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,