

Tina Memo No. 2004-007
Published in Medical Image Analysis
<http://dx.doi.org/10.1016/j.media.2008.10.006>

Cerebral Cortical Thickness Measurements

M.L.J. Scott and N.A. Thacker

Last updated
15 / 2 / 2005



Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Author's note: an updated version of this memo was published in Medical Image Analysis. A preprint is available as TINA Memo No. 2009-001.

Cerebral Cortical Thickness Measurements

M.L.J. Scott and N.A. Thacker
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
neil.thacker@manchester.ac.uk

1 Introduction

1.1 Motivation

The assessment of human cerebral cortical thickness (the thickness of the cortical grey matter ribbon) has massive clinical importance in the determination of pathology and in assessing the processes of “normal” brain maturation and ageing. As the brain develops and matures, from infancy through adulthood to old age, various changes occur in the brain tissues, such that it is difficult to pinpoint when maturation ends and degeneration begins. Sowell et al. [28] have extensively investigated these changes from early childhood using MR. They have found that white matter density and volume continue to increase throughout childhood and adolescence, whereas grey matter density (which they claim to be representative of cortical thickness) increases over the same period, but grey matter volume increases in early childhood and then declines after puberty. These changes could be accounted for by the process of myelination of the white matter which proceeds throughout childhood and adolescence.

Grey matter volume loss (on MR) is seen throughout adulthood to old age, but white matter volume increases until the age of 40, and then declines. Again, this may be due to continued myelination of the white matter, which has been shown using post-mortem studies to continue until at least 30 years of age. In this study Sowell et al. [28] investigate grey matter density (GMD) in 176 normal subjects from 7 to 87 years and find a significant non-linear decrease of GMD with age in the dorsal frontal and parietal cortices, whereas in the left posterior temporal region, GMD increases with age until 30, after which rapid decline occurs.

Double et al. [8] have shown however, using post-mortem studies of normal elderly brains that there appears to be no loss in cortical grey matter volume with age, and that the observed non-significant 2ml/year decline in cerebral volume is consistent with being entirely due to the (also) non-significant decline in white matter volume per year. This white matter loss itself was shown to be significant only in frontal regions. The authors claim that this result is not incompatible with MR studies which do show decreases in regional grey matter.

In addition to cortical volume and thickness, another property of the cortical ribbon which has major implications in disease is the degree of gyrification of the cortex. Gyrification of the human brain begins after the 6th month of gestation, undergoes an increase in activity shortly after birth when there is rapid proliferation of grey matter, and is stable in early childhood. Sulcal/gyral positioning is certainly genetically regulated, but the extent of environmental effects pre and post-natally are poorly known, and not at all in later life. With ageing, the brain typically becomes more atrophic, with decreased grey matter volume and increased cerebro-spinal fluid (CSF). Magnotta et al. [19] showed that the gyri of the cortex become more sharply and steeply curved while the sulci become more flattened and less curved. They also showed that cortical thinning occurs at a greater rate in males than females, begins in the fourth decade of life and continues with age.

As well as appearing to be associated with normal ageing, differences in cortical thickness/volume and gyral complexity are implicated in neurodegenerative and developmental diseases. Double et al. [8] demonstrated in a post-mortem study of Alzheimer’s Disease that loss of white matter was concomitant with normal ageing, but there was a significant reduction in cortical volume compared with normal brains, particularly in the medial temporal lobe. The decrease also appeared to be greater in those subjects with greater cognitive decline. Huntington’s Disease exhibits progressive striatal degeneration, but relatively little is known about cortical degeneration. Rosas et al. [25] demonstrated that Huntington’s Disease subjects show heterogeneous thinning of the cortex, that thinning occurs in the early stages of the disease, differs in the different disease stages and that there is an anterior-posterior trend in thinning with disease progression. The idea of heterogeneous thinning, possibly due to a pathological cascade in the striatal neurons, is also illustrated in post-mortem studies [27]. Sailer et al. [26] show that in Multiple Sclerosis (MS), the mean overall cortical thickness was significantly reduced compared to controls. The MS subjects had significant focal thinning in frontal and temporal regions, and in the motor cortex in longstanding/severe disease. Mean thinning also correlated with disability, disease duration and lesion load.

Converse to the general thinning which appears to be seen in degenerative disease, Ballmaier et al. [2] have demonstrated significant grey matter increases in elderly depressed patients, relative to controls, in the orbitofrontal cortex, parietal cortex and left temporal cortex. However, focal decreases in grey matter were also seen in the orbitofrontal cortex, adjacent to the increased regions.

With respect to developmental diseases, Kuperberg et al. [16] showed that subjects with chronic schizophrenia show cortical thinning, especially in the pre-frontal and temporal cortices, which they claim may reflect neuropathological abnormalities in cortical structure. Narr et al. [22] also show a frontal lobe difference from normals in schizophrenia, namely that male schizophrenic patients have significant increases (compared to their male counterparts) in cortical folding (hypergyria) in the right superior frontal cortex, which they believe to be due to disturbed gyral formation in utero.

Levitt et al. [17] have also shown that cortical sulcal mapping in Autism may be useful in elucidating any abnormalities in cerebral organisation in this disease, showing that there are positional shifts in major sulci which may be indicative of delayed maturation in these brain regions.

1.2 The Concept of Cortical Thickness

Although having discussed diseases in which changes in cortical thinning occur, we have not yet explained what is meant by cortical thickness, nor how people go about measuring it (see section 1.3). Indeed, many approaches do not directly measure thickness, but instead use measures of grey matter density (typical of Voxel Based Morphology (VBM) methods) or probabilities of tissue occurrence. The main difficulties with these techniques have been well discussed in the literature [3, 5, 7]. Aside from the statistical difficulties associated with forming valid conclusions, the variation in topological structure prevents the interpretation of the most significant changes due to a simple volume change. This has been described as [11], “implications for the sensitivity of any morphometric technique to detect changes in regions of high variance”.

The fact that the human cortex is so convoluted is part of the excitement of investigating the brain, but at the the same time, cortical folding provides two specific challenges. The first is the problem of how to measure a thickness of a curved structure, the second is the determination of the boundaries of the grey matter ribbon, particularly when the opposing banks of two gyri are sufficiently close that there is no CSF between them. Assuming for the moment that the grey-white and grey-CSF boundaries have been accurately segmented in 3D, there are various measures of distance between the two surfaces that might be employed. First, if an active shape model algorithm using correspondences has been used to create the surfaces, the distance between the corresponding points on the two surfaces can be used. This is however dependent on the correspondences being accurate, which you could never be sure of. Secondly, the minimum distance (see Fig. 1) between a given point (a) on one surface (S1) and the closest point (b) on surface two (S2) can be used. Thirdly, the distance along the surface normal from a point (a) on S1 to S2 (c) can be used. Of these two methods, this is the least biased, whereas the minimum distance will always produce a shorter average cortical thickness than the surface normal method. However, with both of these methods, on average if you start from the end points on S2 (b and c) and apply the same distance criteria, you will not end up back at the previously determined start points (a) on S1, but at points b' and c', as illustrated.

Another problem with using start points and lines from one surface to the other is that of the spacing of the start points. For example (see Fig. 2), when traversing the grey matter starting at the white-grey matter boundary, and using equally spaced points along the boundary as the starting points for the lines, assuming that surface normals are used, then the curve sulcal fundi will always be oversampled, relative to a flat section of cortex, and the gyral crowns will always be undersampled, which may bias a mean result if there are any gyral/sulcal differences in thickness. Fischl and Sereno [9], and Jones et al. [13] show a decrease in cortical thickness with sulcal depth. The effect would be less severe if the median is taken, providing that it is taken over a sufficiently small cortical region.

One very elegant solution to these problems is to use Laplace's equation [13] to calculate field lines through the cortex (see Fig. 3) These are equivalent to lines of electric potential and as such never cross and always follow the shape of the cortex (they are diffeomorphisms). The thickness of the cortex at any given point now becomes the sum of the lengths of the surface normals at each field line, where the start point at each field line is the end point of the surface normal from the previous field line. This ensures that every point has a unique and reversible point on the opposing surface. The assumed model also has some analogy with reality [1], where the cortical mantle is divided into 6 laminal layers.

Regardless of the algorithms used to detect the surfaces and calculate thickness, it is necessary to segment the tissues of the brain (CSF, GM and WM) in order to detect the surfaces. There are four MR-related issues which should be mentioned here:

1. The contrast between the tissues needs to be sufficient in order to define a boundary (this may not always

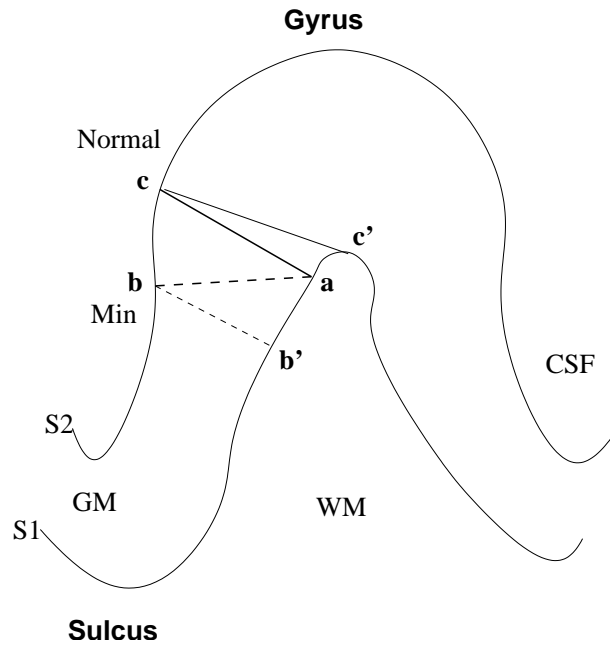


Figure 1: Diagram illustrating the use of surface normals and the minimum distance in estimating cortical thickness.

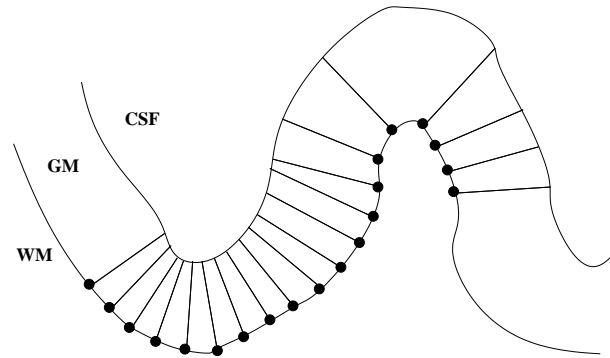


Figure 2: Diagram illustrating that when sampling the grey-white matter boundary evenly using surface normals, the sulcal fundi will be oversampled and the gyral crowns undersampled relative to a flat section of cortex.

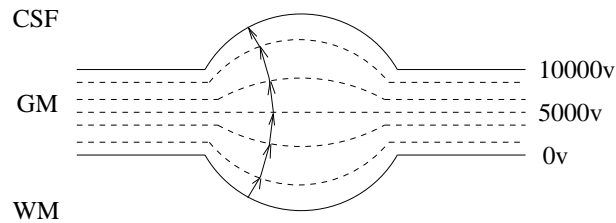


Figure 3: Diagram illustrating the use of electric potential field lines in modelling the cortical grey matter. The grey-white matter boundary is at 0v, the grey-CSF boundary at 10000v, and the distance between the two boundaries can be modelled as the path an electron would take between the two.

be true in disease, the very groups for which cortical thickness is required).

2. The segmentation used must be able to detect partial voluming of tissues, again, the level of partial voluming should not be so great that a boundary cannot be determined.
3. Through-plane inhomogeneities may result in the tissues having different mean values in each imaged slice.
4. In order to perform skull stripping/reliable segmentations, the brain tissue (GM and WM) should be differentiable from the skull. Merely relying upon intervening CSF to separate skull from tissue is not optimal.

Details of the approaches taken by various groups to these problems are given below.

N.B. Jones et al. [13] using the Laplacian approach eliminate through plane magnetic field inhomogeneities by normalising the grey matter peaks to the mode over all the slices. In-plane inhomogeneity is removed using a Fourier transform approach. The ability to detect the grey-CSF boundary in tight sulci is increased by skeletonising the image by subtracting an average image from the original.

1.3 Approaches

The goal of early cortical work appears to have been to represent the whole cortical surface with the topology of the sphere (to flatten out the hemispheres), with the essential point that the surface should not self-intersect where two opposing gyri are in close proximity. Miller et al. [21] developed methods for constructing not actual thicknesses, but probabilities of occurrence of tissues at given sulcal depths. They initially use a Bayesian tissue segmentation on the entire volume using the EM algorithm and tissue mixture densities, and generate surfaces to fit the resulting GM/WM interface using active surface methods and isocontouring. The GM/WM cortical surface is created from the isosurface using tetrahedral decomposition on each voxel [12] (similar to the marching cubes algorithm), such that the surface is composed of triangles that bound the tetrahedra. Then a boundary shell of voxels intersecting with this surface is created and for all those voxels exterior to the shell (those which should be in the cortex) the distance to the GM/WM surface is calculated. Separate histograms for GM, WM and CSF are constructed, containing the frequency of occurrence of the tissue as a function of distance to the GM/WM surface. From these it is possible to calculate the probability that a voxel at a given distance is of a certain tissue type. Joshi et al. [14] illustrate this technique in the macaque brain, and Ratnanather et al. [24] extend the technique to estimate the surface area of the GM/WM interface in human medial prefrontal cortex.

MacDonald et al. [18] appreciate that deformable models can be used to aid segmentation and they demonstrate the use of a multiple surface deformation algorithm (ASP - anatomic segmentation using proximities) to achieve two aims. Firstly the algorithm explicitly prevents non-intersecting surfaces, resulting in topologies isomorphic to a sphere, and because it can deform multiple surfaces (to give in this case the GM/WM and GM/CSF surfaces) with both intersurface proximity and model-based constraints the two surfaces can “guide” each other into a plausible solution. Although there is not an absolute anatomic homology between subjects at each surface vertex, they have taken the averages of the vertices over 150 normal subjects to produce a mean map. They derive three measures of cortical thickness, distance between linked points, distance between closest points on the two surfaces and the distance along the surface normal from the outer surface to the inner.

Kabani et al. [15] perform a validation study using this method in MR brain images from 40 young (18-40) normals, using 10 cortical areas. The left hemisphere was tagged in 20 brains and the right in the other 20. They validate the automatic method against an expert manual cortical thickness estimation, and show an 80% success rate, in that there is no significant difference between the automatic and manual cortical thickness results in 16 of the 20 regions investigated. Both the manual and automatic results are greater than literature values from studies using post-mortem brains, but this is most likely due to dehydration of the fixed samples.

The most popular method of surface reconstruction, and the basis of many cortical thickness studies is that of Dale, Fischl and Sereno [6]. They identify the white-grey matter boundary, having performed a through plane intensity normalisation (by normalising the WM peaks in each image to a set value, assuming the tissue peaks are always in the same ratio) and having segmented the white matter voxels. They remove the brainstem and separate the hemispheres, and fit a surface tessellation to the white matter segmentation to represent the GM-WM. The tessellation is smoothed using a deformable surface algorithm, whereby connectivity is explicitly maintained and self-intersection is prevented. Fischl, Dale and Sereno [10] extend this work to inflate the cortical surface (so sulcal activity can be visualised), flatten an entire hemisphere, and to morph the hemisphere onto a surface, maintaining the topological structure of the surface but ensuring a closed co-ordinate system. Fischl and Dale [9] then go on to use the technique to estimate cortical thickness in 30 subjects (17 male) of age range 20-37 years. They find that the crowns of the gyri (on average 2.7 ± 0.3 mm) are thicker than the fundi of the sulci (2.2 ± 0.3 mm). They also showed good reproducibility, with a mean inter-session standard deviation on the thickness measurements being 0.25mm and a cross-scanner reproducibility of 0.31mm.

2 Methods

2.1 Overview

The approach we have taken to determine estimates of regional cortical thickness is more simplistic than the elaborate methods typically employed to produce surface rendered/expanded cortical thickness representations.

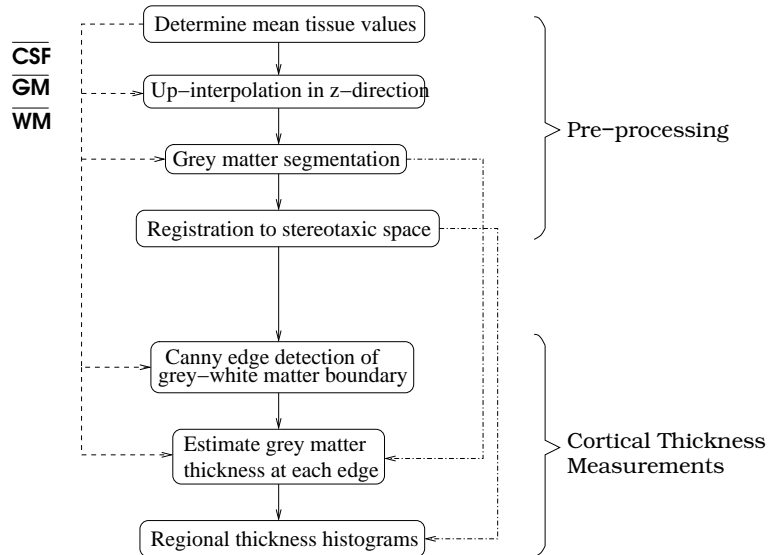


Figure 4: Overview of the cortical thickness method, including pre-processing stages

However, we will demonstrate that the methods we use are robust, result in excellent statistical error and give physiologically plausible results. Figure 4 gives a brief overview of the steps taken. The processing of the data can be divided into two stages, the actual cortical thickness estimation, and the pre-processing to convert the original volume of data into the required form. Initially, the average and standard deviation of the grey level values of the pure tissues (in this case only grey matter GM, white matter WM and cerebro-spinal fluid CSF are considered) in the volume are determined for use in future processing steps. Next, the data is up-interpolated in the z-direction (through plane) using a novel step-interpolation scheme making use of the estimated pure tissue values. The volume of data is then registered to a common stereotaxic space which defines the cortical regions used later in producing regional histograms of the cortical thickness. The last pre-processing stage is the segmentation of the grey matter, again using the mean and standard deviation of the pure tissue values. The segmented GM is used in the cortical thickness estimation.

The cortical thickness estimation itself proceeds by using a modified Canny edge detection routine to determine the grey-white matter boundary. The surface normal to the boundary in 3D at each voxel on the boundary is determined and a search is performed in this direction on the segmented grey matter map until an edge (of some description) is found. The type of edge (a CSF edge or white matter edge if there is no CSF between opposing banks of a sulcus) is determined and the cortical thickness length between the two edges is inserted into the appropriate regional histogram according to the registration into stereotaxic space determined earlier. The median of each regional histogram is calculated in order to give a robust estimation of the average cortical thickness in each region.

2.2 GM, WM, CSF Value Estimation

The MR acquisition used is a T2 inversion recovery sequence, where the CSF appears darkest in the images, grey matter appears grey and white matter appears white. An example histogram of grey-level values from a volume of the sequence is shown in blue in Fig. 6. Specific details of the scan parameters are given in Table 2 in section 4.

In order to both up-interpolate the data (using a novel step-interpolation scheme, briefly described below) and to segment the grey matter for the thickness estimation, it is necessary to determine the image intensity values of the pure tissues of the brain. A histogram is constructed containing those intensity values of voxels within a volume block of brain containing only those tissues of interest (ie, CSF, GM and WM). The approximate position of the volume block is illustrated in Figs. 5(a) and 5(b), although the specific location varies between subjects. The specific positioning of the volume in the frontal lobe was in order to obtain a sufficient number of voxels of the three pure tissue types, whilst attempting to avoid partial voluming. In elderly brains, the differentiation between all three tissue types becomes less distinct for several reasons. The boundary between grey and white matter appears much more diffuse than in younger brains (the gradient between the two become much more shallow) presumably due to demyelination of the white matter axons such that the white matter appears more like the non-myelinated grey matter. In addition, the presence of fluid filled lesions in white matter, and the greater partial voluming of grey matter and CSF (due to increased CSF) results in the appearance of grey matter in the white matter and of

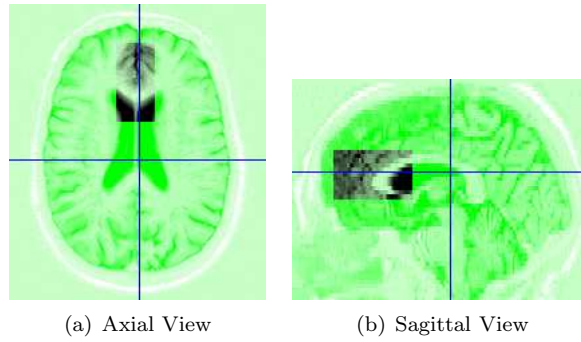


Figure 5: Sections illustrating approximate position of the cuboid from which the tissue histogram is constructed.

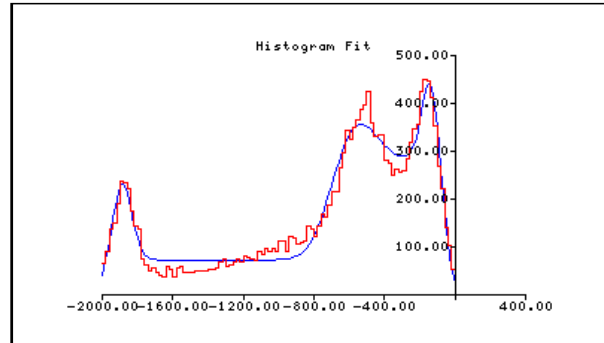


Figure 6: Histogram of voxel intensity values (red) and fit (blue) of the Gaussian mixture model to the histogram.

the actual grey matter appearing to have lower grey level values than is true. In order to obtain realistic estimates of the pure tissue values, CSF is taken from the ventricles, white matter from the corpus callosum, where there does not appear to be a lot of CSF/white matter partial voluming, and the cortex in front of the anterior cingulate, with relatively little CSF/grey matter partial voluming. In addition, use of a volume will to some extent accommodate variations in signal intensity due to through-plane magnetic field inhomogeneities. This may however mean that there will be greater variance on the position of the peaks than if only one slice had been used. The effect will be small as relatively few slices are used (up to 10). The issue of segmentation as a whole is an area of active research in the group.

We assume that the histogram represents both voxels containing pure tissues and partial volumes of tissues (in this case either a mixture of grey matter and CSF or a mixture of white matter and grey matter). We assume the pure tissue voxels have a Gaussian distribution and that the paired tissue partial volume distributions can be modelled for each tissue as the pure tissue Gaussian convolved with a triangular distribution. The model is fitted to the histogram using simplex, to obtain the mean values and standard deviation of the pure Gaussians. Figure 6 shows a typical histogram and fit to the data. A more sophisticated multi-image mixture model algorithm is described in [23].

2.3 Up Interpolation of the Data

In order to obtain clear (ie, with an acceptably low noise level for visual inspection) high resolution images for clinical use, in-plane voxels are typically high resolution (pixel size less than 1mm), whereas the through plane (z-direction) resolution is much coarser (of the order of 3mm), as the volume of images is not viewed (clinically) from the side. It is impractical due to the timing of the acquisition to improve the through-plane resolution for clinical use, therefore instead we employ an up-interpolation scheme for the z-direction in the post-processing. In the future with the increases in scanner power, it will be possible to produce such higher resolution isotropic voxels and obviate the need for up-interpolation.

The partial volume grey level analysis briefly described above will give an estimate of the fractional (volumetric) contributions of two tissues to a voxels grey level. This information provides a constraint on the potential boundaries (separating the two tissues) that could pass through the voxel. It is also possible to estimate the orientation of the boundary using spatial differentiation of the local grey level structure. These constraints are sufficient to determine a linear boundary model for the voxel, up to a directional ambiguity, which can be resolved by selecting

the orientation of the voxel according to the best agreement with the adjacent slices of data. Once the boundary and voxel orientation has been determined, it is possible to interpolate new grey-level values for any subdivision of the voxel. Here we have assumed the pure tissue grey levels as given above, zero inter-slice gap and have divided the voxels into two. See [20] for more details. The main limitation of the technique is in dealing with structures which are smaller than one voxel, where the algorithm will place the structure to one end of the voxel. However, this is not considered to be of concern when performing cortical thickness measurements, as the structures of interest here are much larger than the voxel sizes. A second problem is the need to assume a fixed set of mean grey level values for the tissues. Although this seems to be a good assumption, particularly in normal subjects, it is well known that field inhomogeneities, disease and even the spatially varying characteristics of tissues may have an influence on the measurement. One advantage that we have here is that we only need accurate interpolation and segmentation around the cortex. In addition we have implemented a quality control process in order to monitor the assumption that the grey-matter/white-matter boundary has equivalent value in all regions. We will explain below how it is possible to compute the systematic error on the boundary calculation and even adjust for this if necessary.

2.4 GM Segmentation of the Data

The grey matter in the up-interpolated images is then segmented using the means and standard deviations of the three tissue classes (CSF, GM, WM) determined earlier. The maps produced (of each slice) show the most likely volume estimate for GM in each voxel given the data, based on the defined probability density functions of the image intensity distribution. Voxels with values above 50% are therefore taken to be mainly grey matter and the boundary is defined at a 50% transition.

2.5 Registration to a Stereotaxic Space

We wish to be able to produce regional cortical thickness measurements and to compare these regional measurements between individuals and groups. This necessitates the registration of all the brains into a common stereotaxic space. For consistency we therefore register the up-interpolated brains to an average brain in Talairach space [29]. It was originally intended that different regions of the brain (ie, left-right, front-middle-back) could be scaled differently, as it was recognised that there are hemispheric differences in sizes etc, but here we have simply used a 15 parameter affine model describing pitch, roll, yaw, translation and scale. The regional specificity is therefore quite coarse, and as a precaution only sulcal/gyral structures and larger have been singled out for analysis, as smaller structures may not be consistently placed in all brains.

2.6 Cortical Thickness Estimation

The general approach to the cortical thickness estimation is to first determine the grey-white matter boundary, then to estimate the surface normal (into the grey matter) of the boundary, and then to search along the normal until a grey matter edge is found. It is then necessary to determine whether the edge is at a CSF-grey or grey-white boundary.

2.6.1 Iso Canny Routine

In order to determine the starting grey-white boundary, we have developed an extension to a standard edge detector. Defining the boundary at the average of the two pure tissue (50% transition) values, it is possible to produce a simple 'z-score' measure of each voxels grey level being consistent with this midpoint value and to construct from this value a likelihood. This image will therefore highlight the grey-white matter boundary (and additional edges where there is a sharp transition between CSF and background signal). This can then be used as a replacement for the conventionally sum squared image gradient (edge strength) in the Canny edge detector. The probability is modelled as a Gaussian with a standard deviation of 10 times that of the noise in the input image, and is computed as illustrated in Fig. 7 ¹.

The later stages of the Canny edge detector perform non-maximal suppression and hysteresis thresholding. This results in well localised connected edge strings that persist into low edge strength regions. Edge locations are computed within each slice of the data to sub-pixel accuracy from quadratic fits to the peak in the edge strength

¹This is monotonically related to the true likelihood but has better numerical properties for subsequent processing and sub-voxel peak location.

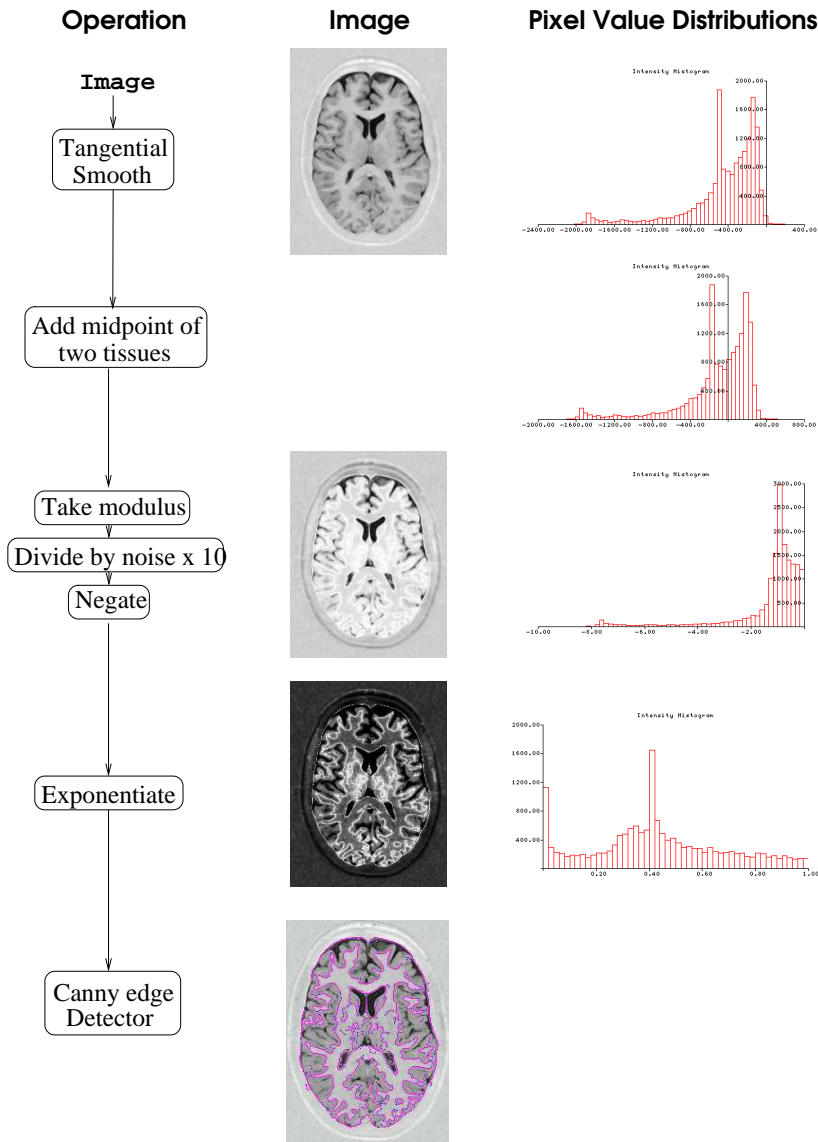


Figure 7: Diagram illustrating the Isocanny routine, showing the image and voxel value distributions at the given stages of the algorithm.

map. This is a necessary step if we are to attempt measurement of changes in cortical thickness of the order of fractions of a millimetre. Typical performance of these techniques corresponds in this case to approximately 0.1 mm reproducibility. The thresholds used here for hysteresis thresholding are: lower threshold = 0.0 (i.e., all edges which can be found are kept; upper threshold = 0.9 (i.e., only very strong edges are kept); minimum length = 5, in order to eliminate noisy edges in the white matter. The iso-canny routine is applied sequentially in 2D to each up-interpolated slice of the image volume. It could be argued that the boundary should be found within the 3D volume rather than the 2D slice, but it must be remembered that using a 2D method the correct boundary is still being found but is just being sampled at a set resolution through plane. The iso-canny routine will still place the boundary in the appropriate place on each slice.

2.6.2 Gaussian Smoothing and Orientation Determination

Having determined the boundary in 2D, it is necessary to determine the surface normal to the edge in 3D. In order to reduce the noise on the calculation, the entire volume of data undergoes a 3D Gaussian smoothing, using a kernel with a sigma of 1/2 voxel and range of 5 voxels. The surface normal is found by taking the local grey level gradient in 3D, using spatial differentiation of the 6 adjacent voxels to the voxel containing the edge of interest. Note that we have not centred the differentiation on the actual central edge position, although the error on the

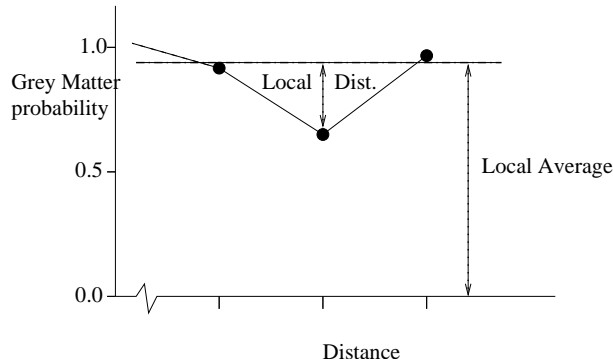


Figure 8: Diagram illustrating the calculation of the dip extent.

resulting orientation should be negligible.²

2.6.3 Search for an Edge/Partial Volume Dip

We can now take the grey matter tissue probability maps and the grey-white matter boundary and edge orientations and determine the distance from the boundary along the orientation direction until another grey matter edge is detected. First, we set a threshold beyond which the search stops. (See section 2.6.4 for optimal determination of parameters). We then search along the orientation direction in steps of 1mm on the grey matter segmentation image volume. An edge is defined as the point at which the grey matter probability drops below 50%, and is calculated as the linear interpolation between the values below 50% in the current voxel of interest and above 50% in the previous voxel. This edge may either be a grey-CSF or grey-white matter interface, if there is no CSF in the intervening sulcus (see later). In addition to these definite edges in the probability images, there may also be CSF or WM edges which are sufficiently narrow that they result in a partial volume voxel. Therefore, we also record the position of the greatest dip along the search orientation in the grey matter probability image.

A dip is defined as where the grey matter segmentation image value at the position of interest is lower than the average of the values at the adjacent positions either side. The value of the dip (local distance) (see Fig. 8) is this average minus the value of the voxel in which the dip occurs. The reason for taking this local distance, rather than the absolute value of the dip is to ensure that the dip is of sufficient depth relative to the surrounding voxels, and that it is not just noise on a low average part of the graph. Again, the type of dip (CSF, WM) can be determined from the original images. The actual position of the dip is calculated as follows. It is assumed that the position of the centre of the voxel in which the dip occurs and the positions of the centres of the adjacent voxels form part of a local quadratic curve, and so the position of the minimum of the dip can be obtained as a linear approximation to a quadratic (see Fig. 9(a)). The dip is due to a partial voluming of grey matter with either CSF or white matter. In reality, there will be a step edge between the tissues (see Fig. 9(b)), but on the scanned images the voxel intensity is merely an average of the signals from all the tissues in the voxel. The consequence of this is that the position of the step edge (the position of the grey matter boundary) is unknown. However the proportion of the voxel that is not grey matter can be estimated as local distance (from Fig. 8) which itself is a proportion. Assuming that there are only two tissues in the voxel and that the tissue partial voluming with the grey matter is a contiguous block and that this block is centred on the position of the minimum of the dip, then the actual boundary will be the *proportion/2* in front of the minimum dip position.

Once the position of the edge/dip boundary has been determined, the value of the voxel in the original grey level image is assessed to determine whether the edge is at a grey-CSF or grey-white matter boundary. If the value is less than the mean grey level value it is a CSF boundary, and the edge/dip position remains intact. Otherwise, it is a white matter boundary. It is therefore assumed that the opposing banks of a sulcus have been traversed and that these banks will on average be the same thickness. The sulcal thickness will therefore be on average half this length. Note that it is not necessary to enter this length in the histogram twice, as the other sulcus will be accounted for when calculating the thickness for the opposing GM-WM edge.

Now, for each orientation direction searched, there are potentially up to two possible positions for a grey matter boundary (ie, an actual edge or a dip). There may however be only one of either type, or none of either. When there is only one choice, we choose that point for the position of the edge. When there are two choices, however,

²At the top and bottom of the volume there will be no slice above and below (respectively) for this calculation, so twice the value between the voxel of interest and the voxel below/above is used. In practice this is an irrelevance, as the imaged volume extends above and below the Talairach regions.

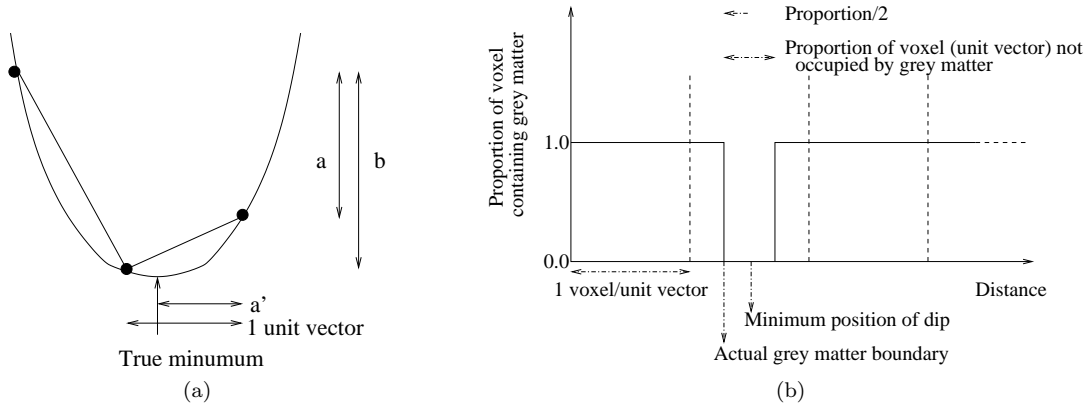


Figure 9: (a) Diagram illustrating the calculation of the position of the trough of a dip, assuming that the dip can be modelled as an approximation to a quadratic curve. The fraction (a') of the voxel at which the dip occurs is equivalent to the fraction a/b . (b) Diagram illustrating the calculation of the position of the trough of a dip, assuming that the dip can be modelled as an approximation to a quadratic curve. The fraction (a') of the voxel at which the dip occurs is equivalent to the fraction a/b .

we must apply some a priori assumptions about the most expected/wanted edge type. A CSF edge always takes precedence, and a white matter dip is always chosen last. We have decided to choose a CSF dip over a white matter edge because the CSF dip will be determined accurately, whereas if the white matter edge were chosen, we would have to use the less accurate assumption that the sulcal thickness is half the distance to the white matter edge.

2.6.4 Optimal Choice of Extent

When searching for an edge or a dip along the surface normal from the GM-WM starting boundary, it is necessary to confine the search to a fixed extent in order to prevent spuriously long, false, edges/dips being found. On the other hand, the search extent has to be sufficiently long to accommodate the true edges/dips. As part of this, it is necessary to allow twice the search distance that would be used to find a GM-CSF edge/dip, in order to find a GM-WM edge/dip, as on average the WM-GM to GM-WM distance will be twice that of a WM-GM to GM-CSF distance. In the following experiments therefore, the distances quoted are those allowed for finding a white matter edge/dip, and half this extent is used for finding CSF edges/dips.

Figures 10(a) - 14(a) and summarised in 15(a) and 15(b) show the effect of increasing search extent, from 5 to 40mm, on the values of the median thicknesses obtained over all the regions, given by lobe for one normal 40 year old male. In all cases, as the extent is increased, the median thickness increases. The reason being that we are working with a long tailed distribution and the greater the search extent, the more likely a longer edge will be found. Figures 10(b) - 14(b) show the effect of increasing the search extent on the numbers of grey/white matter edges found with opposing edges (ie, called "number of lines" in the plots). As can be seen with these plots, as the extent is increased, the number of lines found increases, but for all regions reach a plateau within a search extent of 10-20mm. However, in most regions, the median thickness continues to increase after 20mm. It is therefore impossible to define a unique thickness measurement, we can only standardise on a systematic approach and then investigate the reproducibility of the method (see below). We have therefore chosen 20mm as our cut-off point, as we believe that spurious edges are being detected beyond this point.

2.6.5 Optimal Choice of Threshold Dip

In determining the greatest dip, we could have specified that the dip must be greater than a certain (absolute) value for it to be considered a dip. Obviously, any dip beyond 50% is an edge anyway. In order to determine whether a threshold has an effect on the estimated median thickness, we performed the cortical thickness measurement with a range of minimum dip thresholds from 0 to 0.3. We produced three sets of graphs, given in Figs. 16(a) - 20(c). These show the effect of the dip threshold on median thickness when (a) both edges and dips are entered into the median thickness histograms, (b) edges only are entered and (c) dips only are entered in the histograms.

As can be seen, increasing the threshold of the minimum dip allowed has little effect on the median thickness in any of the regions, when either edges and dips and edges alone are considered. Increasing the threshold beyond

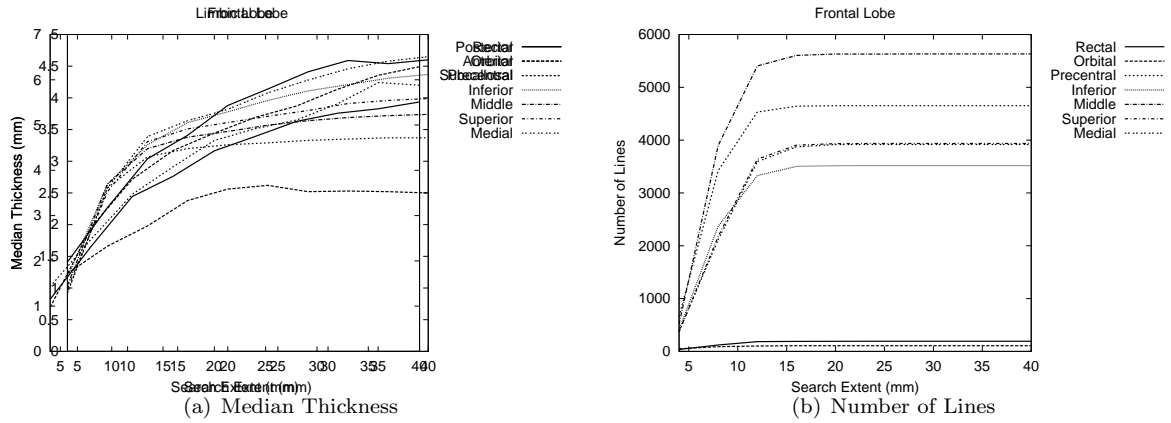


Figure 10: Effect of varying the search extent on the median thickness and number of lines found in Frontal Lobe structures

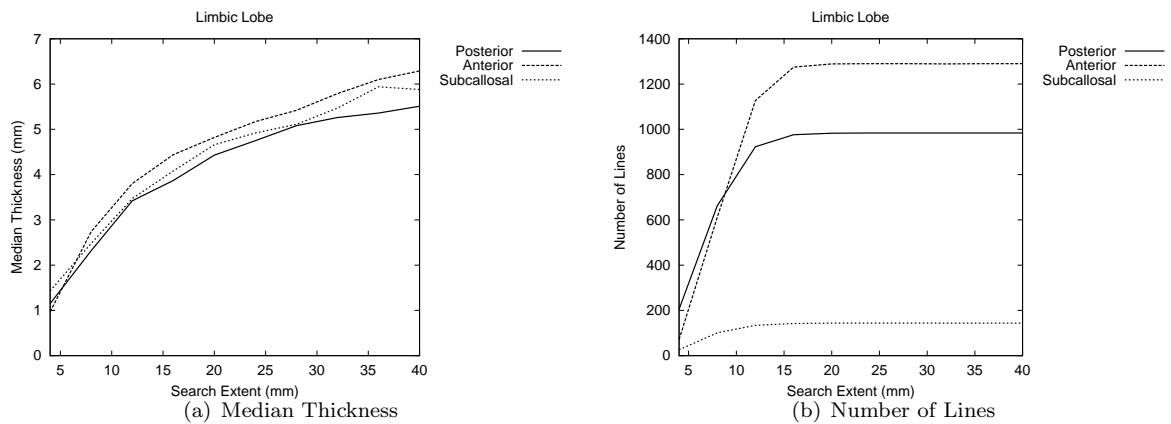


Figure 11: Effect of varying the search extent on the median thickness and number of lines found in Limbic Lobe structures

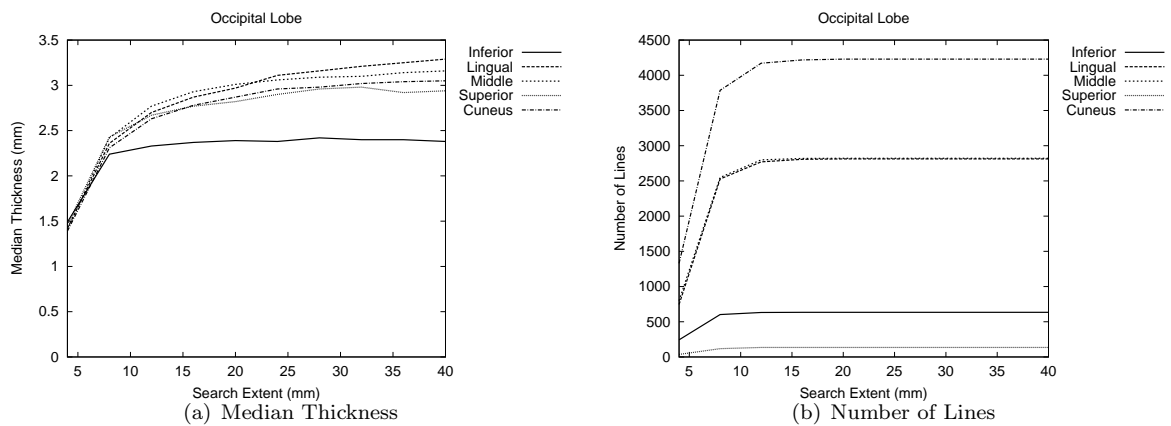


Figure 12: Effect of varying the search extent on the median thickness and number of lines found in Occipital Lobe structures

0.15 results in the median thickness of the regions when only the dips are considered becoming highly erratic, but this is due to the fact that as the threshold is increased, fewer dips are found, resulting in instability in the histograms, rather than indicating a problem with the technique. As there appears to be no change in the overall result by using a threshold, it was decided not to use one.

Note also that the median thickness of the dips alone is slightly greater than that of the edges alone. This could

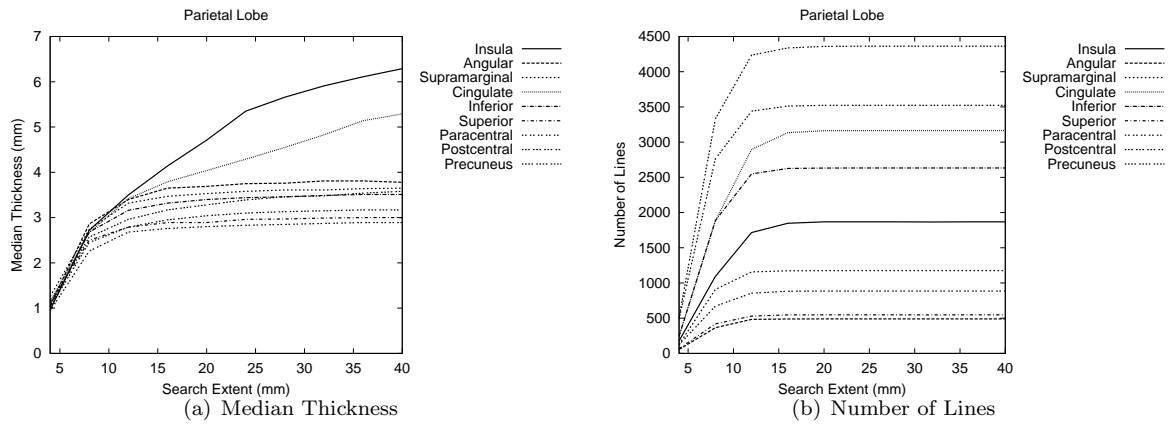


Figure 13: Effect of varying the search extent on the median thickness and number of lines found in Parietal Lobe structures

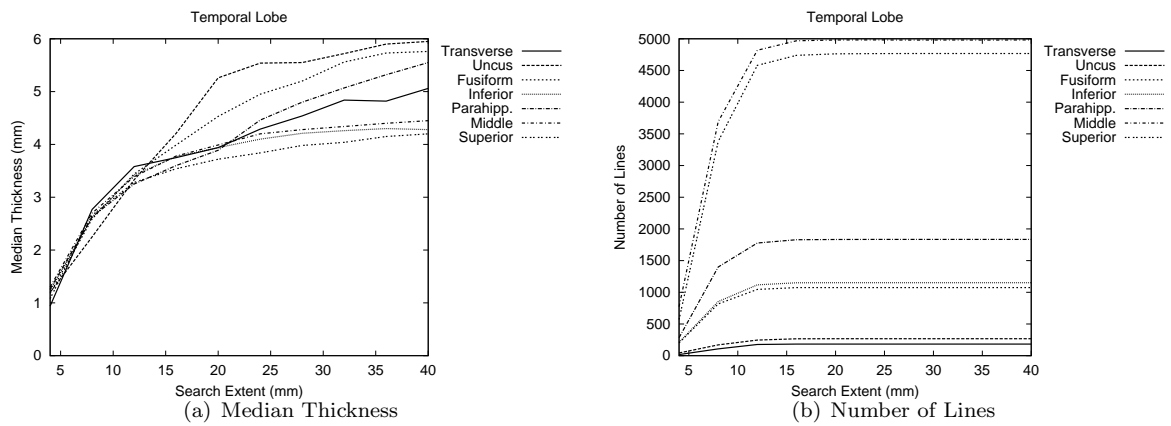


Figure 14: Effect of varying the search extent on the median thickness and number of lines found in Temporal Lobe structures

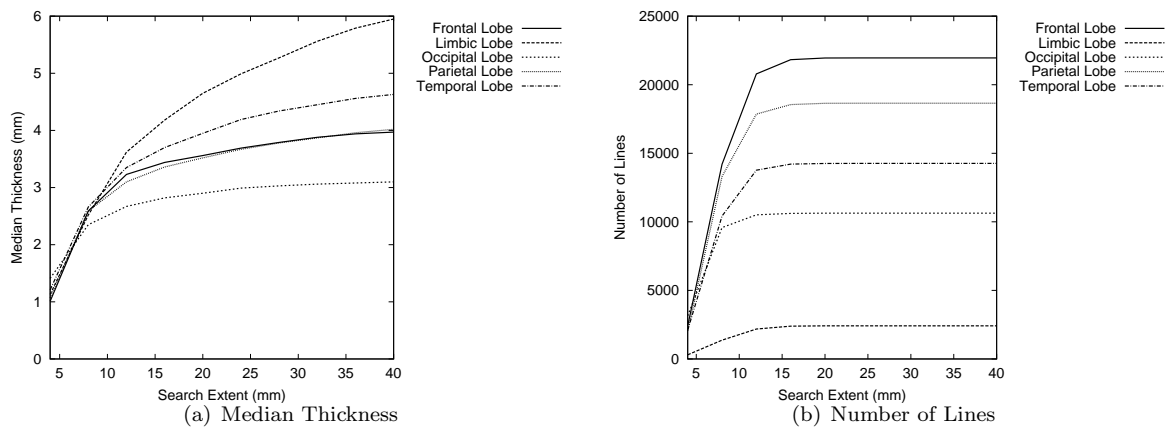


Figure 15: Effect of varying the search extent on the median thickness and number of lines found by cerebral lobe.

be accounted for by the possibility that these plots are slightly biased, in that the length of a thickness due to a dip is only histogrammed if there is no edge which has taken preference, in which case a shorter length due to a dip will not be recorded. However, it might also genuinely be the case that dips are systematically further from the starting GM-WM boundary.

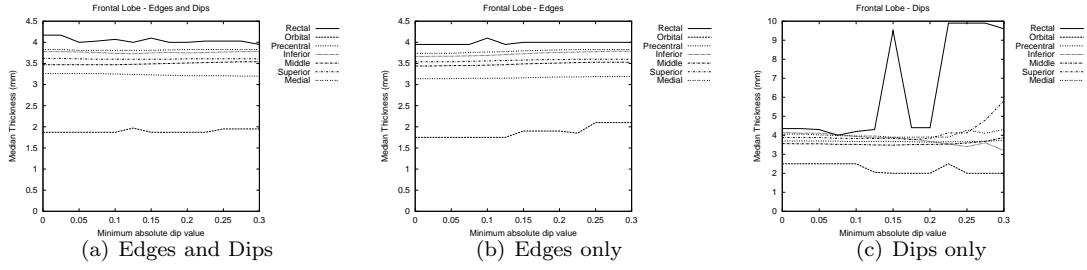


Figure 16: Effect of changing minimum absolute dip threshold on median thickness in the Frontal Lobe.

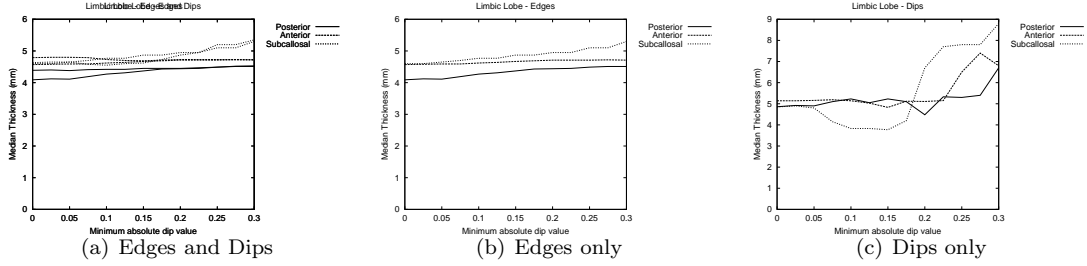


Figure 17: Effect of changing minimum absolute dip threshold on median thickness in the Limbic Lobe.

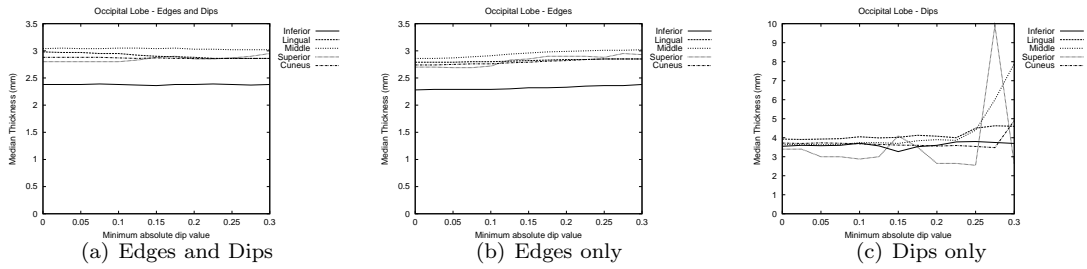


Figure 18: Effect of changing minimum absolute dip threshold on median thickness in the Occipital Lobe.

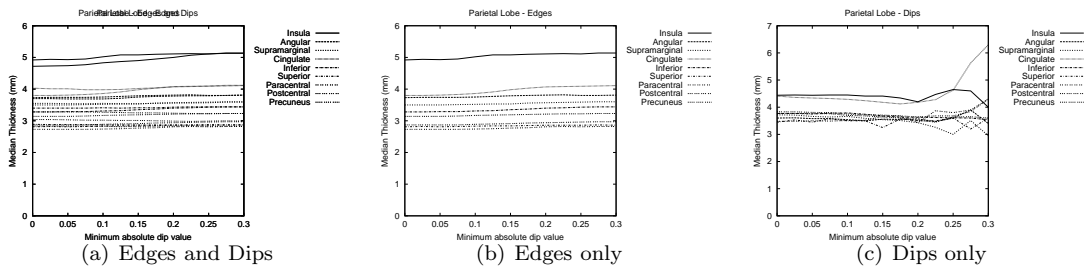


Figure 19: Effect of changing minimum absolute dip threshold on median thickness in the Parietal Lobe.

2.6.6 Cortical Thickness Refinement

All of the processing described so far has been based as closely as possible on individual voxel measurements. However, we expect the surface of the cortex and the boundary between tissues to be a smooth surface. We can exploit this in order to increase the stability of individual cortical thickness measurements. The use of the Gaussian smoothing to reduce the noise on the orientation calculation, the confinement of the search along the surface normal orientation to 20mm and the determination of partial volume dips in the grey matter are strategies which work well in preventing spuriously long thicknesses. However, we also apply a refining algorithm to remove any gross outliers which might remain.

The algorithm makes use of the spatial continuity provided by the edge string detection. It assesses each edge on the edge string returned by the iso-canny routine in comparison to its neighbours. It compares the value of

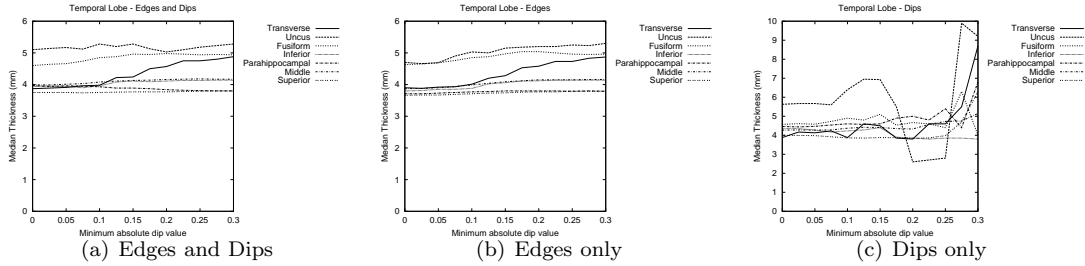


Figure 20: Effect of changing minimum absolute dip threshold on median thickness in the Temporal Lobe.

the cortical thickness when starting the search at that edge with the median of the thicknesses of a window of five values (comprising the edge of interest and the two edges on both sides, except at the beginning and ends of the edge strings, where the first/last five edges respectively are used). If the difference between the two exceeds a threshold (4mm was chosen) then the thickness starting at that edge is replaced by the median value. In the vast majority of cases this results in identification and removal of anomalous boundaries.³

2.7 Histogramming Regions, Median Thickness

The registration of the datasets to the Talairach atlas has already been performed and so it is possible to divide the cortex into sulcal/gyral regions. The thickness values of the edges found in these regions are entered into histograms (from 0-20mm, 50 bins) of the regions and the median and interquartile ranges (25% and 75%) of the histograms determined. The regions for which the cortical thickness was estimated are given in Table 1. Note that they are gyral regions, but they do cover the sulci too.

3 Error Quantification

As we have described above, the measurement of median thickness depends upon the specifics of the algorithm (such as the search extent). The measurement is therefore not an absolute quantity. However, this need not be a problem if our main goal is to compare measurements between anatomical regions and across subjects. In this case reproducibility is the main factor. In order to quantify the reproducibility error on the median thickness estimates, the same person could be scanned twice at two different sittings and the results compared.

Four young normals (see table 2 for the group and scan parameters) were scanned on two occasions, where the repeat scan was within 4-21 days of the first, and both scans took place at approximately the same time in the morning. Each of the eight datasets was assessed independently of the others, undergoing all of the pre-processing and thickness estimation steps. The median thickness was estimated in 31 cortical regions (averaging left and right regions) and the results are shown for all four subjects in Fig. 21. Also shown are the correlations between the first and second scans for each subject, constrained by the origin, as well as the line of equality.

The systematic error on the reproducibility measurements can be calculated as the extent of the deviation of the regression coefficients of the four subjects away from the line of equality⁴. The systematic error on the measurement in a given individual therefore amounts to 6.2%. The main cause of this systematic error is in the determination of the means and standard deviation of the CSF, GM and WM peaks (together with the Bayesian prior terms). Note that the effect of the segmentation on the magnitude of the systematic error is of crucial importance when planning longitudinal studies. Subjects must be scanned under identical conditions (physiological as well as in terms of instruments) and segmentation algorithms must be able to cope with drift in the intensity values obtained from the scanner. The statistical error on the any given regional measurement should be no more than 0.06mm, which itself should be dominated by the sample size.

The results from the above reproducibility study are independent of how voxels are allocated to anatomical regions. In order to get an *estimate* of the reproducibility of the technique for the comparison of equivalent structures we have compared the left-right symmetry of measurements within a brain. This is not only a simple process but

³In addition, the process will occasionally result in an edge with no thickness gaining a length (0.03% frequency), but may also result in the thicknesses of a few edges being erroneously replaced by zero (0.33% frequency). These frequencies are sufficiently low to prevent biasing the regional histograms.

⁴This is calculated as the RMS of the differences between the line of equality and the coefficients, scaled by $\sqrt{2}$ to account for the fact two measurements have been taken.

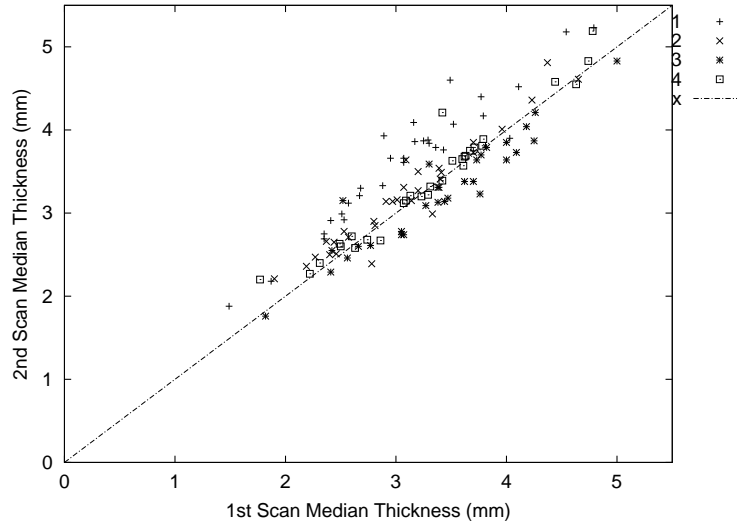


Figure 21: Scatterplot of median thickness values in four subjects for a repeat scan versus the first scan.

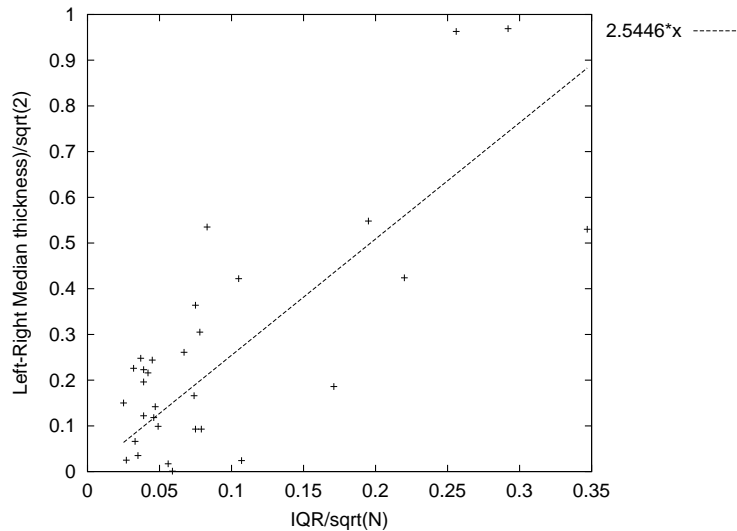


Figure 22: Scatterplot and regression line of repeatability error (absolute difference between left and right regional median thicknesses, over $\sqrt{2}$) against the inter-quartile range normalised to \sqrt{N} . The slope on the regression line is 2.54.

also has the added advantage that we can assess the effects of reproducibility completely independently of errors in grey-level calibration for segmentation and boundary detection identified above.

This variation in reproducibility between left and right can be used to calibrate an equivalent of the standard error for the median, using the interquartile range for the region normalised to the square root of the number of entries in the histogram for that region. This has been done for the same individual as earlier (male, 40). Figure 22 shows the (absolute) systematic error for this individual plotted against the normalised inter-quartile range (in this case the cortical thickness refinement has been applied). Also shown is the regression line on this data, which has a slope of 2.54. For a Gaussian distribution of thickness measurements this would take a value of 1.6. This is sufficiently close agreement that we can say that the majority of the errors observed are due to sample size and not other factors.

Figures 23(a) - 23(e) show for the same individual the median thickness in each cortical region as well as the error bars calculated from the scaled normalised inter-quartile range. Note that for the majority of regions, the systematic errors on the median thicknesses are small, of the order of a quarter to a third of a millimetre, meaning that we should be able to reliably detect differences in cortical thickness of at least 0.5mm in these regions. As the expected decrease in thickness in diseases such as Alzheimer's should be a couple of millimetres (having accounted for age) we should be able to detect changes due to the disease process. Regions with large errors (frontal rectal

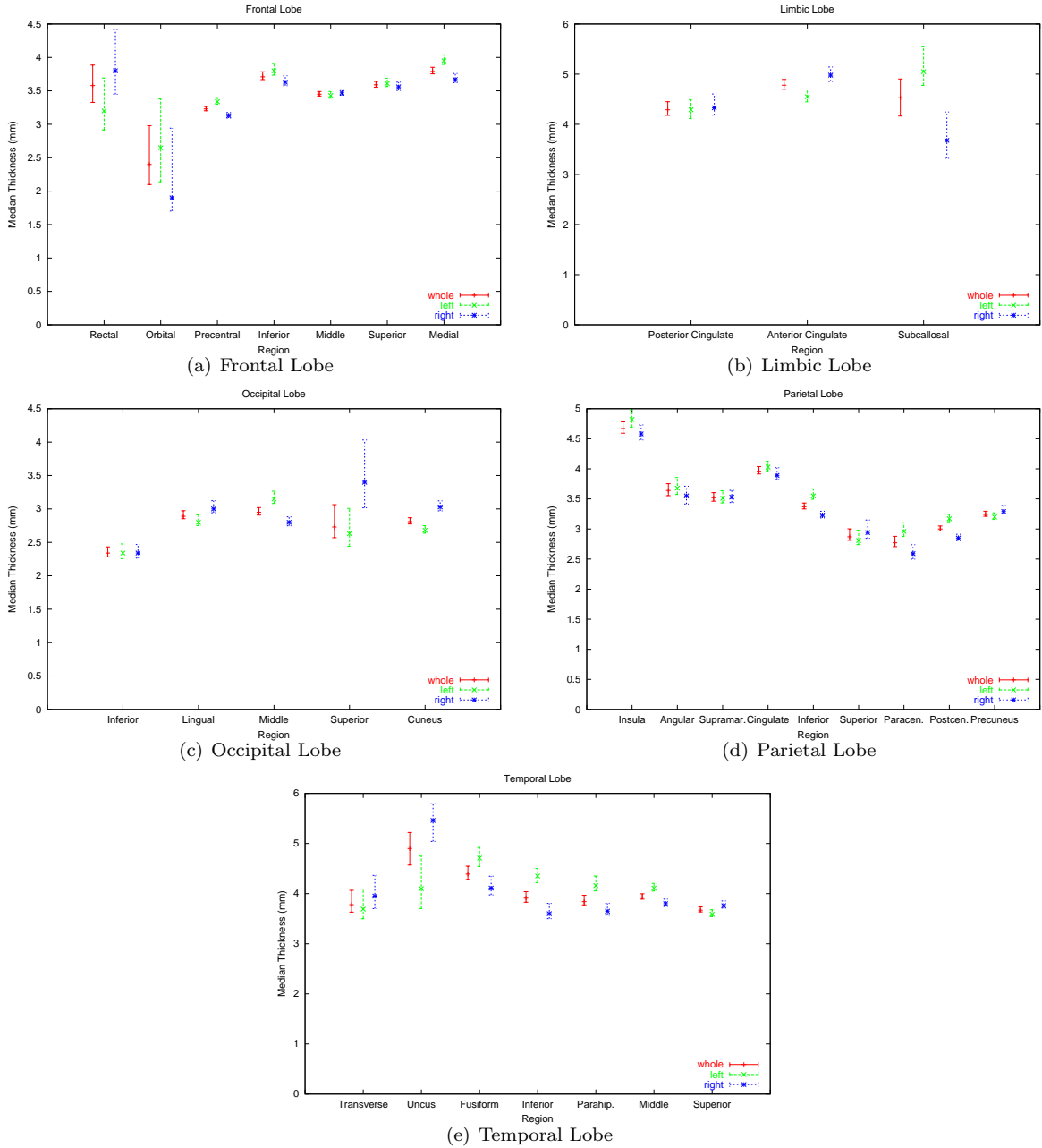


Figure 23: Median cortical thickness and error per region and cerebral lobe.

and orbital gyri, subcallosal limbic gyrus, superior occipital gyrus, transverse temporal gyrus and uncus) are due to the low numbers of entries in the histograms for those regions, due to the fact the structures themselves are quite small.

3.1 Iso Canny post-processing correction

If we consider the effects of errors in the tissue segmentation and boundary detection, the largest source of error is expected to be associated with the detection of the grey-matter white matter boundary location. Here small changes in estimated mean tissue values will have the largest potential effect on the partial volume interpretation (by at least a factor of 4). This will produce systematic errors on the cortical thickness when an incorrect grey-level value for the mid-point of the grey and white matter is used. However, we have developed a technique to monitor this effect and to correct for it *post hoc* if required. In order to calibrate such a correction, using the same normal (although the result may well be different using other normals), and using in each case the same set of grey level and probability images, we perturbed the grey/white midpoint by -60 to $+80$ and investigated the effect this had

Lobe	Region	Slope ($\times 10^{-3}$) (mm/grey level)
Frontal	Precentral Gyrus	-6.1369
	Inferior Frontal Gyrus	-5.4881
	Middle Frontal Gyrus	-6.2738
	Superior Frontal Gyrus	-7.3036
	Medial Frontal Gyrus	-5.8571
Limbic	Posterior Cingulate	-5.3095
	Anterior Cingulate	-7.8988
	Subcallosal Gyrus	3.1786
Occipital	Inferior Occipital Gyrus	-2.274
	Lingual Gyrus	-2.8155
	Middle Occipital Gyrus	-2.9583
	Superior Occipital Gyrus	-2.7560
	Cuneus	-2.4167
Parietal	Insula	-3.6488
	Angular Gyrus	-6.6071
	Supramarginal Gyrus	-5.1964
	Cingulate Gyrus	-4.5476
	Inferior Parietal Lobule	-5.9524
	Superior Parietal Lobule	-5.4405
	Paracentral Lobule	-5.1429
	Postcentral Gyrus	-5.2202
	Precuneus	-5.5536
Temporal	Transverse Temporal Gyrus	-9.732
	Uncus	-2.583
	Fusiform Gyrus	-4.9881
	Inferior Temporal Gyrus	-3.6071
	Parahippocampal Gyrus	-2.7083
	Middle Temporal Gyrus	-4.4048
	Superior Temporal Gyrus	-4.2679

Table 1: Table of the Talairach regions investigated, and the values of the slopes of median thickness against isocanny midpoint used.

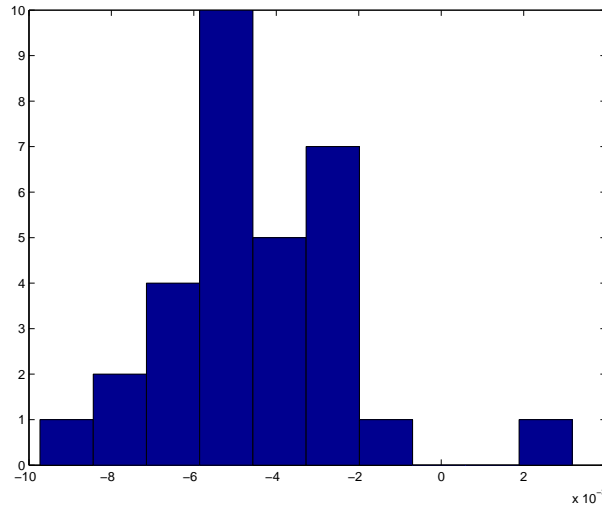


Figure 24: Histogram of slope values from Table 1. The median value is $-0.0051\text{mm/grey level}$.

on the median thickness in each region. Other than the regions with very few histogram entries, this resulted in a negative linear slope of thickness against midpoint value used; these slopes are given in Table 1.

As can be seen from a histogram of these slope values (Fig. 24), they appear to be drawn from a Gaussian distribution, with a median value of $-0.005\text{ mm/grey level}$. Having obtained the slopes from which we can estimate the correction, we have devised a method for determining the position along the slope that each region actually

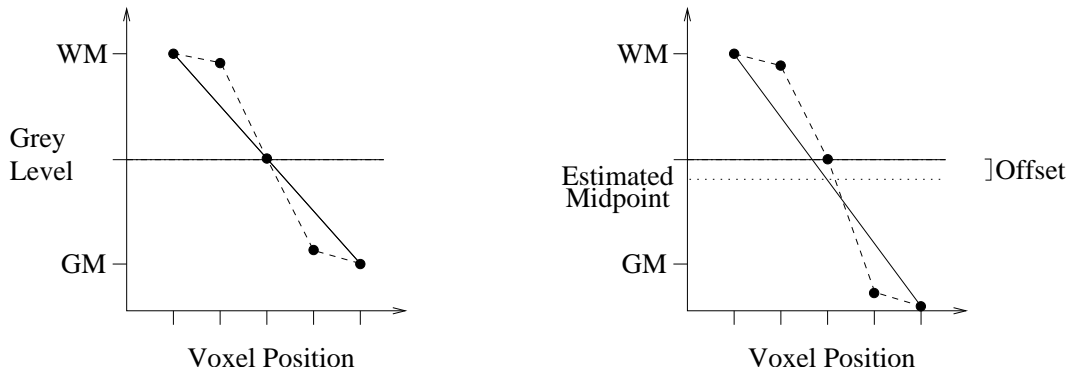


Figure 25: Diagram illustrating the midpoint offset calculation. The plots show voxel position on the x-axis and grey level on the y-axis. Marked are the pure tissue grey level values for grey and white matter and the midpoint between them is shown by the dash-dot line. The central voxel is at the midpoint (edge) and the two voxels on each side are in grey matter and white matter. The left-hand plot shows the ideal case, where the linear interpolation (dashed line) between the grey level values of the voxels two lengths either side corresponds to the midpoint value. The right-hand plot shows an example of the case where the value of the voxel in the grey matter is less than the pure tissue value, so that the interpolated grey level value at the central voxel is less than the midpoint (edge) value.

falls, as follows. The boundary between the grey and white matter should be a step edge, the width of which should not extend over many pixels. Therefore if you traverse the image volume from the determined edge position (using what you believe to be the correct grey-white matter midpoint) along the previously determined orientation, both in the positive and negative direction (in this case 2mm), you would expect to be within pure tissue (grey and white depending upon the direction). Assuming that the image gradient is the same either side of the edge boundary, one would expect the value of the central voxel (ie: at the edge) to be half way between the values at the voxels 2mm away either side (ie, linear interpolation). The difference between the two values is therefore the offset by which you have incorrectly placed the edge (ie, midpoint). This is illustrated in Fig. 25.

In order to get a value for the each region, the offset values for each edge in the region are simply histogrammed in the same way as the cortical thickness measurements. Here it is impractical to take the median of the region to obtain an average, as one can never be sure where the histograms start and end, instead we take the peak of the histogram as the estimate.

The results for all the regions in 8 young and 4 elderly subjects, demonstrate a negative bias (roughly ranging between 0 and -100). This is consistent with the grey matter grey levels being lower than expected; there is more of an emphasis on the lower grey matter values due to partial voluming between the grey matter and CSF ⁵. This suggests that we can use the locations of the peak histogram values in order to perform quality control across subjects, but we should be cautious about applying a correction between regions in the same brain. From this distribution we can say that the dominant systematic error on median thickness measurement due to segmentation errors will be of the order of $25 \text{ grey levels} \times 0.005 \text{ mm/grey level} = 0.125 \text{ mm}$. This is small in comparison to the observed sampling error but might be taken as an upper limit on the accuracy of the results on large regions.

4 Preliminary Results

4.1 Subjects and Scans

The normal subject group is composed of three groups from three different studies, scanned on different occasions on the same scanner (1.5 Tesla whole-body scanner Philips ACS PT 6000 NT, Philips Medical Systems, Best, The Netherlands). Scan parameters and demographics are given in Table 2. The acquisition used was a T2 inversion recovery sequence. All subjects were normal volunteers and gave informed consent at least 24 hours before scanning. Group 1 were part of a perfusion reproducibility study, Group 2 were normal volunteers in an ME study and Group 3 were part of a Wellcome study on amnesic cognitive impairment. The principle reason for using a range of age groups was to investigate the effect of age on cortical thickness.

⁵Equivalent quantities of partial voluming between grey-matter and white matter has less of an effect on grey level due to the smaller difference in mean grey level value

Parameter	Group 1	Group 2	Group 3
Number (males)	4 (4)	9 (5)	106 (43)
Mean Age (range)	40.2 (34-46)	35.4 (19-53)	74.4 (60-86)
TI	300	300	300
TE	18	18	18
TR	6850	6850	6850
Flip Angle	90	90	90
Echo train length	9	9	9
Slice Orientation	Axial	Coronal	Axial
In-plane Resolution	1.80 ²	1.80 ²	1.80 ²
Slice Thickness	3.0	4.0	3.0

Table 2: Subject demographics and scan parameters used.

4.2 Headsize vs age and sex

Headsize is usually correlated with age, with older persons having smaller heads. Females also typically have smaller heads than their male counterparts. There are usually sex effects on age, too, where there tend to be more older females in an elderly volunteer group and more younger males. We used several analysis of variance in order to determine whether there are relationships between head size, sex and age. Here headsize is calculated as the bounding box of the interior of the skull, thus the measure of headsize is independent of any atrophy the brain itself has undergone. It has been shown [4] that there is an excellent correlation between this term and the actual intra-cranial volume.

For this study, there is no significant difference in age with sex ($p=0.1287$, Adj R-sq = 0.01124, p is significant at 0.05). Head size is not significantly correlated with age ($p=0.143$, Adj R-sq = 0.01), and when sex is also included as a factor, becomes less significant once the effects of sex have been accounted for. Sex itself is highly significant in predicting head size ($p=1.41 \times 10^{-10}$, Adj R-sq for the whole model = 0.305), with females having smaller heads than males.

In all the following analyses of regional median thickness, sex is always included in the models in order to account for its effects. Head size and age are also included to test whether there are any correlations between the median thicknesses and these variables.

4.3 Median Thicknesses vs age, sex and head size in normals

We perform an ANOVA for each region, taking the form: Median Thickness = sex + head size + age, where sex is a factor and head size and age are covariates. Type 1 sum of squares are used, meaning that the order in which the variables are inserted in the model is important. The variance explained by any variable in the model is given the variance which has already been accounted for by any prior terms.

Table 3 gives the probabilities that sex, head size and age are significant in predicting group variance in median thickness. Note that in this case, the variance accounted for by head size is that above and beyond the variance accounted for by sex, and the variance accounted for by age is additional to that already accounted for by sex and head size. We have used a p -value significance of 0.05. Note that we have analysed 31 regions, and that we would not expect the cortical thickness in each region to be independent of all others. The reader may wish to impose a more stringent cut-off value than we have used.

As can be seen from the p -values listed in Table 3, there is a highly significant correlation in most regions of median thickness against age. Exceptions to this are the rectal gyrus of the frontal lobe, which has enormous statistical error (see Fig. 23(a)), and the inferior temporal gyrus. This region has comparable statistical accuracy to most other regions, so the fact that we do not see an age trend is suggestive of two reasons: First, there may genuinely be no age correlation; secondly, there may be something specific to the temporal lobe, such as its proximity to the basal ganglia, or the very deep sulci running through it, which renders our method less accurate in determining the cortical thickness in this regions.

Sex is significant in accounting for the variance seen in cortical thickness in ten regions (rectal and superior frontal gyri; anterior cingulate gyrus; angular and cingulate parietal gyri and superior parietal lobule; and the fusiform, parahippocampal, middle and superior gyri of the temporal lobe). In these ten cases, females have greater cortical thickness than males. Indeed, as can be seen from column two of 3, only the uncus shows a thinner median thickness in females compared to males.

Region	Sex		Head Size		Age	
	Constant	P-value	Slope ($\times 10^{-4}$)	P-value	Slope	P-value
Rectal Gyrus	0.213	0.020	-2.836	0.362	-0.008	0.07
Orbital Gyrus	0.187	0.200	-0.759	0.973	-0.018	1.42E-05
Precentral Gyrus	0.061	0.792	-1.043	0.921	-0.020	5.55E-16
Inferior Frontal Gyrus	0.036	0.653	-1.668	0.489	-0.014	1.74E-09
Middle Frontal Gyrus	0.080	0.244	-1.888	0.433	-0.017	1.52E-12
Superior Frontal Gyrus	0.099	0.046	-2.933	0.128	-0.016	5.10E-11
Medial Frontal Gyrus	0.076	0.171	-3.012	0.149	-0.020	7.77E-16
Posterior Cingulate	0.013	0.592	0.280	0.651	-0.013	3.12E-05
Anterior Cingulate	0.337	1.4E-5	-3.041	0.227	-0.013	4.62E-05
Subcallosal Gyrus	0.017	0.311	0.628	0.466	-0.031	6.36E-11
Inferior Occipital Gyrus	0.074	0.219	-2.091	0.387	-0.009	2.89E-03
Lingual Gyrus	0.090	0.138	-1.169	0.623	-0.012	5.72E-08
Middle Occipital Gyrus	0.097	0.107	-0.684	0.778	-0.006	4.49E-03
Superior Occipital Gyrus	0.043	0.772	-1.045	0.820	-0.012	1.66E-04
Cuneus	0.083	0.177	-1.044	0.676	-0.011	5.03E-08
Insula	0.172	0.187	-1.766	0.900	-0.034	2.00E-16
Angular Gyrus	0.145	0.039	-2.234	0.387	-0.018	1.17E-10
Supramarginal Gyrus	0.119	0.090	-3.164	0.208	-0.025	2.00E-16
Cingulate Gyrus	0.094	0.009	-2.843	0.061	-0.007	3.65E-04
Inferior Parietal Lobule	0.118	0.221	-1.449	0.773	-0.022	2.00E-16
Superior Parietal Lobule	0.136	0.036	-1.332	0.619	-0.014	3.15E-09
Paracentral Lobule	0.044	0.705	-2.681	0.371	-0.023	4.47E-14
Postcentral Gyrus	0.046	0.633	-1.932	0.423	-0.019	2.22E-16
Precuneus	0.060	0.182	-2.448	0.186	-0.013	4.07E-09
Transverse Temporal Gyrus	0.158	0.179	-0.422	0.868	-0.016	3.78E-07
Uncus	-0.073	0.809	-6.564	0.026	-0.021	3.45E-07
Fusiform Gyrus	0.151	0.006	-5.239	0.027	-0.020	1.28E-09
Inferior Temporal Gyrus	0.008	0.432	-2.287	0.244	-0.003	0.37
Parahippocampal Gyrus	0.181	0.001	-1.304	0.500	-0.007	1.11E-03
Middle Temporal Gyrus	0.126	0.040	-1.282	0.547	-0.007	4.96E-03
Superior Temporal Gyrus	0.156	0.028	-1.110	0.804	-0.017	1.82E-12

Table 3: Table giving the probabilities that sex, head size and age are significant in predicting variance in median thickness, as given by region. Values in bold are significant at the $p=0.05$ level. Also shown are the slopes for the regression of median thickness against head size and age in the model, and the constant offset in thickness according to sex. (NB, these are offsets for females, the constants for males are zero.)

Head size is significant in accounting for the variance in two regions (the uncus and fusiform gyrus of the temporal lobe); in both cases, the larger the head size, the smaller the thickness (although the correlation is very shallow). The general lack of a correlation of median thickness with head size supports the view that subjects with larger heads, hence larger brains, have increased white matter volume, assuming that the length of the cortical ribbon is roughly constant with head size.

4.4 Comparison of median thicknesses with other groups' results

Figure 26 shows a scatterplot of average median thickness for 20 regions (10 hemispherically symmetrical) from this current study and a semi-automated manual mark-up conducted by Kabani et al. [15]. They used 40 brain MRI volumes from a group aged 18-40. Half were used for assessment of the left-hand regions, and the other half for the right-hand regions. They applied a dual-surface deformation algorithm to determine the WM-GM and GM-CSF surfaces of the cortical ribbon, then a trained neuroanatomist marked out the correspondences between the two surfaces, from which the group averages were calculated. To compare our regional results with these we have taken the 13 young normals (aged 19-53) and taken the averages of the 10 regions for the left and right regions in the same brains.

Note that there is a strong correlation between Kabani's results and ours, lending some weight to the assertion that both methods characterise equivalent trends. The data lie fairly evenly about the line of equality, and there is no

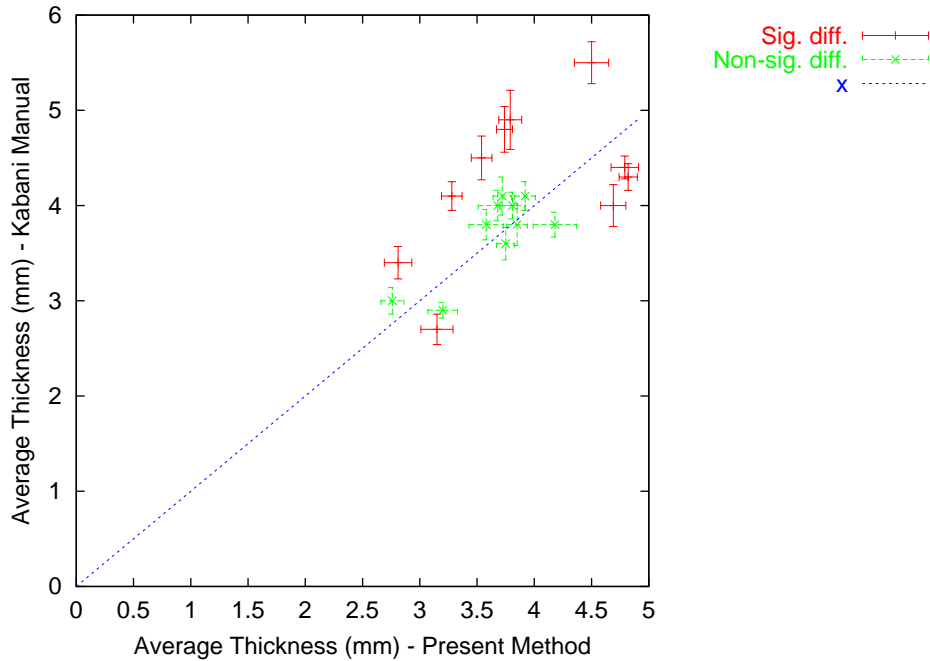


Figure 26: Scatterplot showing a comparison of results from the present method against a semi-automated manual mark-up. On the x-axis are the average median thicknesses (plus bars illustrating group standard error) from the 13 young normals in this study for the 20 regions; on the y-axis are the mean values from a semi-automated manual mark-up of 20 young normals (Kabani et al. [15]). In green are those regions where there is no significant difference (using a t-test with significance at $p=0.05$) between this study and Kabani’s, in red are those regions which are significantly different.

significant difference (at $p=0.05$, $df = 31$) between the results from both studies in half of the regions investigated. The remaining statistical differences are explainable. In particular, manual mark-up studies are often restricted to making measurements at specific orientations and in systematic anatomical locations (such as a limited numbers of points within sulci). Our measurement is defined such that all grey matter/ white matter boundary locations (as defined automatically by a geometric atlas) have an opportunity to contribute to the thickness distribution. This distribution is then characterised by a median, rather than an average for purposes of robustness. Such differences of definition, as described in the introductory sections of this document, are sufficient to prevent direct comparison and these differences need to be considered when comparing published results. In particular, although results of longitudinal change and analyses of correlated changes in group studies should be consistent between methods (i.e. the scientific conclusions such as those given in the tables above, should be equivalent), we would generally caution against direct comparison of absolute values derived from different definitions of thickness.

4.5 Advantages over VBM-style approaches and limitations of the technique

In conventional volumetric analysis of brain structure, results are often presented as significance scores in the form of a high resolution image. Here we only attempt to make measurements within relatively coarsely defined anatomical regions, consistent with the intrinsic accuracy of Talairach atlases. However, the idea that morphology based approaches are spatially accurate representations of the regions of majority tissue volume change would be an error. Spatially varying sensitivity to the movement of tissue boundaries and topological variation, together with the smoothing processes generally applied, should caution against any naive interpretation of such results in this way. Taking this problem into account even VBM results should only be interpreted on the basis of a much more coarsely defined anatomical map. In our method we have simply made this process explicit from the beginning. The advantage of this is that we can largely ignore the variability in brain structure brought about by cortical folding, and concentrate more directly on the quantity of tissue. Clearly, there is one important aspect of structure which is missed by measuring cortical thickness, and that is volume. We have addressed this issue in a previous study using techniques based upon monitoring the relative increases of cranial fluid in coarse regions of the brain [30]. In some respects this new technique, and that one are intended to be complementary.

In contrast to VBM, the technique described here measures a genuine dimensioned (length) change in cortical structure. This was done in an attempt to avoid the confound of anatomical variation in conventional approaches.

In addition, standard VBM requires in the order of hundreds of subjects in order to produce results. With our approach the intention is to provide measurements with sufficient accuracy that cortical thickness distributions can be interpreted for decision support on individuals.

Although there are difficulties in summarising a complex distribution entirely in terms of a median we should at least be able to correctly attribute quantifiable changes in the selected regions of the brain. In addition, VBM techniques are often interpreted without restriction, across the entire of the data set, with the belief that estimates of significant change mean the same thing in each region. Here we make estimates of error based upon the local sample quantity. In addition, we accept that we may wish to restrict use of data from a limited number of volumes, subject to the reliability of measurements and their associated error estimates as a function of anatomical region. This requires further study.

5 Conclusions

This paper presents a group of methods which aim to deliver reproducible estimates of cortical thickness in a wide variety of brain structures. We have investigated the main sources of error associated with this analysis during the course of algorithm design. As a consequence we are able to provide measurements accompanied with the expected error. The main systematic errors on the technique have also been investigated and found to be small in comparison to the error produced by statistical sampling. Even so, more work is still needed in order to minimise systematic factors when applying these techniques to longitudinal studies.

6 Acknowledgements

Thanks to Prof. Jackson and Prof. Rabbitt and Dr. Varma for use of the data. Thanks also to Lewis Griffin for suggesting how to improve the dip positioning calculation.

References

- [1] J Annesse, A Pitiot, I D Dinov, and A W Toga. A myelo-architectonic method for the structural classification of cortical areas. *NeuroImage*, 21:15–26, 2004.
- [2] M Ballmaier, E R Sowell, P M Thompson, A Kumar, K L Narr, H Lavretsky, S E Welcome, H DeLuca, and A W Toga. Mapping brain size and cortical grey matter changes in elderly depression. *Biol. Psychiatry*, 55:382–389, 2004.
- [3] F L Bookstein. “Voxel-based morphometry” should not be used with imperfectly registered images. *NeuroImage*, 14:1454–1462, 2001.
- [4] P A Bromiley, N A Thacker, and A Jackson. Trends in brain volume change with normal ageing. Tina Memo 2004-002, www.tina-vision.net/memos/2004-002.pdf, 2004.
- [5] 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*, 20(3):1425–1437, 2003.
- [6] A M Dale, B Fischl, and M I Sereno. Cortical surface-based analysis I: segmentation and surface reconstruction. *NeuroImage*, 9:174–194, 1999.
- [7] C Davatzikos. Why voxel-based morphometric analysis should be used with great caution when characterising group differences. *NeuroImage*, 23(1):17–20, Sept 2004.
- [8] K L Double, G M Halliday, J J Krill, J A Harasty, K Cullen, W S Brooks, H Creasey, and G A Broe. Topography of brain atrophy during normal aging and Alzheimer’s disease. *Neurobiol. Aging*, 17(4):513–521, 1996.
- [9] B Fischl and A M Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Nat. Acad. Sci. USA*, 97(20):11050–11055, 2000.
- [10] B Fischl, M I Sereno, and A M Dale. Cortical surface-based analysis II: Inflation, flattening and a surface-based coordinate system. *NeuroImage*, 9:195–207, 1999.

- [11] C D Good, I S Johnrude, J Ashburner, R N A Henson, K J Friston, and R S J Frackowiak. A voxel based morphometric study of ageing in 465 normal adult brains. *NeuroImage*, 14:21–36, 2001.
- [12] A Gueziec and R Hummel. Exploiting triangulated surface extraction using tetrahedral decomposition. *IEEE T. Vis. Comput. Graph.*, 1:328–334, 1995.
- [13] S E Jones, B R Buchbinder, and I Aharon. Three-dimensional mapping of cortical thickness using Laplace’s equation. *Hum. Br. Map.*, 11:12–32, 2000.
- [14] M Joshi, J Cui, K Doolittle, S Joshi, D Van Essen, L Wang, and M I Miller. Brain segmentation and the generation of cortical surfaces. *NeuroImage*, 9:461–476, 1999.
- [15] N Kabani, G Le Goualher, D MacDonald, and A C Evans. Measurement of cortical thickness using an automated 3-D algorithm: A validation study. *NeuroImage*, 13:375–380, 2001.
- [16] G R Kuperberg, M R Broome, P K McGuire, A S David, M Eddy, F Ozawa, D Goff, C West, S C R Williams, A J W van der Kouwe, D H Salat, A M Dale, and B Fischl. Regionally localized thinning of the cerebral cortex in schizophrenia. *Arch. Gen. Psychiatry*, 60:878–888, Sept 2003.
- [17] J G Levitt, R E Blanton, S Smalley, P M Thompson, D Guthrie, J T McCracken, T Sadoun, L Heinichen, and A W Toga. Cortical sulcal maps in autism. *Cereb. Cortex*, 13:728–735, 2003.
- [18] D MacDonald, N Kabani, D Avis, and A C Evans. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*, 12:340–356, 2000.
- [19] V A Magnotta, N C Andreasen, S K Schultz, G Harris, T Cizaldo, D Heckel, P Nopoulos, and M Flaum. Quantitative in vivo measurement of gyrification in the human brain: changes associated with aging. *Cereb. Cortex*, 9:151–160, 1999.
- [20] S McKie and N A Thacker. Step interpolation of MR images with inter-slice gap correction. Tina Memo 2003-010, www.tina-vision.net/memos/2003-010.pdf, 2003.
- [21] M I Miller, A B Massie, T Ratnanather, K N Botteron, and J G Csernansky. Bayesian construction of geometrically based cortical thickness metrics. *NeuroImage*, 12:676–687, 2000.
- [22] K L Narr, R M Bilder, S Kim, P M Thompson, P Szeszko, D Robinson, E Luders, and A W Toga. Abnormal gyral complexity in first-episode schizophrenia. *Biol. Psychiatry*, 55:859–867, 2004.
- [23] M Pokrić, N Thacker, M L J Scott, and A Jackson. Multi-dimensional medical image segmentation with partial voluming. In *Proc. MIUA*, pages 77–81, 2001.
- [24] J T Ratnanather, K N Botteron, T Nishino, A B Massie, R M Lal, S G Patel, S Peddi, R D Todd, and M I Miller. Validating cortical surface analysis of medial prefrontal cortex. *NeuroImage*, 14:1058–1069, 2001.
- [25] H D Rosas, A K Liu, S Hersch, M Glessner, R J Ferrante, D H Salat, A van der Kouwe, B G Jenkins, A M Dale, and B Fischl. Regional and progressive thinning of the cortical ribbon in Huntington’s disease. *Neurol.*, 58:695–701, 2002.
- [26] M Sailer, B Fischl, D Salat, C Tempelmann, M A Schonfeld, E Busa, N Bodammer, H-J Heinze, and A Dale. Focal thinning of the cerebral cortex in multiple sclerosis. *Brain*, 126:1734–1744, 2003.
- [27] L D Selemon, G Rajkowska, and P S Goldman-Rakic. Evidence for progression in frontal cortical pathology in late-stage Huntington’s disease. *J. Comp. Neurol.*, 468:190–204, 2004.
- [28] E R Sowell, B S Peterson, P M Thompson, S E Welcome, A L Henkenius, and A W Toga. Mapping cortical change across the human life span. *Nature Neuroscience*, 6(3):309–315, Mar 2003.
- [29] J Talairach and P Tournoux. *Co-planar stereotaxic atlas of the human brain. 3-dimensional proportional system: An approach to cerebral imaging*. Thieme medical publishers, New York, 1988.
- [30] N A Thacker, A R Varma, D Bathgate, S Stivaros, J S Snowden, D Neary, and A Jackson. Dementing disorders: Volumetric measurement of cerebrospinal fluid to distinguish normal from pathologic findings - feasibility study. *Radiology*, 224(1):278–285, Jul 2002.