NeuroImage 59 (2012) 3178–3186

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Automatic hippocampal segmentation in temporal lobe epilepsy: Impact of developmental abnormalities

Hosung Kim [a], Marie Chupin [b], Olivier Colliot [b], Boris C. Bernhardt [a], Neda Bernasconi [a], Andrea Bernasconi [a,*]

[a] Neuroimaging of epilepsy laboratory, McConnell Brain Imaging Center, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada
[b] Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, UMR-S975, INSERM, U975, UPMC Université Paris 6, CNRS, UMR 7225, ICM, Paris F-75013, France

ARTICLE INFO

Article history:
Received 29 June 2011
Revised 8 November 2011
Accepted 14 November 2011
Available online 25 November 2011

Keywords:
Hippocampus
Automatic segmentation
Developmental abnormalities
Epilepsy

ABSTRACT

In drug-resistant temporal lobe epilepsy (TLE), detecting hippocampal atrophy on MRI is important as it allows defining the surgical target. The performance of automatic segmentation in TLE has so far been considered unsatisfactory. In addition to atrophy, about 40% of patients present with developmental abnormalities (referred to as malrotation) characterized by atypical morphologies of the hippocampus and collateral sulcus. Our purpose was to evaluate the impact of malrotation and atrophy on the performance of three state-of-the-art automated algorithms. We segmented the hippocampus in 66 patients and 35 sex- and age-matched healthy subjects using a region-growing algorithm constrained by anatomical priors (SACHA), a freely available atlas-based software (FreeSurfer) and a multi-atlas approach (ANIMAL-multi). To quantify malrotation, we generated 3D models from manual hippocampal labels and automatically extracted collateral sulci. The accuracy of automated techniques was evaluated relative to manual labeling using the Dice similarity index and surface-based shape mapping, for which we computed vertex-wise displacement vectors between automated and manual segmentations. We then correlated segmentation accuracy with malrotation features and atrophy. ANIMAL-multi demonstrated similar accuracy in patients and healthy controls ($p>0.1$), whereas SACHA and FreeSurfer were less accurate in patients ($p<0.05$). Surface-based analysis of contour accuracy revealed that SACHA over-estimated the lateral border of malrotated hippocampi ($r=0.61$; $p<0.0001$), but performed well in the presence of atrophy ($|r|<0.34$; $p>0.2$). Conversely, FreeSurfer and ANIMAL-multi were affected by both malrotation (FreeSurfer: $r=0.57$; $p=0.02$, ANIMAL-multi: $r=0.50$; $p=0.05$) and atrophy (FreeSurfer: $r=0.78$, $p<0.0001$, ANIMAL-multi: $r=0.61$; $p<0.0001$). Compared to manual volumetry, automated procedures underestimated the magnitude of atrophy (Cohen's *d*: manual: 1.68; ANIMAL-multi: 1.11; SACHA: 1.10; FreeSurfer: 0.90, $p<0.0001$). In addition, they tended to lateralize the seizure focus less accurately in the presence of malrotation (manual: 64%; ANIMAL-multi: 55%, $p=0.4$; SACHA: 50%, $p=0.1$; FreeSurfer: 41%, $p=0.05$). Hippocampal developmental anomalies and atrophy had a negative impact on the segmentation performance of three state-of-the-art automated methods. These shape variants should be taken into account when designing segmentation algorithms.

Introduction

Temporal lobe epilepsy (TLE) is the most frequent form of drug-resistant epilepsy. The majority of patients display hippocampal sclerosis, a process characterized by various degrees of neuronal loss and astrocytic gliosis (Babb and Brown, 1987). On MRI, hippocampal sclerosis generally appears as atrophy and signal changes (Jackson et al., 1990). Detecting hippocampal sclerosis is clinically relevant, as it allows the definition of the surgical target and is associated with favorable outcome in more than 70% of patients (Schramm and Clusmann, 2008).

Manual delineation of the hippocampus is the current gold standard, as it is accurate, reproducible and able to detect atrophy with high sensitivity (Bernasconi et al., 2003; Jackson et al., 1993; Kuzniecky et al., 1999). On the other hand, time requirement, rater-bias, and increased demand to study large cohorts of healthy and diseased populations have motivated the development of automated segmentation procedures. Most methods employ deformable (Kelemen et al., 1999; Yang and Duncan, 2004), appearance- (Avants et al., 2010; Duchesne et al., 2002) or atlas-based approaches (Collins et al., 1995; Fischl et al., 2002; Khan et al., 2008). Modeling spatial relationships and texture has improved accuracy (Avants et al., 2010; Chupin et al., 2009b). Alternatively, optimizing atlas-based techniques with graph-cuts may impact favorably results in patients with Alzheimer's disease and hippocampal atrophy (van der Lijn et al., 2008; Wolz et al., 2010). Recently developed multi-template (or

* Corresponding author at: Montreal Neurological Institute, 3801 University Street, Montreal, Quebec, Canada H3A 2B4. Fax: +1 514 398 2975.
E-mail address: andrea@bic.mni.mcgill.ca (A. Bernasconi).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.11.040

template library and label fusion) techniques account for structural variability by selecting from a database a subset that best describe anatomical characteristics of the target structure (Aljabar et al., 2009; Avants et al., 2010; Collins and Pruessner, 2010; Lötjönen et al., 2010). Although algorithms, study groups, imaging type and performance metrics vary across studies, results in healthy controls have generally been satisfactory, with kappa agreement indices ranging from 0.75 to 0.89 (Aljabar et al., 2009; Avants et al., 2010; Chupin et al., 2009b; Collins and Pruessner, 2010; Coupé et al., 2011; Heckemann et al., 2006; Khan et al., 2008; Lötjönen et al., 2011; Morey et al., 2009; Pohl et al., 2007; van der Lijn et al., 2008).

In TLE, agreements between manual labeling and automated segmentation have been low compared to healthy controls, with kappa indices ranging from 0.63 to 0.77 (Akhondi-Asl et al., 2011; Avants et al., 2010; Chupin et al., 2009b; Hammers et al., 2007; Heckemann et al., 2010; Pardoe et al., 2009). The reduced accuracy likely stems from factors other than atrophy, as previous approaches achieved a performance similar to controls in patients with Alzheimer's disease (Barnes et al., 2008; Chupin et al., 2009a; Leung et al., 2010). Indeed, studies in early and declared forms of this condition have reported hippocampal volume reductions ranging from 23 to 34% (Frisoni et al., 1999; Jack et al., 1992; Lehericy et al., 1994; Xu et al., 2000). Notably, the degree of atrophy is generally larger in declared Alzheimer's disease than in TLE, in which the effect size is in the order of 20%. In addition to atrophy, about 40% of TLE patients show atypical shape and positioning of the hippocampus (Bernasconi et al., 2005; Voets et al., 2011). These features, commonly referred to as malrotation, are considered markers of neurodevelopmental anomalies (Baulac et al., 1998; Voets et al., 2011) and may contribute to the pathogenesis of this condition (Blumcke et al., 2002; Sloviter et al., 2004). They are mainly characterized by a rounder appearance and atypical orientation of the hippocampus, and an abnormally deep and verticalized collateral sulcus (Baulac et al., 1998; Bernasconi et al., 2005). Thus, malrotation features not only alter hippocampal morphology, but also modify its spatial relationship with surrounding structures.

Our purpose was to evaluate the impact of malrotation, quantified through 3D descriptive models (Kim et al., 2006; Voets et al., 2011), on the performance of three fully automated hippocampal segmentation algorithms. We chose two algorithms previously used in TLE: SACHA, a region growing approach that utilizes rule-based detection of anatomical landmarks (Chupin et al., 2009b) and FreeSurfer (Fischl et al., 2002), a freely available algorithm based on the nonlinear warp of a target image to a probabilistic atlas (Akhondi-Asl et al., 2011; Pardoe et al., 2009). In addition, we evaluated a multi-atlas approach based on ANIMAL registration technique (Collins and Pruessner, 2010) that is among the most performant algorithms in healthy controls, but has not been applied to TLE. Performance was assessed relative to manual labeling using overlap indices and surface-based shape mapping. The ability of automated methods to lateralize the seizure focus was evaluated using linear discriminant analysis.

Methods

Subjects

We studied 66 consecutive patients (36 males; 16–44 years, mean age 36 ± 10 years) referred to our hospital for the investigation of drug-resistant TLE. The lateralization of the seizure focus was based on a standard clinical evaluation including detailed history of seizure semiology, recording of seizures by means of video-EEG monitoring and radiological assessment of hippocampal sclerosis through visual estimation of atrophy and increased T2 signal. Based on the convergence of these exams, patients were classified into left TLE (LTLE; $n = 35$) and right TLE (RTLE; $n = 31$). None of the patients had a mass lesion (tumor or vascular malformation), developmental malformation of the neocortex (cortical dysplasia, heterotopia or polymicrogyria), or traumatic brain injury. Forty-eight patients underwent surgery. Mean follow-up time was 3.1 ± 3.4 years. We determined surgical outcome according to Engel's modified classification (Engel et al., 1993). Thirty-four (71%) patients had Class I outcome, 5 (10%) Class II, 5 (10%) Class III and 4 (8%) Class IV. Following qualitative histopathological analysis (Meencke and Veith, 1991), hippocampal sclerosis was detected in 41/48 (85%) of patients in whom a hippocampal specimen was available. In the remaining seven, specimens were either incomplete or unsuitable for histopathology.

The control group consisted of 35 age- and sex-matched healthy individuals (19 males; 20–56 years, mean age 32 ± 12 years). The Ethics Committee of the Montreal Neurological Institute and Hospital approved the study, and written informed consent was obtained from all participants.

MRI acquisition

MR images were acquired on a 1.5 T Gyroscan (Philips Medical Systems, Eindhoven, The Netherlands) using a 3D T1-fast field echo sequence (TR = 18 ms; TE = 10 ms; NEX = 1; flip angle = 30°; matrix size = 256×256; FOV = 256 mm; slice thickness = 1 mm), providing an isotropic voxel volume of 1 mm^3. Prior to processing, images underwent automated correction for intensity non-uniformity and intensity standardization (Sled et al., 1998).

The hippocampus was segmented manually according to our previously published protocol (Bernasconi et al., 2003). Prior to segmentation, MR images were registered into the MNI ICBM-152 nonlinear template (Fonov et al., 2011) using 9 parameter linear transformation (Collins et al., 1994).

Automatic hippocampal segmentation

1. SACHA. This algorithm simultaneously segments the hippocampus and the amygdala based on a competitive region deformation constrained by automatically detected anatomical landmarks (Chupin et al., 2007) and probabilistic priors (Chupin et al., 2009b). During the deformation, voxels along the boundaries of the object are iteratively reclassified guided by anatomical priors. In our study, we modified the initialization step. Instead of registering probabilistic atlases of the hippocampus and amygdala (constructed from 16 healthy subjects) to a given target image in native space using the original nonlinear discrete cosine basis registration (Ashburner and Friston, 1999), we employed ANIMAL that combines linear transformation and non-linear warping based on a piece-wise linear coarse-to-fine deformation (Collins et al., 1995). The choice of ANIMAL registration was empirical, as we found an improvement in the segmentation performance of SACHA in a set of 10 healthy controls (Dice index = 82.2 ± 3.3 *vs.* 80.5 ± 3.2, $t = 3.2$, $p = 0.005$).
2. FreeSurfer. In this approach, the hippocampus is segmented using a nonlinear template matching (Fischl et al., 2002). After linearly registering the test image to the template, the algorithm estimates the nonlinear transformation between a given MRI and a probabilistic atlas of the hippocampus constructed from a cohort of 14 young and middle-aged subjects using a maximum likelihood criterion. Probabilistic labels are warped back to the individual MRI using the inverse of this transform. The final segmentation is accomplished by maximizing the *a posteriori* probability in the Bayes formula at each voxel. Voxel-wise probabilistic labels and their predicted image intensities serve as the prior term, while the intensity similarity between the target image and the template serves as the likelihood term.
3. Multi-atlas approach based on ANIMAL registration (Collins and Pruessner, 2010) (henceforth denoted ANIMAL-multi). In brief,

MR images of controls and patients were linearly registered to the MNI ICBM-152 nonlinear template (Fonov et al., 2011) using the transformation parameters generated for manual labeling. We created a template library including these images and corresponding manual labels. The performance was evaluated using a leave-one-out strategy in which we iteratively excluded the image to segment from the template library. For a given test image, we first nonlinearly warped each individual template to it using the ANIMAL registration (Collins et al., 1995). Then, we computed a similarity index between the warped template and the test image defined by the normalized mutual information within a volume of interest centered on the hippocampus. We selected the n most similar templates and obtained the final segmentation using label fusion on a voting strategy (Aljabar et al., 2009; Heckemann et al., 2006; Rohlfing et al., 2004) in which segmentations were averaged and thresholded at 0.5.

Models of hippocampal malrotation

The most representative indicators of hippocampal malrotation are medial positioning, vertical orientation, and increased depth of the collateral sulcus (Baulac et al., 1998; Bernasconi et al., 2005). To quantify these characteristics, in each subject we generated 3D models of the left and right manual hippocampal labels, and automatically extracted sulci (Kim et al., 2008; Voets et al., 2011). We chose the MNI ICBM-152 nonlinear template (Fonov et al., 2011) on which we manually segmented the hippocampus as the reference for computing malrotation features.

1. Sagittal translation (Supplementary Fig. 1A), measuring the position of the hippocampus relative to the midline, was calculated as the distance between the geometric centre of the hippocampus and the mid-sagittal plane (determined at 0 on the x-axis of the Talairach coordinate system).
2. Axial rotation (θ_{axial}; Supplementary Fig. 1B), reflecting a medial–lateral deflection of the hippocampus relative to its geometric centre, measured the degree of rotation with respect to the z-axis between the 1st principal (*i.e.*, longitudinal) axis ($\mathbf{v}_{1,T,}$) of a given hippocampus and the 1st principal axis ($\mathbf{v}_{1,R}$) of the reference. θ_{axial} was positive/negative, when the caudal portion of the hippocampus was medially/laterally positioned relative to its rostral portion.
3. Longitudinal rotation (θ_{long}; Supplementary Fig. 1C), indicating a relative vertical deviation of the hippocampus from its normal horizontal orientation. By aligning the 1st principal axis of a target hippocampus ($\mathbf{v}_{1,T}$) to that of the reference ($\mathbf{v}_{1,R}$) using a rotation matrix (so that $\mathbf{v'}_{1,T}=\mathbf{v}_{1,R}$), the 2nd principal axis of the target ($\mathbf{v}_{2,T}$) was rotated to $\mathbf{v'}_{2,T}$. θ_{long} was measured as the angle between the 2nd principal axis of the target ($\mathbf{v'}_{2,T}$) and the reference ($\mathbf{v}_{2,R}$). θ_{long} was positive, respectively negative, when the target hippocampus was more vertically oriented than the reference due to upward, respectively downward, rotation.
4. Collateral sulcus depth. Using BrainVISA (Riviere et al., 2002), we constructed surfaces corresponding to GM–WM and GM–CSF interfaces and automatically extracted the collateral sulcus. The 3D sulcal surface was then binarized into voxels. A rater (HK) inspected the accuracy of collateral sulcus extraction. In addition, the rater verified that the deepest voxel of the sulcal line on consecutive coronal sections was correctly located at the fundus, and performed manual corrections in case of mislocation. Sulcal depth was determined by calculating the shortest distance between the outer cortical surface and the voxel at the sulcal fundus (Smith, 2002). Lastly, we averaged the depths of all sulcal bottom points on consecutive coronal slices within the extent of the hippocampus. We determined the prevalence of hippocampal malrotation based on a 2 standard deviation cutoff from the distribution of healthy controls (Supplementary Table).

Performance evaluation of automated segmentation algorithms

To evaluate the segmentation accuracy and perform statistical testing, SACHA and FreeSurfer segmentations were mapped to the MNI ICBM-152 nonlinear template (Fonov et al., 2011) using the linear transformation parameters generated for manual labeling.

1. Volume-based agreements. We performed correlation analyses between manual and automated segmentations, and constructed Bland-Altman plots (Bland and Altmann, 1986), on which for each subject the difference between the two measurements is drawn against their mean.
   To assess the segmentation accuracy, we computed the Dice similarity index: $D=2\times v(M\cap A)/(v(M)+v(A))$, where M/A are the voxels comprising manual/automated labels; "M∩A" are voxels in the intersection of M and A; v $(\cdot)$ is the volume operator. For each algorithm, we compared Dice indices between patient groups (*i.e.* LTLE and RTLE) and controls using Student's t-tests.
   In a separate analysis, we assessed the ability of the automatic algorithms to detect atrophy in TLE groups relative to controls by computing Cohen's d ((mean volume controls − mean volume TLE)/pooled SD) that measures the effect size of a between-group difference, and calculated the significance of the observed effect using t-tests. Significances were adjusted for multiple comparisons using Bonferroni correction.
2. Surface-based analysis of contour accuracy. To assess potential systematic shape biases, manual and automated labels were converted to surface meshes and parameterized using an area-preserving, distortion-minimizing mapping technique based on spherical harmonics (SPHARM) (Styner et al., 2006). The uniform icosahedron-subdivision of the SPHARM allows obtaining a point distribution model (PDM) that guarantees shape-inherent vertex-wise correspondence across subjects. For each algorithm, we pooled controls and patients, computed vertex-wise surface-normal component of the displacement vector between the automated and manual label, and computed t-tests on the differences (Morey et al., 2009). In addition, we mapped the SD of the displacement vector at each vertex. Differences in SD between SACHA, FreeSurfer, and ANIMAL-multi were assessed using F-tests. Results were thresholded for statistical significance using the False Discovery Rate (FDR) correction at $q<0.05$ (Benjamini and Hochberg, 1995).
3. Performance evaluation with respect to atrophy and malrotation. For both volume- and surface-based analyses, we used linear models to investigate the effect of hippocampal atrophy and malrotation on variations of the Dice index and the surface normal components of the displacement vector, respectively.

Seizure focus lateralization

As stated earlier, the lateralization of the seizure focus is based on the convergence of clinical, electrographic and radiological data. To evaluate the yield of MRI volumetry alone to lateralize the seizure focus we performed a linear discriminant analysis. For each set of hippocampal labels (*i.e.*, manual and automated), we calculated an asymmetry ratio as $2(L-R)/(L+R)$, where L/R stands for volume of the left/right hippocampus. This ratio was standardized using a z-transform relative to the distribution of controls. To determine the side of seizure focus in each patient, we input the standardized ratio into the classifier. To maximize specificity, we defined decision margins so that no classification fell within the asymmetry range of healthy controls. Cross-validation was performed using a leave-one-out approach. This procedure, by which an individual patient is classified on the basis of the remaining patients, allows an unbiased assessment of lateralization performance for previously unseen TLE cases.

Table 1
Comparison between automated segmentation methods and manual labeling using Dice similarity index.

	SACHA		FreeSurfer		ANIMAL-multi	
	Left	Right	Left	Right	Left	Right
Controls (n = 35)	81.2 (±3.3)	82.6 (±3.1)	70.1 (±2.5)*	73.5 (±2.6)	83.5 (±4.8)	85.4 (±4.3)
LTLE (n = 35)	**77.4 (±4.3)***	80.8 (±3.0)	**62.2 (±5.4)***	70.4 (±3.4)	80.7 (±6.5)*	84.1 (±8.6)
RTLE (n = 31)	79.9 (±4.9)	79.8 (±5.6)	68.2 (±3.9)	**68.4 (±5.5)**	82.4 (±5.0)	82.3 (±5.7)

Dice indices are presented in % mean (±SD); LTLE/RTLE: left/right temporal lobe epilepsy; for each algorithm differences in accuracy between controls and patients are in bold.
* Indicates hemispheric asymmetry; significances are adjusted for multiple comparisons using Bonferroni correction ($p = 0.05/18 = 0.003$).

Results

We observed strong correlations between manual and automated segmentations (SACHA: $r = 0.86$; FreeSurfer: $r = 0.90$; ANIMAL-multi: $r = 0.83$, $p < 10^{-10}$; Supplementary Fig. 2). The Bland–Altman plots showed that differences in volume between manual and automated procedures did not differ ($t < 2.0$, $p > 0.1$; Supplementary Fig. 3).

Volume-based assessment of segmentation accuracy

Dice similarity indices between automated and manual hippocampal segmentations in healthy controls and TLE patients are detailed in Table 1. Overall, ANIMAL-multi (mean Dice index 83.1 ± 5.7) was superior to the other algorithms (*vs.* SACHA: 79.4 ± 6.2; $t = 4.4$, $p < 0.001$; *vs.* FreeSurfer: 68.9 ± 5.4; $t = 17.1$, $p < 0.001$). ANIMAL-multi and SACHA performed better than FreeSurfer in all groups ($p < 0.001$). ANIMAL-multi performed equally in patients and controls ($p > 0.1$), whereas SACHA and FreeSurfer showed lower accuracy in patients ($p < 0.05$).

For SACHA, Dice index in healthy controls did not correlate with variations in hippocampal volume or malrotation ($|r| < 0.34$, $p > 0.2$; Table 2). On the other hand, in patients, a lower Dice index was associated with increased degrees of hippocampal longitudinal rotation and collateral sulcus depth ($|r| > 0.5$, $p < 0.04$). While this effect was seen ipsilaterally in LTLE, correlations were bilateral in RTLE.

For FreeSurfer, Dice index in controls was positively correlated with left hippocampal volume ($r = 0.55$, $p = 0.02$). In TLE patients, it was influenced by both atrophy ($r > 0.78$, $p < 0.001$) and longitudinal rotation ($|r| > 0.55$, $p < 0.02$) bilaterally. Interaction analysis, however, revealed a larger effect of atrophy than malrotation ($t > 5.4$, $p < 0.00001$) on the Dice index.

For ANIMAL-multi, Dice index in healthy controls was not influenced by variations in hippocampal volume or malrotation ($|r| < 0.34$, $p > 0.2$), whereas in TLE patients a lower index was associated with more severe atrophy ($r > 0.58$, $p < 0.001$) and longitudinal rotation ($|r| > 0.49$; $p < 0.05$) ipsilateral to the seizure focus. These associations were weaker than those observed for FreeSurfer ($t > 3.7$, $p < 0.001$).

Shape analysis of segmentations

Shape differences between manual and automated hippocampal tracings are shown in Supplementary Fig. 4. SACHA and ANIMAL-multi produced segmentation results that were globally similar to manual labeling. However, SACHA systematically underestimated the supero-medial margin (mean error: −0.9 ± 0.4 mm, FDR < 0.05) and overestimated the infero-medial boundary (mean error: 1.2 ± 0.8 mm, FDR < 0.05) corresponding to the subiculum. ANIMAL-multi underestimated the superior and medial hippocampal margin (mean error: −0.6 ± 0.3 mm, FDR < 0.05). FreeSurfer, on the other hand, globally overestimated the hippocampus (mean error: 1.7 ± 2.1 mm, FDR < 0.05). Moreover, the SD of the displacement vectors was overall higher for FreeSurfer (2.1 mm) than the two other algorithms (*vs.* SACHA: 0.8 mm, $p < 0.00001$; *vs.* ANIMAL-multi: 0.7 mm, $p < 0.00001$).

Impact of malrotation and atrophy upon segmentation accuracy

For SACHA, the overestimation of the lateral boundary of left hippocampus was related to a lower Dice index ($r = -0.67$; FDR < 0.001, Fig. 1A). In this region, we found a positive correlation between the segmentation error and the depth of the collateral sulcus ($r = 0.61$; FDR < 0.001, Fig. 1B). As illustrated in Fig. 2, the deeper the collateral sulcus, the more it protrudes vertically towards the wall of the lateral ventricle, thus coming in contact with the lateral border of the hippocampus. In this case, the automatic procedure tends to erroneously include the fundus of the sulcus into the hippocampal tracing. Atrophy on the other hand, did not have a negative impact on segmentation accuracy.

Table 2
Correlations between automatic segmentation accuracy and hippocampal volume/malrotation features.

	Volume		Sagittal translation		Axial rotation		Longitudinal rotation		Collateral sulcus depth	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
SACHA										
Controls	0.18	0.09	−0.08	−0.06	−0.34	−0.22	−0.25	−0.32	−0.19	−0.13
LTLE	0.37	0.30	−0.12	−0.10	−0.35	−0.17	**−0.51**	−0.41	**−0.55**	−0.37
RTLE	0.31	0.43	−0.23	−0.04	−0.13	−0.11	−0.44	**−0.52**	**−0.61**	**−0.68**
FreeSurfer										
Controls	**0.55**	0.48	−0.05	0.03	0.05	−0.11	−0.04	−0.12	−0.24	−0.09
LTLE	**0.77**	**0.79**	−0.17	−0.12	−0.05	−0.33	**−0.55**	−0.45	−0.36	−0.31
RTLE	**0.78**	**0.79**	−0.13	0.05	−0.11	−0.03	**−0.60**	**−0.55**	−0.31	−0.35
ANIMAL-multi										
Controls	0.34	0.17	−0.20	−0.23	0.24	−0.14	−0.04	−0.30	−0.14	−0.22
LTLE	**0.58**	0.32	−0.37	−0.21	−0.05	−0.15	**−0.49**	−0.33	−0.37	−0.40
RTLE	0.35	**0.64**	−0.13	−0.33	−0.11	−0.36	−0.39	**−0.50**	−0.31	−0.34

Significances (in bold) are adjusted for multiple comparisons using Bonferroni correction ($p = 0.05/90 = 0.0006$); LTLE/RTLE: left/right temporal lobe epilepsy

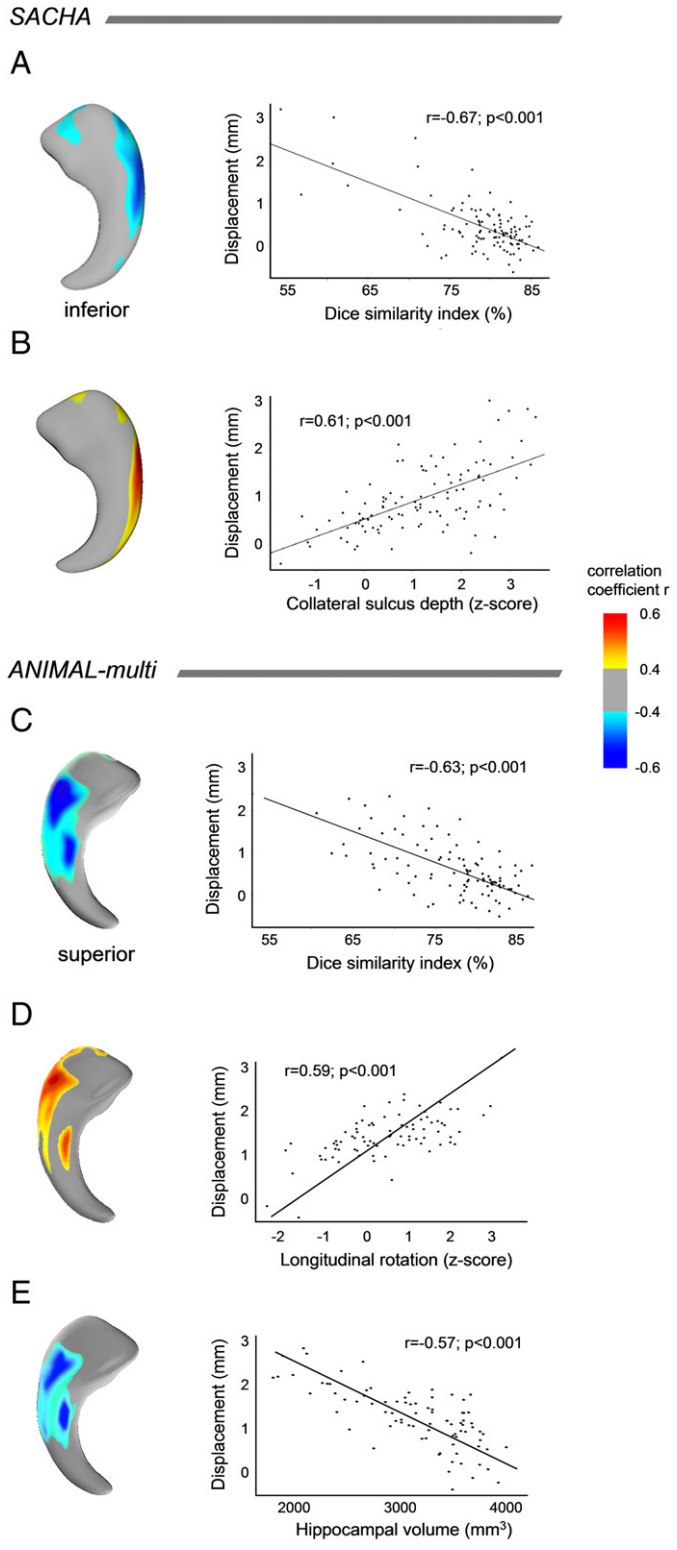

Fig. 1. Surface-based analysis of contour accuracy. Correlation maps and plots of vertex-wise surface-normal components of the displacement vectors between manual tracing and automated segmentation with Dice similarity index (A, C), collateral sulcus depth (B), hippocampal longitudinal rotation (D) and hippocampal volume (E). The left hippocampus is shown. In A, the five outliers with a Dice index below 65% are patients with hippocampal malrotation.

For ANIMAL-multi, the overestimation of the supero-lateral hippocampal boundary was related to a lower Dice index (left: $r=-0.63$, FDR<0.001; right: $r=-0.56$, FDR=0.002). The overestimation was related to longitudinal rotation (left: $r=0.59$, FDR<0.001; right: $r=0.53$, FDR=0.005). We found a negative correlation between the segmentation error and hippocampal volume on the superior boundary, *i.e.* the more atrophic the hippocampus, the larger the error (left: $r=-0.57$, FDR<0.001; right: $r=-0.51$; FDR=0.01). Fig. 1 C to E shows results for the left hippocampus.

For FreeSurfer, the overestimation of hippocampal volume was associated with lower Dice index across all vertices bilaterally ($r>-0.71$; FDR<0.001). We also found a negative correlation between the segmentation error and volume ($r=-0.78$; FDR<0.001). On the other hand, there was no association between the segmentation error and malrotation ($r<0.35$; FDR>0.1).

Figs. 2 and 3 show examples of segmentation errors related to malrotation and atrophy.

Ability of automated methods to lateralize the seizure focus in patients

Group analysis

Both manual and automated approaches identified hippocampal atrophy ipsilateral to the seizure focus (Table 3). However, compared to manual volumetry (Cohen's *d* LTLE/RTLE: 1.96/1.41, t=7.9/4.9), automated methods detected a smaller magnitude of volume reduction (*vs.* SACHA: 1.26/0.94, t=5.2/3.8; *vs.* FreeSurfer: 1.02/0.77, t=3.9/3.6; *vs.* ANIMAL-multi: 1.21/1.01, t=4.9/4.0).

Individual analysis

Results are shown in Fig. 4. Using decision margins that enforced non-classification of patients in whom asymmetry values fell within the range of healthy controls, manual volumetry correctly lateralized the side of the focus in 67% (44/66) of patients. Automated methods lateralized the seizure focus in fewer patients (Fisher's exact test; SACHA: 61%=40/66, p=0.29; FreeSurfer: 56%=37/66, p=0.14; ANIMAL-multi: 58%=38/66, p=0.19). SACHA and ANIMAL-multi misclassified one RTLE patient as LTLE. In this patient with malrotation (longitudinal rotation: z=2.1), the volume of the right hippocampus was overestimated because SACHA erroneously included the fundus of the collateral sulcus in the segmentation and ANIMAL-multi incorporated voxels adjacent to the lateral boundary of the hippocampus.

Among the 23 patients with hippocampal malrotation, 14 (64%) were correctly lateralized using manual volumetry, 12 (55%) with ANIMAL-multi, 11 (50%) by SACHA, and 9 (41%) using FreeSurfer (Fisher's exact test; ANIMAL-multi: p=0.38; SACHA: p=0.15; FreeSurfer: p=0.05).

Discussion

We quantitatively evaluated the influence of hippocampal shape and positioning anomalies on the performance of three state-of-the-art automatic segmentation algorithms. Overall, our results suggest that the accuracy and clinical utility of automated methods are suboptimal in patients in whom shape grossly deviates from the range of normal variants, but also in case of marked atrophy. SACHA overestimated the volume of the hippocampus in its lateral border in cases with malrotation, but behaved relatively well in the presence of atrophy. On the other hand, ANIMAL-multi and FreeSurfer were affected by hippocampal malrotation and atrophy, although the magnitude of these anomalies were smaller in the former. Compared to manual segmentation, automated procedures underestimated the magnitude of atrophy. Importantly, the automated methods did not determine the side of the seizure focus in four (SACHA), six (ANIMAL-multi) and seven (FreeSurfer) patients who were correctly lateralized using manual labeling, and the presence of malrotation lead to wrong lateralization in one patient. In a clinical setting where decisions concerning surgery are taken on a patient-by-patient basis, even a relatively small decrease in sensitivity and lateralization performance may have detrimental consequences.

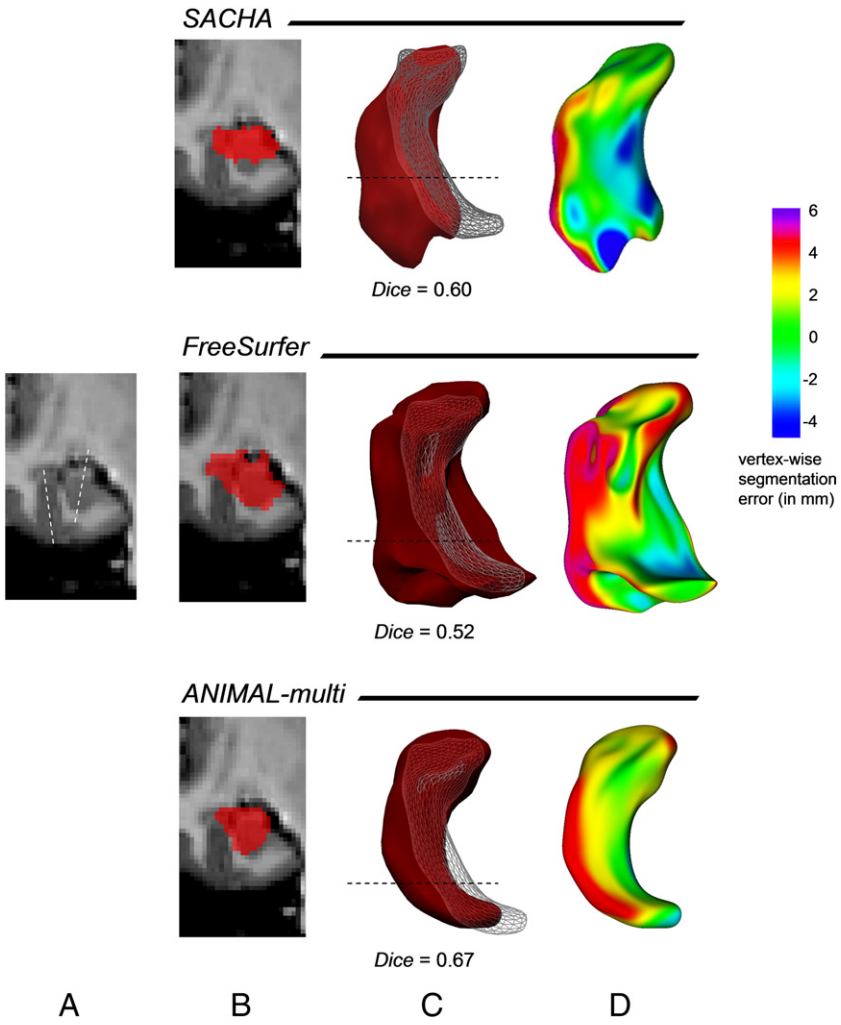

Fig. 2. Impact of hippocampal malrotation on automated algorithms. A) Coronal MRI section of a patient with atypical positioning characterized by vertical orientation of both the hippocampus and the collateral sulcus (dotted lines). B) Automatically generated hippocampal labels overlaid on MRI. C) Parameterized surfaces of automatic labels (red) overlaid on the manual tracing (wireframe). The horizontal dotted lines in the hippocampal body correspond to the level of the coronal sections shown in panel A and B. D). Vertex-wise differences in contour (*i.e.*, segmentation error) between manual and automated labels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Compared to manual segmentation, FreeSurfer performed rather poorly in both controls and patients, a finding in agreement with two previous studies examining smaller cohorts (Akhondi-Asl et al., 2011; Pardoe et al., 2009). While differences in segmentation protocols may have contributed to these unsatisfactory results (Konrad et al., 2009), consistently reduced performance is more likely driven by the use of a single-template in this algorithm. Such template, derived from 14 healthy subjects, may not capture salient structural variations. Our surface-based analysis of contour accuracy revealed that the magnitude of random errors in segmentation was larger in FreeSurfer than the two other algorithms, as shown by the overall higher standard deviation of the displacement vectors. In addition, the accuracy of FreeSurfer was influenced by both atrophy and longitudinal rotation, and this association was stronger than that of ANIMAL-multi. Thus, FreeSurfer's low performance throughout the hippocampus makes a bias related exclusively to segmentation protocols unlikely. On the other hand, FreeSurfer's performance with respect to seizure focus lateralization was satisfactory. Thus, a comprehensive benchmark assessment evaluating not only accuracy but also the clinical usefulness likely provides a more realistic view of the global performance of a given automated technique.

In SACHA, the initial segmentation resulting from the non-linear warping is further refined through local deformation guided by anatomical priors. This process may reduce, at least to some extent, biases related to the use of a single-template. Nevertheless, priors derived from healthy subjects may be unsuitable when the hippocampus does not display the usual relationships with its surroundings. Indeed, surface-based shape mapping revealed errors in the presence of a deep collateral sulcus protruding towards the wall of the ventricle and thus coming in close contact with the lateral border of the hippocampus. This atypical anatomical arrangement results in a partial volume effect of the collateral sulcus in the white matter of the parahippocampal gyrus, which appears darkened, with intensity gradients resembling that of the hippocampus. In this scenario, the region growing process falsely penetrates the sulcus and includes its fundus into the tracing of the hippocampus. These intensity changes, however, are not likely to mislead manual delineation, since the white matter remains distinguishable from the hippocampus on visual inspection. In addition, the rater generally utilizes multiple sources of information to determine anatomical landmarks.

By building expert prior-based models of shape and selecting automatically those that best fit the structure to segment, algorithms using multi-template libraries and label fusion (Aljabar et al., 2009; Collins and Pruessner, 2010; Heckemann et al., 2006) have the potential to overcome the limitations of individual or averaged template techniques. These multi-template techniques have so far been rarely

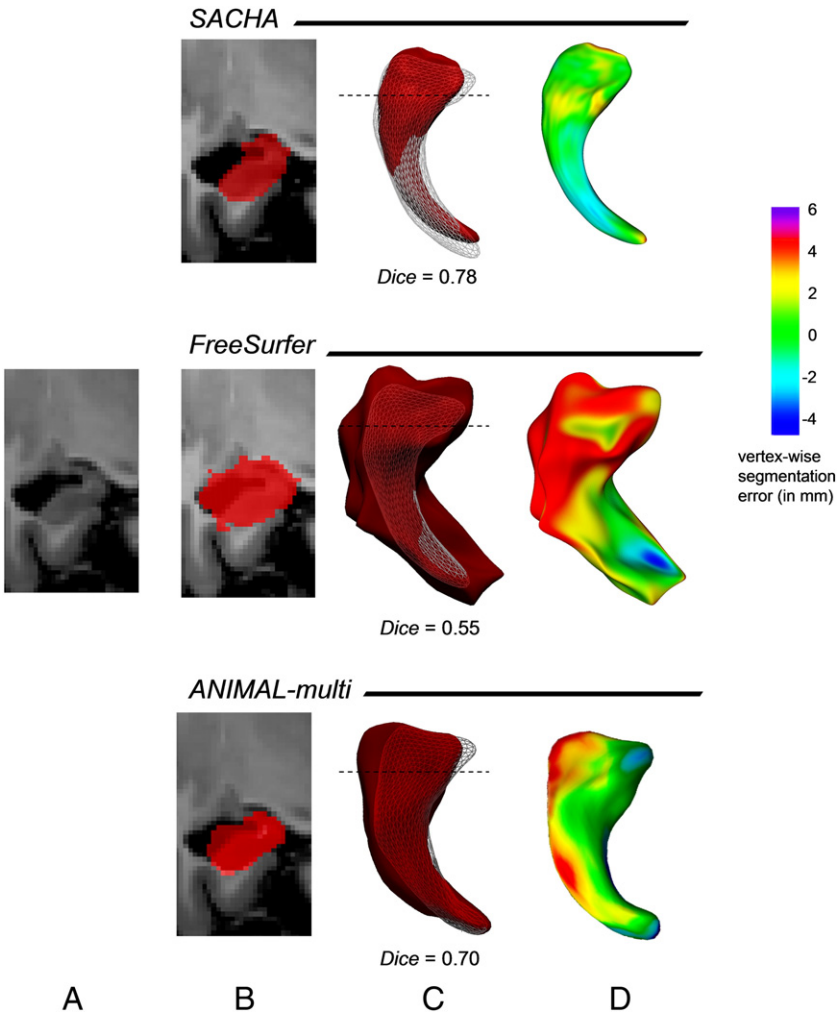

Fig. 3. Impact of hippocampal atrophy on automated algorithms. A) Coronal MRI section of a patient with hippocampal atrophy. B) Automatically generated hippocampal labels overlaid on MRI. C) Parameterized surfaces of automatic labels (red) overlaid on the manual tracing (wireframe). The horizontal dotted lines in the hippocampal head correspond to the level of the coronal sections in A and B. D) Vertex-wise differences in contour (*i.e.*, segmentation error) between manual and automated labels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

used to segment pathological hippocampi (Barnes et al., 2008; Leung et al., 2010). Compared to a previous study based on 18 TLE patients that yielded a Dice index of 78.3% (Hammers et al., 2007), ANIMAL-multi's performance was 82.4% in our study. The higher performance likely stems from our database constructed from a large sample of patients and controls. In addition, the template selection strategy of ANIMAL-multi using mutual information as image similarity metric has been shown to outperform the traditional random subset selection approach (Aljabar et al., 2009).

Hippocampal developmental anomalies are prevalent in TLE (Baulac et al., 1998; Bernasconi et al., 2005; Depondt et al., 2002; Sloviter et al., 2004; Stiers et al., 2010; Voets et al., 2011), but also found in up to 60% of patient with various types of malformations of cortical development (Baulac et al., 1998; Bernasconi et al., 2005; Sato et al., 2001), and have been documented in schizophrenia (Connor et al., 2004) and autism (Salmond et al., 2002). We therefore believe that malrotation should be taken into account when designing automated segmentation methods. Noteworthy, in our study, the use of a large cohort of patients including those with malrotation likely broadened the spectrum of shape variations within the template library, allowing ANIMAL-multi to achieve the highest segmentation accuracy. Statistical texture models (Hamarneh and Li, 2009;

Table 3
Group analysis.

	Controls		LTLE		RTLE	
	Left	Right	Left	Right	Left	Right
Manual	3668 ± 368	3924 ± 434	**2551 ± 747** (1.96)*	3593 ± 727 (0.62)	3463 ± 479 (0.49)	**3134 ± 842** (1.41)*
SACHA	3954 ± 572	3906 ± 540	**2979 ± 947** (1.26)*	3709 ± 723 (0.31)	3800 ± 524 (0.28)	**3205 ± 998** (0.94)*
FreeSurfer	6216 ± 573	6147 ± 547	**5391 ± 1129** (1.02)*	6301 ± 725 (−0.03)	6149 ± 769 (−0.03)	**5301 ± 1274** (0.77)*
ANIMAL-multi	3558 ± 345	3885 ± 470	**2985 ± 595** (1.21)*	3649 ± 966 (0.33)	3418 ± 591 (0.30)	**3301 ± 703** (1.01)*

Hippocampal volume in mm^3 is presented in mean ± SD; values in parentheses indicate Cohen's *d* index, *i.e.* the strength of the effect size of hippocampal atrophy (0.2 indicates a small effect, 0.5 a medium effect, and >0.8 a large effect); group-wise significances in volumes (bold) and effect size (*) are adjusted for multiple comparisons using Bonferroni correction ($p = 0.05/24 = 0.0021$).

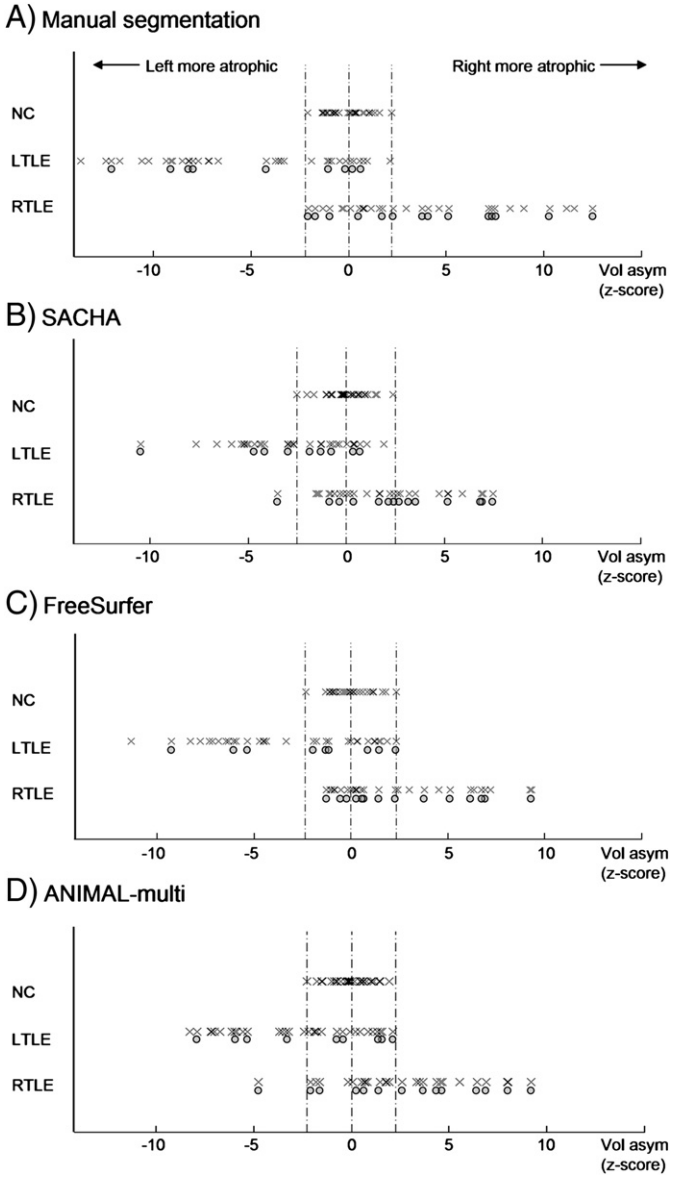

Fig. 4. Seizure focus lateralization in TLE. For each segmentation method, the hippocampal asymmetry ratio was standardized using z-transformation relative to the distribution of healthy controls and then fed into a linear discriminant function classifier. Crosses denote individuals in controls (NC) and patients with left/right temporal lobe epilepsy (LTLE/RTLE). Circles below the crosses identify patients with hippocampal malrotation. To maximize the specificity of the classifier, we defined decision margins (broken lines) so that no classification fell within the asymmetry range of controls.

Paragios and Deriche, 2002) may further capture loco-regional anatomical variability. Finally, modeling shape is likely to guide the deformation within the range of the template library.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2011.11.040.

Acknowledgment

This work was supported by the Canadian Institutes of Health Research (CIHR MOP-57840, CIHR MOP-93815).

References

Akhondi-Asl, A., Jafari-Khouzani, K., Elisevich, K., Soltanian-Zadeh, H., 2011. Hippocampal volumetry for lateralization of temporal lobe epilepsy: automated versus manual methods. Neuroimage 54, S218–S226.

Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46, 726–738.

Ashburner, J., Friston, K.J., 1999. Nonlinear spatial normalization using basis functions. Hum. Brain Mapp. 7, 254–266.

Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C., 2010. The optimal template effect in hippocampus studies of diseased populations. Neuroimage 49, 2457–2466.

Babb, T.L., Brown, W.J., 1987. Pathological findings in epilepsy. In: Engel Jr., J. (Ed.), Surgical Treatment of the Epilepsies. Raven, New York, NY, pp. 511–540.

Barnes, J., Foster, J., Boyes, R.G., Pepple, T., Moore, E.K., Schott, J.M., Frost, C., Scahill, R.I., Fox, N.C., 2008. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. Neuroimage 40, 1655–1671.

Baulac, M., De Grissac, N., Hasboun, D., Oppenheim, C., Adam, C., Arzimanoglou, A., Semah, F., Lehericy, S., Clemenceau, S., Berger, B., 1998. Hippocampal developmental changes in patients with partial epilepsy: magnetic resonance imaging and clinical aspects. Ann. Neurol. 44, 223–233.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57, 289–300.

Bernasconi, N., Bernasconi, A., Caramanos, Z., Antel, S.B., Andermann, F., Arnold, D.L., 2003. Mesial temporal damage in temporal lobe epilepsy: a volumetric MRI study of the hippocampus, amygdala and parahippocampal region. Brain 126, 462–469.

Bernasconi, N., Kinay, D., Andermann, F., Antel, S., Bernasconi, A., 2005. Analysis of shape and positioning of the hippocampal formation: an MRI study in patients with partial epilepsy and healthy controls. Brain 128, 2442–2452.

Bland, J.M., Altmann, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurements. Lancet 327, 307–310.

Blumcke, I., Thom, M., Wiestler, O.D., 2002. Ammon's horn sclerosis: a maldevelopmental disorder associated with temporal lobe epilepsy. Brain Pathol. 12, 199–211.

Chupin, M., Mukuna-Bantumbakulu, A.R., Hasboun, D., Bardinet, E., Baillet, S., Kinkingnehun, S., Lemieux, L., Dubois, B., Garnero, L., 2007. Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: method and validation on controls and patients with Alzheimer's disease. Neuroimage 34, 996–1019.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O., 2009a. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19, 579–587.

Chupin, M., Hammers, A., Liu, R.S., Colliot, O., Burdett, J., Bardinet, E., Duncan, J.S., Garnero, L., Lemieux, L., 2009b. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. Neuroimage 46.

Collins, D.L., Pruessner, J.C., 2010. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. Neuroimage 52, 1355–1366.

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.

Collins, D.L., Holmes, C.J., Peters, T.M., Evans, A.C., 1995. Automatic 3-D model-based neuroanatomical segmentation. Hum. Brain Mapp. 3, 190–208.

Connor, S.E.J., Ng, V., McDonald, C., Schulze, K., Morgan, K., Dazzan, P., Murray, R.M., 2004. A study of hippocampal shape anomaly in schizophrenia and in families multiply affected by schizophrenia or bipolar disorder. Neuroradiology 46, 523–534.

Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L., 2011. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. Neuroimage 54, 940–954.

Depondt, C., Van Paesschen, W., Matthijs, G., Legius, E., Martens, K., Demaerel, P., Wilms, G., 2002. Familial temporal lobe epilepsy with febrile seizures. Neurology 58, 1429–1433.

Duchesne, S., Pruessner, J., Collins, D.L., 2002. Appearance-based segmentation of medial temporal lobe structures. Neuroimage 17, 515–531.

Engel Jr., J., Van Ness, P.C., Rasmussen, T., Ojemann, L.M., 1993. Outcome with respect to epileptic seizures. In: Engel Jr., J. (Ed.), Surgical Treatment of the Epilepsies. Raven, New York, pp. 609–621.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der, K.A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation. Automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., 2011. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 54, 313–327.

Frisoni, G.B., Laakso, M.P., Beltramello, A., Geroldi, C., Bianchetti, A., Soininen, H., Trabucchi, M., 1999. Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease. Neurology 52, 91–100.

Hamarneh, G., Li, X., 2009. Watershed segmentation using prior shape and appearance knowledge. Image Vis. Comput. 27, 59–68.

Hammers, A., Heckemann, R., Koepp, M.J., Duncan, J.S., Hajnal, J.V., Rueckert, D., Aljabar, P., 2007. Automatic detection and quantification of hippocampal atrophy on MRI in temporal lobe epilepsy: a proof-of-principle study. Neuroimage 36, 38–47.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33, 115–126.

Heckemann, R.A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J.V., Hammers, A., 2010. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. Neuroimage 51, 221–227.

Jack Jr., C.R., Petersen, R.C., O'Brien, P.C., Tangalos, E.G., 1992. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42, 183–188.

Jackson, G.D., Berkovic, S.F., Tress, B.M., Kalnins, R.M., Fabinyi, G.C., Bladin, P.F., 1990. Hippocampal sclerosis can be reliably detected by magnetic resonance imaging. Neurology 40, 1869–1875.
Jackson, G.D., Berkovic, S.F., Duncan, J.S., Connelly, A., 1993. Optimizing the diagnosis of hippocampal sclerosis using MR imaging. AJNR 14, 753–762.
Kelemen, A., Szekely, G., Gerig, G., 1999. Elastic model-based segmentation of 3-D neuroradiological data sets. IEEE Trans. Med. Imaging 18, 828–839.
Khan, A.R., Wang, L., Beg, M.F., 2008. FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using Large Deformation Diffeomorphic Metric Mapping. Neuroimage 41, 735–746.
Kim, H., Bernasconi, N., Bernahardt, B., Colliot, O., Bernasconi, A., 2006. Quantitative analysis of shape and positioning of the hippocampal formation in patients with temporal lobe epilepsy. Epilepsia 47, 72.
Kim, H., Bernasconi, N., Bernhardt, B., Colliot, O., Bernasconi, A., 2008. Basal temporal sulcal morphology in healthy controls and patients with temporal lobe epilepsy. Neurology 70, 2159–2165.
Konrad, C., Ukas, T., Nebel, C., Arolt, V., Toga, A.W., Narr, K.L., 2009. Defining the human hippocampus in cerebral magnetic resonance images — an overview of current segmentation protocols. Neuroimage 47, 1185–1195.
Kuzniecky, R., Bilir, E., Gilliam, F., Faught, E., Martin, R., Hugg, J., 1999. Quantitative MRI in temporal lobe epilepsy: evidence for fornix atrophy. Neurology 53, 496–501.
Lehericy, S., Baulac, M., Chiras, J., Pierot, L., Martin, N., Pillon, B., Deweer, B., Dubois, B., Marsault, C., 1994. Amygdalohippocampal MR volume measurements in the early stages of Alzheimer disease. AJNR Am. J. Neuroradiol. 15, 929–937.
Leung, K.K., Barnes, J., Ridgway, G.R., Bartlett, J.W., Clarkson, M.J., Macdonald, K., Schuff, N., Fox, N.C., Ourselin, S., 2010. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. Neuroimage 51, 1345–1359.
Lötjönen, J.M.P., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage 49, 2352–2365.
Lötjönen, J., Wolz, R., Koikkalainen, J., Julkunen, V., Thurfjell, L., Lundqvist, R., Waldemar, G., Soininen, H., Rueckert, D., 2011. Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease. Neuroimage 56, 185–196.
Meencke, H.J., Veith, G., 1991. Hippocampal sclerosis in epilepsy. In: Lüders, H. (Ed.), Epilepsy Surgery. Raven Press, New York, pp. 705–715.
Morey, R.A., Petty, C.M., Xu, Y., Pannu Hayes, J., Wagner Ii, H.R., Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G., 2009. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. Neuroimage 45, 855–866.
Paragios, N., Deriche, R., 2002. Geodesic active regions and level set methods for supervised texture segmentation. Int. J. Comput. Vis. 46, 223–247.
Pardoe, H.R., Pell, G.S., Abbott, D.F., Jackson, G.D., 2009. Hippocampal volume assessment in temporal lobe epilepsy: how good is automated segmentation? Epilepsia 50.
Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M., 2007. A hierarchical algorithm for MR brain image parcellation. IEEE Trans. Med. Imaging 26, 1201–1212.
Riviere, D., Mangin, J.F., Papadopoulos-Orfanos, D., Martinez, J.M., Frouin, V., Regis, J., 2002. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. Med. Image Anal. 6, 77–92.
Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 21, 1428–1442.
Salmond, C.H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D.G., Friston, K.J., 2002. Distributional assumptions in voxel-based morphometry. Neuroimage 17, 1027–1030.
Sato, N., Hatakeyama, S., Shimizu, N., Hikima, A., Aoki, J., Endo, K., 2001. MR evaluation of the hippocampus in patients with congenital malformations of the brain. AJNR Am. J. Neuroradiol. 22, 389–393.
Schramm, J., Clusmann, H., 2008. The surgery of epilepsy. Neurosurgery 62 (Suppl 2), 463–481 (discussion 481).
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.
Sloviter, R.S., Kudrimoti, H.S., Laxer, K.D., Barbaro, N.M., Chan, S., Hirsch, L.J., Goodman, R.R., Pedley, T.A., 2004. "Tectonic" hippocampal malformations in patients with temporal lobe epilepsy. Epilepsy Res. 59, 123–153.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.
Stiers, P., Fonteyne, A., Wouters, H., D'Agostino, E., Sunaert, S., Lagae, L., 2010. Hippocampal malrotation in pediatric patients with epilepsy associated with complex prefrontal dysfunction. Epilepsia 51, 546–555.
Styner, M., Oguz, I., Xu, S., Brechb¨uhler, C., Pantazis, D., Gerig, G., 2006. Statistical Shape Analysis of Brain Structures using SPHARM-PDM. MICCAI 2006 Opensource workshop.
van der Lijn, F., den Heijer, T., Breteler, M.M.B., Niessen, W.J., 2008. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. Neuroimage 43, 708–720.
Voets, N.L., Bernhardt, B.C., Kim, H., Yoon, U., Bernasconi, N., 2011. Increased temporolimbic cortical folding complexity in temporal lobe epilepsy. Neurology 76, 138–144.
Wolz, R., Heckemann, R.A., Aljabar, P., Hajnal, J.V., Hammers, A., Lotjonen, J., Rueckert, D., 2010. Measurement of hippocampal atrophy using 4D graph-cut segmentation: application to ADNI. Neuroimage 52, 109–118.
Xu, Y., Jack Jr., C.R., O'Brien, P.C., Kokmen, E., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, R.G., Petersen, R.C., 2000. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. Neurology 54, 1760–1767.
Yang, J., Duncan, J.S., 2004. 3D image segmentation of deformable objects with joint shape-intensity prior models using level sets. Med. Image Anal. 8, 285–294.