms were obtained using the volume- and surface-based registration approaches that mapped the T1-volume or the FreeSurfer surface of each participant to all other participants via templates. These transforms were then applied to the volume or surface labels from each participant (31 participants × 6 volumes × 30 participants × 4 registration methods). The volume registrations were computed using a 12-parameter affine transform (using an algorithm developed by Avi Snyder and incorporated and distributed in FreeSurfer), SPM 5 nonlinear normalization and diffeomorphic registration using ANTS. SPM's nonlinear normalization and the affine transform are two of the most common volumetric group-normalization approaches reported in the neuroimaging literature. The ANTS algorithm was shown to be significantly superior to the other two commonly used volumetric methods in adults (Klein et al., 2009). For evaluating the ANTS algorithm, we created a group-specific, skull-stripped template from these participants using the brain-only images generated by FreeSurfer and warped each participant to that template. The template can be downloaded from: http://www.mit.edu/~satra/research/pubdata. The affine and SPM normalization routines registered each participant to an average adult template (affine: 711-2C_as_mni_average_305.4dfp.img distributed with FreeSurfer; SPM: T1.nii distributed with SPM 5 and 8) in MNI space. A visual comparison of these templates is available as supplementary material. The cortical surfaces were matched using spherical registration (Fischl et al., 1999) using an existing adult atlas distributed with FreeSurfer. In order to compute the overlap measures described below, the surface labels were also converted into volumes where all the vertices in gray matter were labeled. We computed several metrics to determine how well the labels aligned across participants. These metrics were computed for all pairwise registrations. Details of all of these metrics can be found in an evaluation study (Klein et al., 2009). They are summarized here.

Modified Hausdorff distance (lower → better registration)—The modified Hausdorff distance (Dubuisson and Jain, 1994) is a distance measure that computes the similarity between two shapes and has been used previously for template-based image matching. If S and T are point sets from the source and target volume respectively, then the modified Hausdorff distance H_d between them is given by:

$$H_d(S,T) = \max(d(S,T), d(T,S))$$
$$d(X,Y) = \frac{1}{N_X} \sum_{x \in X} \min_{y \in Y} \|x - y\|_2$$

where $\| \cdot \|_2$ denotes the L2-norm and N_X denotes the number of elements in set X.

Typically, the point sets S and T are determined by finding all the voxels that lie on the boundary of the labels in the volumes corresponding to the source and the target. If the boundaries match exactly this distance will be 0 and will increase with greater differences between the boundaries. Thus lower values indicate better registration.

Jaccard coefficient, Dice coefficient and target overlap (higher → better registration)—The overlap metrics represent the extent of overlap between a target label T and the source label S when the label S is mapped via registration to the target volume. These metrics evaluate to 0 for no overlap and to 1 for perfect overlap. Thus higher values indicate better registration. The Jaccard and Dice coefficients are also referred to as "union" and "mean" overlap respectively.

$$\text{Jaccard coefficient} = \frac{|S \cap T|}{|S \cup T|}$$
$$\text{Dice coefficient} = 2\frac{|S \cap T|}{|S| + |T|}$$
$$\text{Target overlap} = \frac{|S \cap T|}{|T|}$$

where $| \cdot |$ indicates set count.

False positive and false negative error (lower → better registration)—In addition to the overlap metrics, two error measures were defined. A false positive error is a measure based on voxels that are labeled as belonging to the target label when the source was mapped to the target even though these voxels are not part of the target. A false negative error is a measure based on voxels of the target that were not labeled with the target label when the source was mapped to the target. A value of 0 for both error metrics implies a perfect registration. In practice, lower values indicate better registration.

$$\text{False positive} = \frac{|S \setminus T|}{|S|}$$
$$\text{False negative} = \frac{|T \setminus S|}{|T|}$$

where $\cdot \setminus \cdot$ indicates set difference and $| \cdot |$ indicates set count.

Each metric quantifies different components or features of registration accuracy. The union, mean and target overlap quantify intersection or overlap, but normalize the amount of overlap by different quantities and their use is dependent on the problem at hand. The false negative and false positive measures provide information about misclassification. Finally, the Hausdorff distance metric focuses on the similarity and proximity of boundaries between two regions.

Statistical analysis

Dependent variables were the computed outcome of the distance and overlap measures. Independent variables were participant Age, cortical Region (label), cortical Hemisphere and the registration Methods. A non-parametric resampling with replacement method with 10,000 replications was used to test the main effects of the independent variables and the interaction effects of Age with Method, Region and Hemisphere using a factorial design with Age as a between-subjects factor and Region, Hemisphere, and Method as within-subject factors. This non-parametric approach allowed treating subjects as random effects *even though* the dependent metrics were derived in a pairwise cross-subjects manner. In order to maintain treating subjects as random effects, age was treated as a dichotomous effect. The sample was divided into an older group, n=17, and a younger group, n=14, based on the median age of the whole group. This is similar to the comparison between the child group and the adult group in the Burgund et al. (2002) study.

Results

This section presents the results of calculating the distance and overlap between corresponding manually labeled brain regions from 31 children after volume-based and surface-based registration. Two main effects were observed. First, there was no significant effect of age on the similarity measures of any kind of registration. Second, the kind of registration did influence measures (higher overlap and smaller distance) of cortical registration accuracy, with surface-based registration providing significantly higher accuracy than any form of volume based registration, and, diffeomorphic registration providing significantly higher accuracy than the other two volume-based approaches (as in Klein et al. 2009).

Distance Measure

The surface-based approach was significantly more accurate than the volume-based approaches (mean difference = 1.56mm) when quantified using modified Hausdorff distance ($F = 912.8$; $p < 0.001$; Table 1). This difference was observed across all labeled regions in both hemispheres (Figure 2). There was a significant effect of Region ($F = 20.3$; $p < 0.001$) on modified Hausdorff distance, with the Sylvian fissure having the lowest distance. There was also a significant effect of hemisphere ($F = 5.6$; $p < 0.018$), with the right hemisphere regions showing better proximity after registration than the left hemisphere regions. There was no significant effect of Age or interactions of Age with any of the other three independent variables. For both the surface-based and the poorest volume based registration (the affine), the error in registration was smaller than typical fMRI voxel sizes, when comparing specific individuals and a specific region (Figure 3) or across groups and regions (Figure 4). In addition to treating age as a dichotomous effect, we compared Hausdorff distance between overlapped regions in the two hemispheres separately for the youngest and the oldest participant to all other participants, thus comparing it as a function age. No age-related trend was observed (Figure 5). ANTS performed significantly better than SPM nonlinear normalization ($F = 868.6$; $p < 0.001$) and affine normalization ($F = 363.7$; $p < 0.001$).

Overlap Measures

The surface-based approach was significantly more accurate than the volume-based approaches when quantified using Target Overlap ($F = 2103.4$; $p < .001$). There were also significant effects of Hemisphere and Region (Figure 6; Table 2). Similar to the results from the distance measure, the right hemisphere regions showed greater overlap than the left hemisphere regions and the Sylvian fissure showed the greatest overlap among regions.

There was no significant effect of Age or interactions of Age with any of the other independent variables. The results from comparing the other overlap measures (see Methods section for details) across surface-based and volume-based registration showed a similar pattern of results as Target Overlap. These results across groups and regions are available as supplementary material.

Discussion

Technological advances in non-invasive neuroimaging offer tremendous potential for improving our understanding of typical and atypical cognitive development. This requires quantitative comparison of structural and functional data from children whose brain morphology changes well into adolescence. Such a comparison is facilitated by the alignment or registration of individual brain into a common reference frame. The current results demonstrate that spatial normalization into a common reference frame is feasible for participants ranging from 4.2 to 11 years of age, without the introduction of age-related biases. Taken together with evidence for the validity of including children 7–8 years of age and adults in a common reference frame (Burgund et al., 2002), the present findings indicate that it is appropriate to study structural and functional brain development in a common reference frame.

The results also show that FreeSurfer's surface-based registration improves the accuracy of aligning cortical landmarks in children's brains compared to commonly-used volume-based registration approaches. This improvement stems from FreeSurfer's approach to registration. FreeSurfer extracts the cortical sheet from a brain image as parameterized-surfaces and therefore can generate additional information about morphological properties such as curvature and thickness of cortex. When registering one brain to another, such information systematically improves the accuracy of aligning structural (e.g., sulci, gyri, connectivity) and functional properties (e.g., brain activity patterns, gene expression). Therefore surface-based registration using spherical coordinates should provide a more accurate reference frame for comparison of functional and structural imaging results from cortical regions across children and adults. The relative registration-accuracy across the volume-registration methods was similar to those observed for adults (Klein et al. (2009). In a comparison of surface- and volume- registration techniques applied to labeled brains of adult participants, Klein et al. (2010) observed no significant difference between FreeSurfer and ANTS. However, the results from the present study on children indicate that FreeSurfer is more accurate than ANTS. Therefore, topographic properties used by FreeSurfer (e.g., curvature) may provide better features for matching across participants in this age range than the intensity-derived properties used by the volume-registration algorithms.

For all registration methods, the mean Hausdorff distance (Figure 4) between registered regions is less than the acquisition voxel sizes used in the vast majority of fMRI studies (typically greater than 3mm) and much less than the effective functional resolution obtained after smoothing the data. FreeSurfer and ANTS performed significantly better than the other volume registration techniques, they are also more computationally intensive. Therefore, in certain situations where the underlying functional data is highly smoothed, the analysis may not benefit from the additional computational complexity of FreeSurfer and ANTS over the simpler methods. However, for analysis of high-resolution fMRI with minimal smoothing, these advanced registration methods may be more appropriate.

In many studies, individual-specific functional and structural region of interest (ROI)-based analysis approaches are typically used to alleviate problems that arise from misalignment during whole-brain registration (e.g., in development, Golarai et al., 2007). The registration methods evaluated here would be more useful in studies where no specific functional

localizers are available or structural delineation is too time-consuming, or no strong *a priori* reasons exist for expecting an effect to be contained within an easily defined ROI. In such studies, functional and structural ROIs are often delineated using automatic methods (e.g., Fischl et al., 2004; Hinds et al., 2009) that typically rely on registering a brain to an average template or pairwise registration of brains.

In this study two small, but significant, effects of hemisphere and region on the similarity measures were observed. First, the mean registration error of the Sylvian fissure was smaller than that of the calcarine and central sulci independent of the methods used. Klein et al. (2009) observed a similar effect as a function of size of regions for overlap measures. Overlap measures may show region-size related bias (e.g., two small regions may show a large change if a few voxels change and vice-versa for large regions). On the other hand, distance measures should not be subject to such a bias. But the present finding is observed for both distance and overlap measures. This finding may reflect superior accuracy for larger regions (Sylvian fissures) than smaller regions (calcarine and central sulci), because increased registration accuracy measured by distance may stem from the fact that larger regions can provide more "information" (whether intensity or curvature) to the registration algorithms. Hence these algorithms may work better for larger regions that afford greater details.

Second, the mean registration accuracy of right hemisphere regions was significantly greater than the mean accuracy of left hemisphere regions. Although numerous studies have investigated hemispheric differences, these have primarily been based on shape, extent, or volume differences between hemispheres. These results do not directly translate to corresponding differences in similarity metrics between hemispheres. The present results would indicate that the right hemisphere has a more stable folding pattern in this age-range. Speculatively, this may result from greater influences of language experience on the language-dominant left hemisphere (although the lack of an interaction with age suggest that these experiences occur before age 5). However, given the limited set of regions, methods and participants, further studies are needed to validate these observations.

Caveats and limitations of the study

A surface-based approach, as evaluated here, ignores subcortical structures (e.g., basal ganglia, cerebellum) that are often critically important towards understanding brain function and malfunction in several neurodevelopmental disorders. Registration approaches that combine surface and volume-based methods (e.g., Zollei et al., 2010, Joshi et al., 2007, Postelnicu et al., 2009) provide an alternative that includes subcortical structures. It is also important to note that this study compared only three volume-based registration methods to one surface-based registration approach applied to a limited-dataset containing three primary fissures in each hemisphere. These sulci were chosen based on their cross-subject stability, as well as to be distributed across disparate parts of the brain. It is possible that one might observe larger variations than those observed here for other parts of cortex. Although sulcal identification was performed independently of the registration procedures that were evaluated, surface-based registration algorithms do rely on topographic properties of the surface (e.g., curvature). Expert labelers typically use similar geometric and topographic information to label regions. Therefore, any labels derived from sulcal and gyral landmarks, as was the case in this study, are likely to benefit surface-based registration algorithms more than volume-based ones. However, sulci and gyri are the most prominent macroanatomical features of the cerebral cortex and we believe that these are best identified on surface-based representations and not on slice-based representations. Finally, we created a group-specific template to evaluate ANTS, while FreeSurfer and the other volume-registration methods relied on their own average templates based on adults. The FreeSurfer affine template and the SPM T1 template contain skull-related information, are blurrier and therefore are not

optimal for using with ANTS. The ANTS template cannot be used with SPM and FreeSurfer for the opposite reasons (too sharp, no skull). Since the intent of the study was to compare algorithms and not optimize their use, we ran the software as commonly reflected in the literature. However, prior studies have reported that custom templates that are representative of the participant population provide better registration targets compared to templates supplied with current registration software packages (Yoon et al., 2009; Klein et al., 2010). Thus, creating custom group-specific template may improve registration accuracy for the other methods. However, based on the results of Klein et al. (2009), we do not expect the SPM normalization and the affine registration methods to outperform ANTS even after custom template creation.

Conclusions

The results of this study indicate that one can register or normalize brains from a group of children between 4.2 and 11 years of age without the introduction of an age-related bias in accuracy of normalization. Furthermore, increased accuracy of such registration can be obtained using FreeSurfer's surface-based registration technique as compared with volume normalization methods. As functional data is obtained with greater consistency, reliability and at higher resolution, it becomes vital to compare such data from multiple subjects and studies with as accurate an alignment between them as possible. In particular, these registration approaches enable accurate comparison of structure and function from children. Being able to compare the wide array of pediatric neuroimaging data should greatly enhance in our understanding of typical development and developmental disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

## References

Amunts K, Schleicher A, Brgel U, Mohlberg H, Uylings HB, Zilles K. Broca's region revisited: cytoarchitecture and intersubject variability. J Comp Neurol. 1999; 412(2):319–341. [PubMed: 10441759]

Ardekani BA, Guckemus S, Bachman A, Hoptman MJ, Wojtaszek M, Nierenberg J. Quantitative comparison of algorithms for inter-subject registration of 3d volumetric brain MRI scans. J Neurosci Methods. 2005; 142(1):67–76. [PubMed: 15652618]

Ashburner J, Andersson JL, Friston KJ. High-dimensional image registration using symmetric priors. Neuroimage. 1999; 9(6 Pt 1):619–628. [PubMed: 10334905]

Avants BB, Schoenemann PT, Gee JC. Lagrangian frame diffeomorphic image registration: Morphometric comparison of human and chimpanzee cortex. Med Image Anal. 2006; 10:397–412. [PubMed: 15948659]

Burgund ED, Kang HC, Kelly JE, Buckner RL, Snyder AZ, Petersen SE, Schlaggar BL. The feasibility of a common stereotactic space for children and adults in fMRI studies of development. Neuroimage. 2002; 17(1):184–200. [PubMed: 12482076]

Cointepas Y, Mangin J-F, Garnero L, Poline J-B, Benali H. BrainVISA: software platform for visualization and analysis of multi-modality brain data. NeuroImage. 2001 June.13(6):98.

Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr. 1994; 18(2):192–205. [PubMed: 8126267]

Cowan WM. The development of the brain. Sci Am. 1979; 241(3):113–133. [PubMed: 493917]

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage. 1999; 9:179–194. [PubMed: 9931268]

Dale AM, Sereno MI. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J Cogn Neurosci. 1993; 5:162–176.

Davatzikos C, Prince JL, Bryan RN. Image registration based on boundary mapping. IEEE Trans Med Imaging. 1996; 15(1):112–115. [PubMed: 18215894]

Desikan RS, Sgonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006; 31(3):968–980. [PubMed: 16530430]

Drury HA, Essen DCV, Anderson CH, Lee CW, Coogan TA, Lewis JW. Computerized mappings of the cerebral cortex: a multiresolution flattening method and a surface-based coordinate system. J Cogn Neurosci. 1996; 8(1):1–28. [PubMed: 11539144]

Dubuisson M, Jain A. A modified hausdorff distance for object matching. Pattern Recognition. 1994

Evans AC, Collins DL, Milner B. An MRI-based stereotactic atlas from 250 young normal subjects. Journal Soc Neurosci Abstr. 1992; 18:408.

Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM. 3D statistical neuroanatomical models from 305 MRI volumes. Proc IEEE-Nuclear Science Symposium and Medical Imaging Conference. 1993:1813–1817.

Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A. 2000; 97:11050–11055. [PubMed: 10984517]

Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans Med Imaging. 2001; 20:70–80. [PubMed: 11293693]

Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo BTT, Mohlberg H, Amunts K, Zilles K. Cortical folding patterns and predicting cytoarchitecture. Cereb Cortex. 2008; 18(8):1973–1980. [PubMed: 18079129]

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33:341–355. [PubMed: 11832223]

Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT, Dale AM. Sequence-independent segmentation of magnetic resonance images. Neuroimage. 2004a; 23(Suppl 1):S69–84. [PubMed: 15501102]

Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage. 1999a; 9:195–207. [PubMed: 9931269]

Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp. 1999b; 8:272–284. [PubMed: 10619420]

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. Automatically parcellating the human cerebral cortex. Cereb Cortex. 2004b; 14:11–22. [PubMed: 14654453]

Gholipour A, Kehtarnavaz N, Briggs R, Devous M, Gopinath K. Brain functional localization: a survey of image registration techniques. IEEE Trans Med Imaging. 2007; 26(4):427–451. [PubMed: 17427731]

Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JD, Grill-Spector K. Differential development of high-level visual cortex correlates with category-specific recognition memory. Nat Neurosci. 2007; 10(4):512–522. [PubMed: 17351637]

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage. 2006; 32:180–194. [PubMed: 16651008]

Hinds O, Polimeni JR, Rajendran N, Balasubramanian M, Amunts K, Zilles K, Schwartz EL, Fischl B, Triantafyllou C. Locating the functional and anatomical boundaries of human primary visual cortex. Neuroimage. 2009

Hinds OP, Rajendran N, Polimeni JR, Augustinack JC, Wiggins G, Wald LL, Rosas HD, Potthast A, Schwartz EL, Fischl B. Accurate prediction of v1 location from cortical folds in a surface coordinate system. Neuroimage. 2008; 39(4):1585–1599. [PubMed: 18055222]

Hömke L. A multigrid method for anisotropic PDEs in elastic image registration. Numerical Linear Algebra with Applications. 2006; 13(2–3):215–229.

Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 2002; 17(2):825–841. [PubMed: 12377157]

Joshi AA, Shattuck DW, Thompson PM, Leahy RM. Surface Constrained Volumetric Brain Registration Using Harmonic Mappings. IEEE Transactions on Medical Imaging. 2007; 26(12): 1657–1669. [PubMed: 18092736]

Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage. 2009; 46(3):786–802. [PubMed: 19195496]

Klein A, Ghosh SS, Avants B, Yeo BTT, Fischl B, Ardekani B, Gee JC, Mann JJ, Parsey RV. Evaluation of volume-based and surface-based brain image registration methods. NeuroImage. 2010; 51(1):214–220. [PubMed: 20123029]

Miller MI, Christensen GE, Amit Y, Grenander U. Mathematical textbook of deformable neuroanatomies. Proc Natl Acad Sci U S A. 1993; 90(24):11944–11948. [PubMed: 8265653]

Nieto-Castanon A, Ghosh SS, Tourville JA, Guenther FH. Region of interest based analysis of functional imaging data. Neuroimage. 2003; 19(4):1303–1316. [PubMed: 12948689]

Postelnicu G, Zollei L, Fischl B. Combined volumetric and surface registration. IEEE Trans Med Imaging. 2009; 28(4):508–22. [PubMed: 19273000]

Qiu A, Miller MI. Cortical hemisphere registration via large deformation diffeomorphic metric curve mapping. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv. 2007; 10:186–193.

Schormann T, Zilles K. Three-dimensional linear and nonlinear transformations: an integration of light microscopical and MRI data. Hum Brain Mapp. 1998; 6(5–6):339–347. [PubMed: 9788070]

Segonne F, Pacheco J, Fischl B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans Med Imaging. 2007; 26:518–529. [PubMed: 17427739]

Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 1998; 17:87–97. [PubMed: 9617910]

Rademacher J, Caviness VS, Steinmetz H, Galaburda AM. Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cereb Cortex. 1993; 3(4):313–329. [PubMed: 8400809]

Talairach, J.; Tournoux, P. Co-planar stereotaxic atlas of the human brain. Thieme publishers; Germany: 1988.

Thompson P, Toga AW. A surface-based technique for warping three-dimensional images of the brain. IEEE Trans Med Imaging. 1996; 15(4):402–17. [PubMed: 18215923]

Toga AW, Thompson PM, Sowell ER. Mapping brain maturation. Trends Neurosci. 2006; 29(3):148–159. [PubMed: 16472876]

Tosun D, Prince JL. A geometry-driven optical flow warping for spatial normalization of cortical surfaces. IEEE Trans Med Imaging. 2008; 27(12):1739–53. [PubMed: 19033090]

Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH. An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Inform Assoc : JAMIA. 2001; 8 (5): 443–459.

Van Essen DC. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. NeuroImage. 2005; 28(3):635–662. [PubMed: 16172003]

Yoon U, Fonov VS, Perusse D, Evans AC. The effect of template choice on morphometric analysis of pediatric brain data. Neuroimage. 2009; 45(3):769–777. [PubMed: 19167509]

Zollei L, Stevens A, Huber K, Kakunoori S, Fischl B. Improved Tractography Alignment Using Combined Volumetric and Surface Registration. Neuroimage. 2010
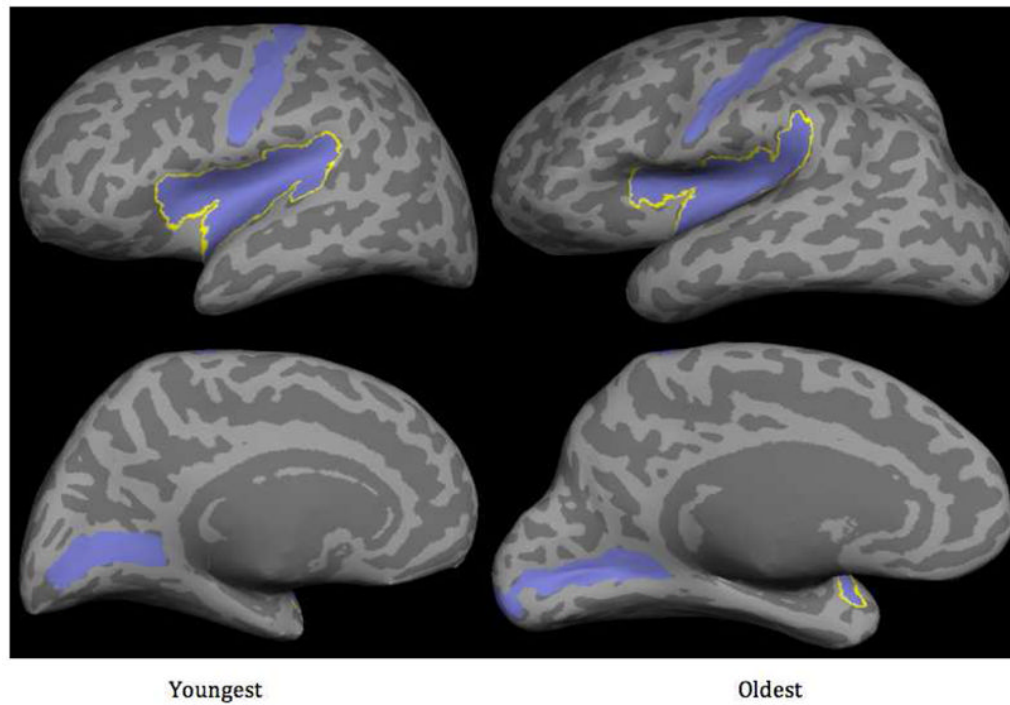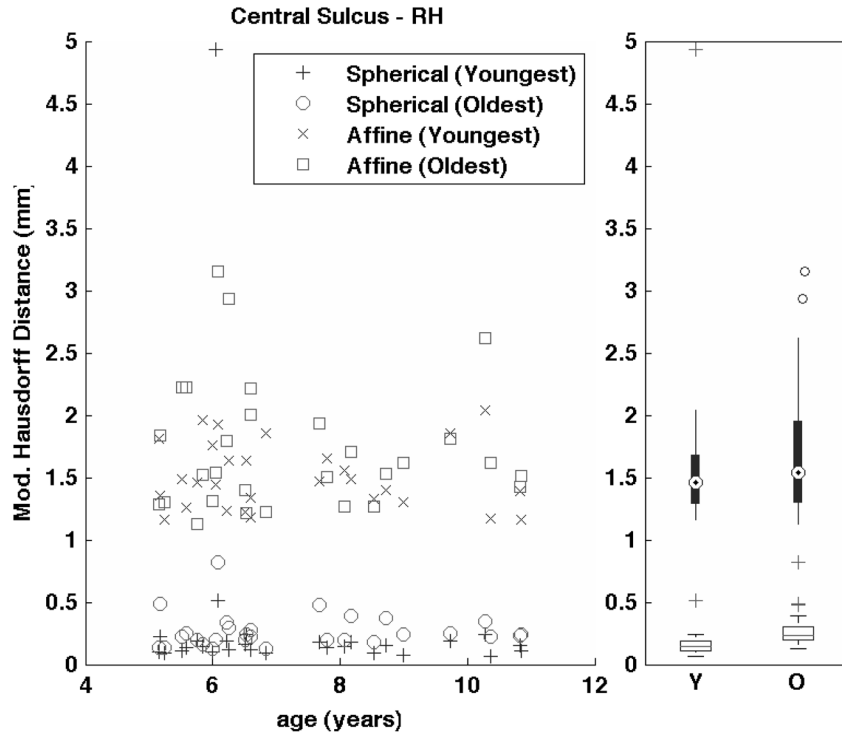
Youngest   Oldest

Figure 1.
Manual labels displayed on inflated left hemisphere surfaces of the youngest (4.2 yrs; left) and oldest participant (11.16 yrs; right). The central sulcus and sylvian fissure (yellow outline) labels are displayed in the top row and the calcarine sulcus labels are displayed in the bottom row.
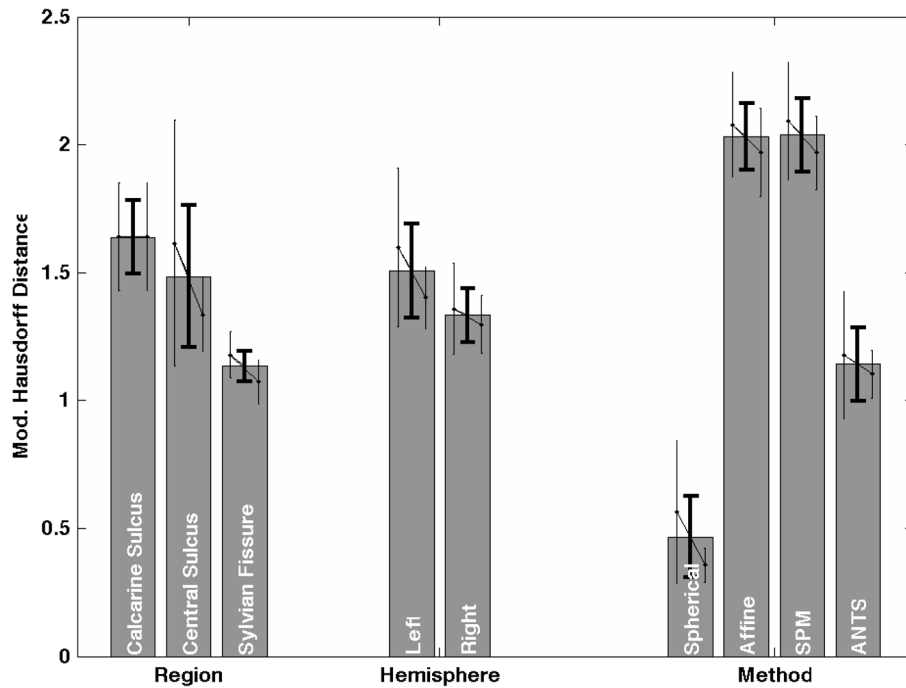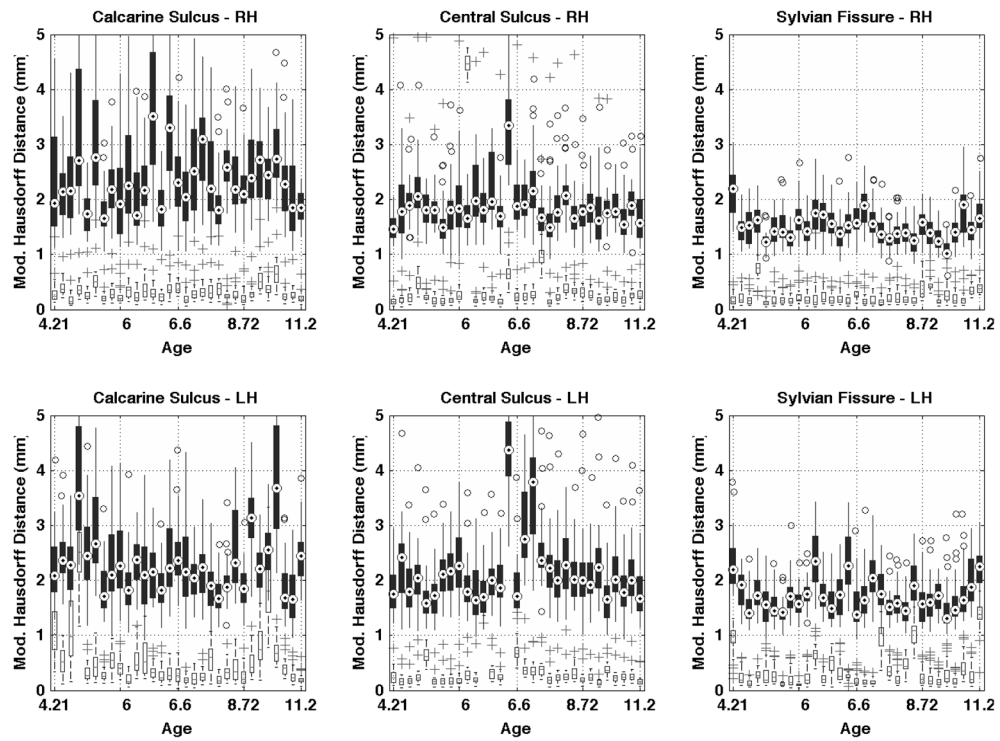
Figure 2.
Comparison of surface and volume registration accuracy using modified Hausdorff distance. Mean effect sizes and confidence intervals from a resampling analysis of Region, Hemisphere and Method factors and their interactions with Age on the modified-Hausdorff distance. The volume registration refers to the 12-parameter affine transform. The middle error bar on each column reflects the 90% confidence interval for each simple main effect. The left and right error bars on each column reflect the 90% confidence interval for that effect in the young age group and the old age group respectively. Lower values reflect more accurate registration.
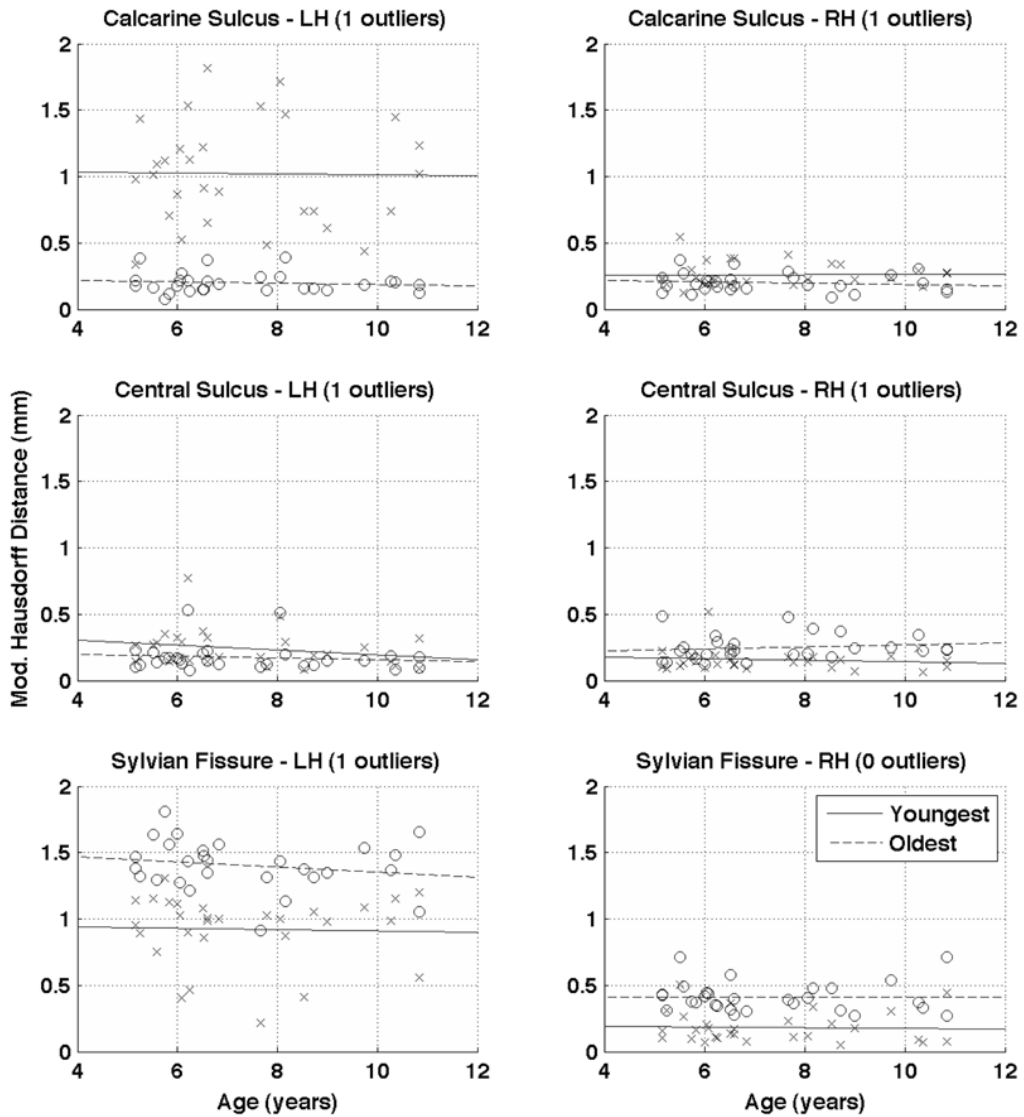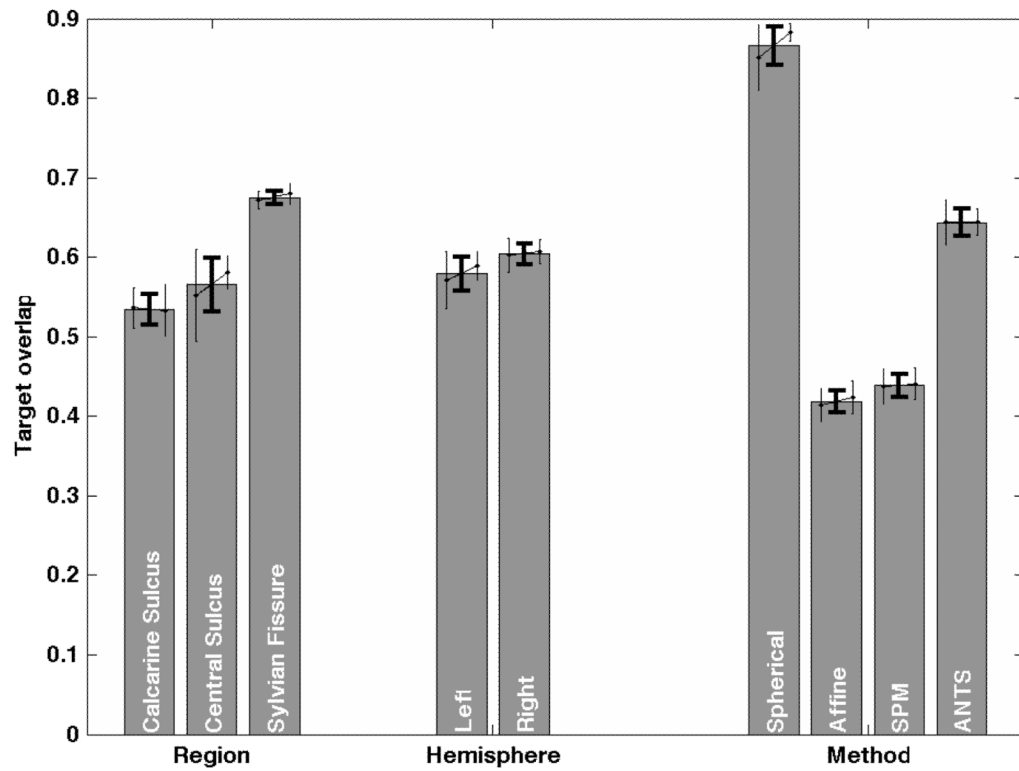
Figure 3.
Modified Hausdorff Distance between the central sulcus in the youngest (4.2 yrs) and in the oldest participant (11.16 yrs) and the central sulcus in all of the other participants. To preserve clarity, only affine volume-based (worst accuracy) and surface registration (best accuracy) distances are reported. The left panel shows the individual distances for each pairwise comparison. The right panel shows a summary boxplot of the same data with the affine registration results in the shaded plots and the surface-based registration data in unshaded boxes. The median values are shown with the filled circle for affine registration and the horizontal line inside the box for the surface-based registration. The length of the shaded part and the length of the box denote the interquartile range. Outliers for the affine registration are denoted with the symbol 'o' while those for the surface-based registration are denoted with the symbol '+'. Lower values indicate more accurate registration.

Figure 4.
Modified Hausdorff distance between a hemispheric region from one participant and the corresponding region of all other participants. The summary plots are arranged as a function of increasing age of participants. To preserve clarity, only the affine volume-based registration results are reported with the shaded plots and the surface-based registration data with unshaded boxes. See Figure 3 for details of the boxplot. The top and bottom rows show the distance of the right and left hemisphere regions respectively. The three columns show the distances computed for the calcarine sulcus, central sulcus and Sylvian fissure. Lower values indicate more accurate registration.

Figure 5.
Graphical relation between modified Hausdorff distance and age. Each subplot shows the regression slopes for the modified Hausdorff distance between regions from the youngest and the oldest participants to the corresponding regions in all other participants after undergoing spherical registration. To eliminate the effect of strong outliers, data points were eliminated if they were greater than 3 standard deviations from the mean. The outlier count (out of 29) is listed in the title for each subplot. (LH = Left hemisphere, RH = Right hemisphere, x - distance from youngest participant, o - distance from oldest participant)

Figure 6.
Comparison of surface and volume registration accuracy using target overlap. Mean effect sizes and confidence intervals from a resampling analysis of Region, Hemisphere and Method factors and their interactions with Age on the Target Overlap Measure. The middle error bar on each column reflects the 90% confidence interval for each simple main effect. The left and right error bars on each column reflect the 90% confidence interval for that effect in the young age group and the old age group respectively. Higher values reflect more accurate registration.

Table 1

Statistical results from a resampling analysis of main effects of Method, Hemisphere, Region and Age and their interactions on the modified-Hausdorff distance (Figure 2).

	Mean	F	P
Method	2.64	912.8	0.000
Hemisphere	0.17	5.6	0.018
Region	0.38	20.3	0.000
Age	0.13	1.0	0.307
Age*Method	0.20	1.6	0.203
Age*Hemi	0.14	1.0	0.306
Age*Region	0.33	1.1	0.286

Table 2

Statistical results from a resampling analysis of main effects of Method, Hemisphere, Region and Age and their interactions on the Target Overlap measure (Figure 6).

	Mean	**F**	**P**
Method	0.72	2103.4	0.000
Hemisphere	0.02	6.9	0.009
Region	0.11	92.9	0.000
Age	0.01	0.6	0.454
Age*Method	0.05	3.3	0.069
Age*Hemi	0.01	0.6	0.443
Age*Region	0.04	0.8	0.357