

# Fully Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease and Mild Cognitive Impairment Applied on Data From ADNI

Marie Chupin,<sup>1,2\*</sup> Emilie Gérardin,<sup>1,2</sup> Rémi Cuingnet,<sup>1,2,3</sup> Claire Boutet,<sup>1,2,5</sup> Louis Lemieux,<sup>6</sup> Stéphane Lehericy,<sup>1,2,4,5</sup> Habib Benali,<sup>3</sup> Line Garnero,<sup>1,2</sup> and Olivier Colliot<sup>1,2</sup>  
and the Alzheimer's Disease Neuroimaging Initiative

**ABSTRACT:** The hippocampus is among the first structures affected in Alzheimer's disease (AD). Hippocampal magnetic resonance imaging volumetry is a potential biomarker for AD but is hindered by the limitations of manual segmentation. We proposed a fully automatic method using probabilistic and anatomical priors for hippocampus segmentation. Probabilistic information is derived from 16 young controls and anatomical knowledge is modeled with automatically detected landmarks. The results were previously evaluated by comparison with manual segmentation on data from the 16 young healthy controls, with a leave-one-out strategy, and eight patients with AD. High accuracy was found for both groups (volume error 6 and 7%, overlap 87 and 86%, respectively). In this article, the method was used to segment 145 patients with AD, 294 patients with mild cognitive impairment (MCI), and 166 elderly normal subjects from the Alzheimer's Disease Neuroimaging Initiative database. On the basis of a qualitative rating protocol, the segmentation proved acceptable in 94% of the cases. We used the obtained hippocampal volumes to automatically discriminate between AD patients, MCI patients, and elderly controls. The classification proved accurate: 76% of the patients with AD and 71% of the MCI converting to AD before 18 months were correctly classified with respect to the elderly controls, using only hippocampal volume. © 2009 Wiley-Liss, Inc.

**KEY WORDS:** segmentation; classification; Alzheimer's disease; MCI

## INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia; its early and accurate diagnosis is challenging. The hippocampus is a gray

<sup>1</sup> Université Pierre et Marie Curie-Paris6, CNRS, UMR-S7225, Paris, France; <sup>2</sup> Inserm, U975, Paris, France; <sup>3</sup> INSERM UMR S 678, Université Pierre et Marie Curie-Paris6, Paris, France; <sup>4</sup> Center for Neuroimaging Research-CENIR, Hôpital de la Pitié-Salpêtrière, France; <sup>5</sup> Department of Neuroradiology, Hôpital de la Pitié-Salpêtrière, France; <sup>6</sup> Department of Clinical and Experimental Epilepsy, Institute of Neurology, University College London, United Kingdom

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([www.loni.ucla.edu/ADNI](http://www.loni.ucla.edu/ADNI)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at [www.loni.ucla.edu/ADNI/Collaboration/ADNI\\_Authorship\\_list.pdf](http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf)).

Grant sponsor: NIH; Grant number: U01 AG024904.

\*Correspondence to: Marie Chupin, Université Pierre et Marie Curie-Paris6, CNRS, UMR-S7225, Paris F-75013, France. E-mail: [marie.chupin@upmc.fr](mailto:marie.chupin@upmc.fr)

Accepted for publication 23 March 2009

DOI 10.1002/hipo.20626

Published online 12 May 2009 in Wiley InterScience ([www.interscience.wiley.com](http://www.interscience.wiley.com)).

matter structure of the temporal lobe known to be affected at the earliest stage of AD, even before the diagnosis can be made, at the stage of mild cognitive impairment (MCI) (Braak and Braak, 1995). Hippocampal volumetry on magnetic resonance images (MRI) can thus constitute a useful diagnostic tool (Dubois et al., 2007). Till now, hippocampal volumetry mostly relies on highly time-consuming manual segmentation, which is rater-dependent, and not feasible in clinical routine.

Automatic segmentation of the hippocampus would overcome these limitations and provide a useful biomarker of AD. The incomplete definition of hippocampal boundaries on MRI scans makes the use of prior information necessary for accuracy and robustness. Prior knowledge can come from statistical information on shape (Kelemen et al., 1999; Shen et al., 2002), deformations (Duchesne et al., 2002), or from registering a single subject atlas template (Csernansky et al., 2000); nevertheless, these methods may be unsuitable for diseased structures. Segmentation using probabilistic information (Fischl et al., 2002; Heckemann et al., 2006) offers more thorough global spatial knowledge compared to single object atlases.

We proposed a fully-automatic method (Chupin et al., 2007, 2009) for the segmentation of the hippocampus (Hc) and the amygdala (Am), based on simultaneous region deformation driven by both anatomical and probabilistic priors. Anatomical information (Bloch et al., 2005) is derived from local anatomical patterns that are stable in controls and patients, around landmarks automatically detected during the deformation (Chupin et al., 2007). Probabilistic information is derived from an atlas built from the registration of manually segmented Hc and Am for 16 young healthy subjects (Chupin et al., 2009). Initialization is obtained from global information and deformation is constrained by local anatomical and probabilistic information.

The goal of this article is to further evaluate this segmentation method in patients with AD, MCI, and elderly controls from the ADNI database (Alzheimer's Disease Neuroimaging Initiative). It will assess the method's robustness with respect to different MRI scan-

ners and acquisition parameters, and with respect to pathology, some of the patients with AD having highly atrophic Hc. Segmentation accuracy will be evaluated with a qualitative rating protocol. Furthermore, we will assess the ability of resulting Hc volumes to discriminate patients with AD and MCI from elderly controls. We will also study the influence of some of image pre-processing steps, and of age group and normalization by total intracranial volume. A preliminary version of this work has been presented at the MICCAI CAPH'08 workshop.

## METHOD

The segmentation is based on the alternate deformation of two objects, one for Hc and one for Am, from two initial objects, through homotopic region deformation. It is modeled in a Bayesian framework, the deformation being driven by an iterative energy minimization. This energy is defined with a functional made of five terms: global and local data attachment, regularization, and volume and surface terms (Chupin et al., 2007). The initial objects are determined from the probabilistic atlases, inside an automatically extracted bounding box (Chupin et al., 2009). The energy functional is then iteratively minimized for Hc and Am, with additional constraints derived from the anatomical and probabilistic priors.

### Probabilistic Atlases

The datasets from  $N$  (here  $N = 16$ ) young healthy subjects were manually segmented by an expert following a protocol ensuring coherence in the three planes. For each of the atlas subject,  $\{S_i, i = 1 \dots N\}$ , the transformation  $T_i$  to the MNI standard space is then obtained through the unified registration and segmentation module of SPM5 (Ashburner and Friston, 2005) using the native data. The transformation (expressed on a basis of  $\sim 1,000$  cosine functions) is then used to propagate the manually labeled binary masks ( $Hc_i$  and  $Am_i$ ) to the MNI space. The atlases  $PA_{Hc}$  and  $PA_{Am}$  are created only once as follows, in the MRI set  $\Omega$ :

$$\begin{aligned} \forall v \in \Omega, PA_{Hc}(v) &= \frac{1}{N} \sum_{i=1}^N T_i(Hc_i)(v) \\ PA_{Am}(v) &= \frac{1}{N} \sum_{i=1}^N T_i(Am_i)(v) \end{aligned} \quad (1)$$

where  $PA_{Hc}(v)$  and  $PA_{Am}(v)$  are the probabilities that  $v$  belongs to Hc and Am.

### Initialization

The first step is to compute forward and backward transformations  $T$  and  $T^{-1}$  between native and MNI spaces. Individual atlases  $IPA_{Hc}$  and  $IPA_{Am}$  are created by back-registering the atlases  $PA_{Hc}$  and  $PA_{Am}$  using  $T^{-1}$ .  $IPA_{Hc}$  and  $IPA_{Am}$  are used to auto-

matically define left and right bounding boxes around the structures of interest, as the smallest boxes embedding the nonnull probability values in both atlases, with an extra one-voxel margin, for the left and right hemispheres, as illustrated in Figure 1a.

Atlas local registration failure in the bounding box is automatically detected and corrected when necessary. For this, it is assumed that, if  $IPA_{Hc}$  is locally misregistered, the 0.5-probability object  $\{v, IPA_{Hc}(v) \geq 0.5\}$  will cover a wider intensity range than if  $IPA_{Hc}$  is correctly registered. Details are given in Chupin et al. (2009).

The last step of the initialization procedure is to create initial objects for Hc and Am. Each probability map is pruned through thresholding while ensuring that the object is still topologically a single object (conditional pruning) to ensure obtaining one smooth and connected object corresponding to the region with maximal probability in each probability map. The two objects thus obtained are then eroded to create the initial objects (Fig. 1b).

### Deformation

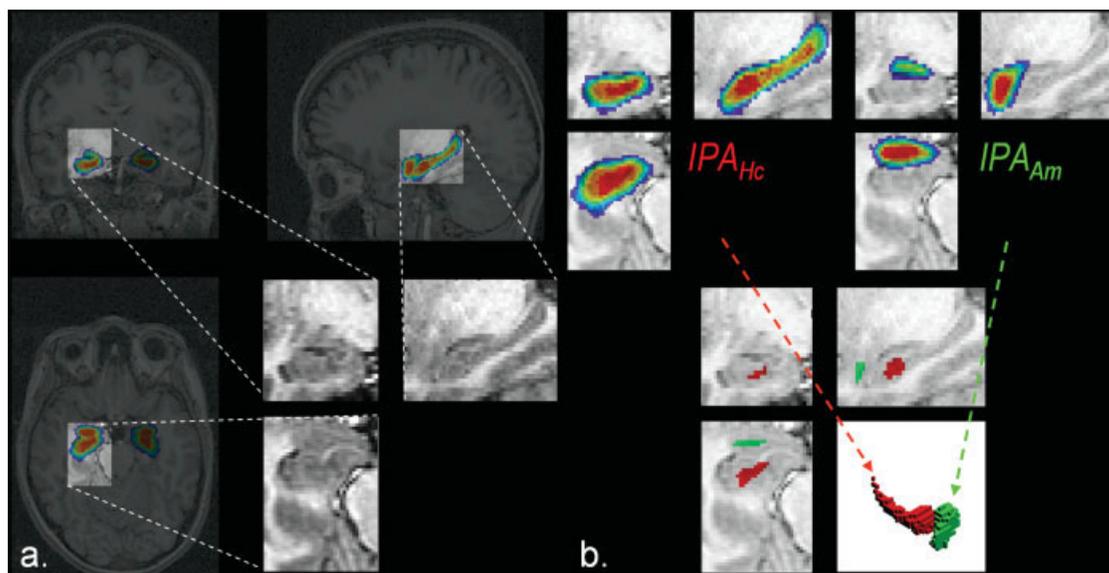
The deformation is then driven by the iterative minimization of the energy functional. At each iteration, voxel candidates are selected at the border of the deforming objects, for which reclassification will be considered; meta-regions are automatically detected during the deformation, these regions being the interface between Hc and Am and 11 families of anatomical landmarks at the border of Hc and Am (Chupin et al., 2007). The energy is then minimized on the voxel candidates through an Iterated Conditional Modes procedure. Low and high likelihood zones are defined around the anatomical landmarks from intensity and spatial local relationships, and three zones are derived from the probability maps:  $PZ^0 = \{v, IPA(v) = 0\}$ ,  $PZ^1 = \{v, IPA(v) = 1\}$ ,  $PZ^{0.75} = \{v, 0.75 \leq IPA(v) < 1\}$ . These specific features are modeled in the regularization term, comparing  $N^O(v)$  the number of  $O$ -labeled neighbors of  $v$  with a standard value  $\tilde{N}$ , with respect to a tolerance  $\sigma_I$  to prevent holes and wires:

$$E_O^I(v) = \frac{(\tilde{N} - \gamma_O^{PZ}(v)\gamma_O^{AZ}(v) \cdot (N_O(v) + \alpha_O^T(v)))^5}{\sigma^I} \quad (2)$$

where  $\alpha^T = 0$ , except for voxels detected as ‘‘tail of Hc’’ given by a local pattern ( $\alpha^T$  then increases from 0 to 16 in the bounding box posterior half).  $\gamma^{AZ} = 1$ , except for voxels in low and high likelihood zones ( $\gamma^{AZ} = 0.5$  in  $O$ -unlikely and 2 in  $O$ -likely zones).  $\gamma^{PZ} = 1$ , except for voxels in the three probability zones ( $\gamma^{PZ}(v) = 0.75$  in  $PZ^0$ ,  $\gamma^{PZ}(v) = 2$  in  $PZ^1$ ,  $\gamma^{PZ}(v) = 1.5$  in  $PZ^{0.75}$ ). These parameters constrain the deformation by decreasing the regularization energy in  $O$ -likely zones and vice versa, as detailed in Chupin et al. (2009).

### Data Analysis: Segmentation and Classification

Using the fully automatic method, we segmented the hippocampus and the amygdala in all subjects using the atlas built



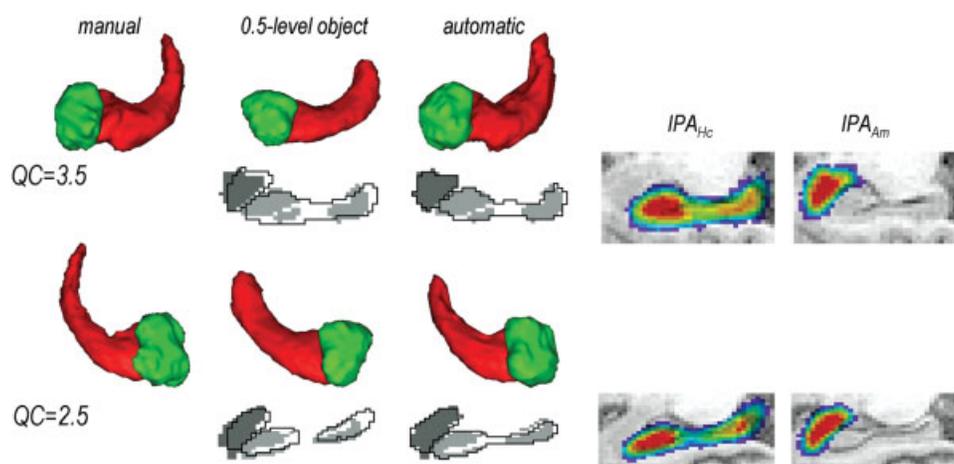
**FIGURE 1.** Initialization. (a) Extraction of the bounding boxes from the probabilistic atlases; (b) extraction of the initial objects from each probabilistic atlas through conditional pruning. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

from the 16 young controls, and the parameters of the algorithm as described in Chupin et al. (2008). The images first underwent SPM5 bias correction, which is available with the unified segmentation module. The automatic segmentations were quality-controlled for the hippocampus with a scale from 0 (unsatisfactory) to 4 (perfect), to estimate if the computed volumes were reliable. The three observers (EG, CB, and MC) were trained on a common subset of 30 subjects, to ensure coherence between the ratings, and blinded to the clinical diagnosis of the subjects.

Volumes were normalized by the total intracranial volume (TIV) computed by summing SPM5 segmentation maps of

gray matter, white matter, and cerebrospinal fluid (CSF), inside a bounding box defined in standard space to obtain a systematic inferior limit. For more robustness with respect to segmentation errors, left and right volumes were averaged. Group differences were assessed using Student's  $t$ -test.

For the classification of patients versus controls, each participant was assigned to the closest group as follows. Robust estimates of classification rate, sensitivity, and specificity were computed with a bootstrap approach for training set selection. In this procedure, we drew without replacement  $\sim 75\%$  of each group to obtain a training set. On this training set, we estimated the mean normalized hippocampal volume for each



**FIGURE 2.** 3D-renderings of manual, atlas-derived and automatic segmentations, overlap between segmentations (manual segmentations in shades of gray) and probabilistic atlases, for the best and worst results among eight patients with AD. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

TABLE 1.

## Experiment 1: Study Population

Subjects	Number	Age	MMS	Centres
CN	30	74 ± 4 (65–85)	29 ± 1 (26–30)	13
AD	29	77 ± 7 (56–89)	23 ± 2 (20–26)	15

CN, cognitively normal subjects; AD, patients with AD. Values: average ± standard deviation (range), MMS, mini-mental state.

group. Each participant in the remaining 25% was then assigned to the group which mean was closest to the volume of this participant. The procedure was repeated 5,000 times.

## EXPERIMENTS AND RESULTS

### Evaluation of segmentation accuracy with respect to manual tracing

Segmentation accuracy for the hippocampus was evaluated by comparing automatic segmentation (S) with a reference manual segmentation (R) with two quantitative indices:  $RV(S, R) = 2|V_S - V_R| / (V_S + V_R)$ , the error on volumes and  $DO(S, R) = 2V_{S \cap R} / (V_S + V_R)$ , the Dice overlap. We compared the performance of the fully automatic approach with an “atlas-derived” segmentation given by the 0.5-level probability object.

Complete results are given in Chupin et al. (2008). In summary, for the 16 young controls used to create the atlas, for the automatic segmentation with a leave-one-out strategy,  $RV = 6\%$  and  $DO = 87\%$ . For eight patients with AD, fully described in Chupin et al. (2007),  $RV = 7\%$  and  $DO = 86\%$ . For the same patients, if we consider the objects derived from the registered atlases,  $RV = 27\%$  and  $DO = 68\%$ . Two examples of segmentation and atlas registration are given in Figure 2. The quality control evaluated the segmentation as correct ( $\geq 3$ ) for 13 Hc (81%), acceptable ( $\geq 2$ ) for 3 Hc (19%) and unsatisfactory ( $< 2$ ) for none.

### Segmentation and classification of subjects from the ADNI database

To assess whether our automatic segmentation method can provide a biomarker for AD, we tested the ability of Hc volumes to discriminate between patients with AD, patients with MCI and elderly controls.

#### Subjects

Data were obtained from the ADNI database ([www.loni.ucla.edu/ADNI](http://www.loni.ucla.edu/ADNI)). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical com-

panies, and nonprofit organizations, as a \$60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The principal investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55–90, to participate in the research, ~200 cognitively normal (CN) older individuals to be followed for 3 yrs, 400 people with MCI to be followed for 3 yrs, and 200 people with early AD to be followed for 2 yrs. For up-to-date information see [www.adni-info.org](http://www.adni-info.org).

#### Experiments

Three different experiments were considered, to evaluate the influence of preprocessing, normalization by TIV, age groups, and segmentation method on classification results. MRI acquisition was done according to the ADNI acquisition protocol (Jack et al., 2008). ADNI images with B1 and “grad warp” corrections were used, as they seemed to correspond to the best quality that could be obtained in clinical routine.

**Experiment 1: Feasibility and segmentation quality.** Sixty subjects (30 patients with AD and 30 CN subjects) were randomly selected from the ADNI database (Table 1). Thirty patients with AD were initially selected, but one reverted to CN during follow-up and was therefore excluded from our analysis. The 59 images came from 18 centers, resulting in 58 images on 1.5T scanners (GE and SIEMENS) and one image on a 3T scanner (SIEMENS). Images were selected at random among the available scanning sessions (baseline or screening) for each subject; only volumes derived from images with preprocessing and normalized by TIV were used.

TABLE 2.

## Experiment 2: Study Population

Subjects	Number	Age	MMS	Centres
CN	139	76 ± 5 (60–90)	29 ± 1 (26–30)	37
AD	124	76 ± 7 (55–91)	23 ± 2 (18–27)	39

CN, cognitively normal subjects; AD, patients with AD. Values: average ± standard deviation (range), MMS, mini-mental state.

TABLE 3.

Experiment 3: Study Population

Subjects	Number	Number with 18 months follow up	Age	MMS	Centres
CN	166	162	76 ± 5 (60–90)	29 ± 1 (25–30)	40
AD	145	137	76 ± 8 (55–91)	23 ± 2 (18–27)	39
MCI	294	210	75 ± 7 (55–90)	27 ± 2 (23–30)	40
MCI <sup>nc</sup>		134	74 ± 7 (58–88)	27 ± 2 (24–30)	35
MCI <sup>c</sup>		76	75 ± 7 (55–88)	26 ± 2 (23–30)	28

CN, cognitively normal subjects; AD, patients with AD; MCI, patients with mild cognitive impairment; nc, non converter at 18 months; c, converter at 18 months; MMS, mini-mental state.

**Experiment 2: Effect of preprocessing, group age and normalization by TIV.** AD and CN subjects (124 patients with AD and 139 CN subjects) with and without the preprocessing steps available at the time of the study were selected from the ADNI database (Table 2). The 263 images were chosen at random from all the available scanning sessions (baseline or screening) for each subject. They came from 41 centers (262 images on 1.5T scanners (GE and SIEMENS) and one image on a 3T scanner (SIEMENS)).

**Experiment 3: Segmentation of the full database with preprocessing.** We selected all the subjects for whom preprocessed images were available at the time of the study. As a result, 605 subjects (145 patients with AD, 166 CN subjects, and 294 patients with MCI) were selected (Table 3). For each subject, we used the MRI scan from the baseline visit when available and from the screening visit otherwise. We only used images acquired at 1.5T. The 605 images came from 41 centers. Among the 210 patients with MCI for whom 18-months follow up was available, 76 converted to AD before 18 months.

We also compared classification results with hippocampal volumes given by our method and those available in ADNI. Semiautomated hippocampal volumetry was carried out using a commercially available high-dimensional brain mapping tool (Medtronic Surgical Navigation Technologies, Louisville, CO), that has previously been validated and compared to manual tracing (Hsu et al., 2002). SNT hippocampal volume is measured first by placing manually global landmarks and 22 local landmarks per hippocampus on each MRI scan. Second, fluid image transformation is used to match each brain to a template brain. Note that the segmentation was also manually edited if the result was not satisfactory.

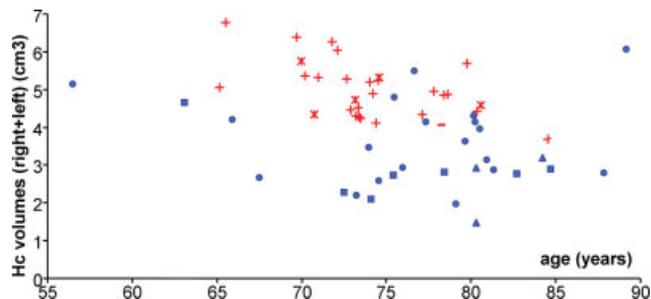


FIGURE 3. Experiment 1: automatically computed (left + right) volumes for the CN (red + for correct segmentations, \* for acceptable segmentations, and – for the unsatisfactory segmentation) and AD (blue discs for correct segmentations, squares for acceptable segmentations, and triangles for unsatisfactory segmentations) as a function of age. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

Results

**Experiment 1.** For the 29 patients with AD, the segmentation proved correct ( $\geq 3$ ) for 19 patients (66%), acceptable ( $\geq 2$ ) for seven patients (24%), and unsatisfactory ( $< 2$ ) for three patients (10%). For the 30 elderly controls, the segmentation proved correct for 24 controls (80%), acceptable for five controls (17%), and not satisfactory for one control (3%). The volumes obtained with the automatic segmentation of the hippocampus (left + right) are displayed in Figure 3. Note that the segmentations which were considered as unsatisfactory still give volumes which are coherent with the classification, and the patients with AD who are likely to be misclassified in fact correspond to reliable segmentations.

The results of group analysis and individual classification using all the segmented volumes are presented in Table 4 for Hc. We also found a significant group difference for Am between AD and controls (0.95 vs. 1.11, 14% atrophy,  $P < 0.05$ ), but using Am volume with Hc volume in a linear SVM analysis did not improve classification results.

**Experiment 2.** For the second experiment, the goal was to compare the classification performances between several conditions: with and without TIV normalization, with and without preprocessing and in a reduced age range (between 70- and 80-yr-old, with 99 CN and 60 AD subjects). We chose not to keep subjects under 70-yr-old, because the populations were highly unbalanced (more AD than CN subjects), and above

TABLE 4.

Experiment 1: Results

Hc volume (cm <sup>3</sup> )	1.69 vs. 2.49	Class. rate	82%
Mean vol. reduction	–32%	Sensitivity	75%
Statistical significance	$P < 0.001$	Specificity	89%

Two left columns: group comparisons of Hc volumes. Two right columns: classification rate, sensitivity, and specificity for classification between AD and CN.

TABLE 5.

## Experiment 2: Results

	No prepro – no TIV	Prepro – no TIV	Prepro – TIV	Prepro, no TIV 70–80
Hc volume (cm <sup>3</sup> )	1.84 vs. 2.45	1.77 vs. 2.43	1.76 vs. 2.49	1.81 vs. 2.47
Mean vol. reduction	–25%**	–27%**	–29%**	–27%**
Class. rate	75%	78%	78%	82%
Sensitivity	73%	76%	77%	81%
Specificity	76%	80%	78%	84%

Top two rows: group comparisons of Hc volumes. Bottom three rows: classification rate, sensitivity, and specificity for classification between AD and CN (see text for details).

\*\* $P < 0.001$ .

80-yr-old, because very elderly CN subjects are highly heterogeneous.

Some dependence with age can be observed in all three conditions. Group analyses indicate significant differences in all three conditions, as shown in Table 5, and within the 70–80 year age range. Classification results are better with preprocessing; they are equivalent with and without TIV normalization. Restricting the classification problem on a 70–80 year age group increases the classification rate to 82%.

**Experiment 3.** Among the patients with AD, the segmentation proved correct ( $\geq 3$ ) for 69 (48%), acceptable ( $\geq 2$ ) for 57

(39%) and unsatisfactory ( $< 2$ ) for 19 (13%) patients. Among the MCI patients, the segmentation proved correct for 185 (63%), acceptable for 92 (31%), and unsatisfactory for 17 (6%) patients. Among the CN, the segmentation proved correct for 127 (77%), acceptable for 37 (22%), and unsatisfactory for 2 (1%) subjects. Volumes are shown in Figure 4, as a function of age, for the pairs of groups for which classification is considered.

Group analyses for the whole cohort are displayed in the first rows of Table 6. Quality control results show that hippocampal segmentation seems to perform better for CN than for AD; on the other hand, segmentation quality is more difficult to assess for highly atrophic structures. Hippocampal volume significantly differs between AD, MCI, or MCI converters and CN, and, more importantly, hippocampal volume significantly differs between MCI converters and nonconverters, with a 14% atrophy between the groups. Note that MMS difference is very small between the two MCI groups, whereas hippocampal volume difference is large and highly significant. Classification results are coherent with those of experiment 2 for AD versus CN. Note that the classification results of MCI converters versus CN are only slightly lower than for AD versus CN, and coherent with the volume difference.

As in experiment 2, classification was also studied for a smaller group of subjects between 70- and 80-yr-old. This resulted in 67 AD, 143 MCI, and 123 CN, and in 69 AD, 42 MCI converters, 65 MCI nonconverters, and 119 CN with 18 months follow up. Group analyses are displayed in the first four rows of Table 7 and show the same patterns as on the

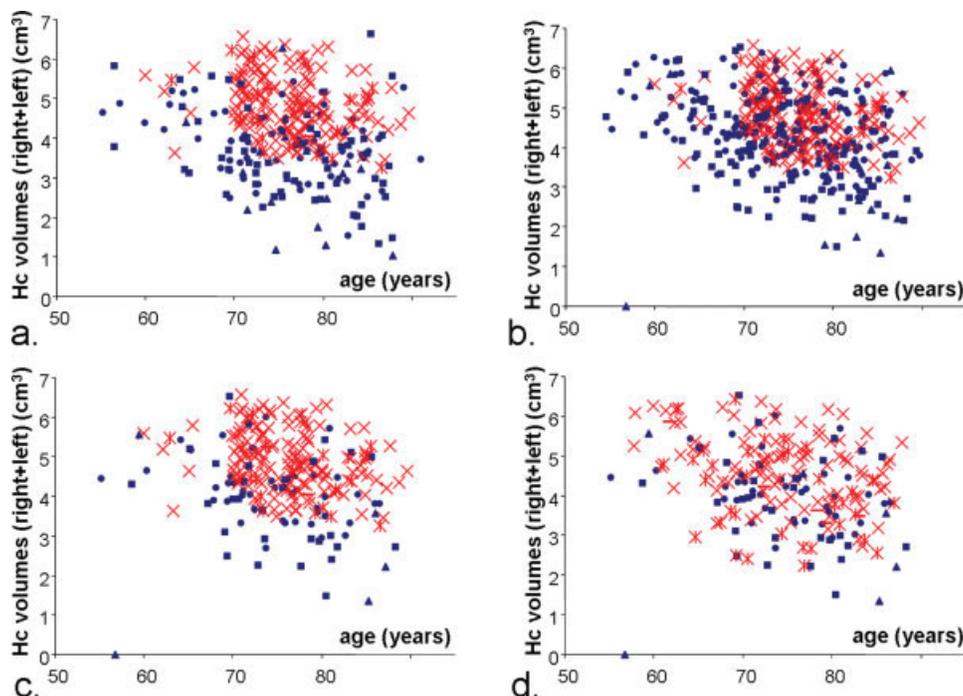


FIGURE 4. Experiment 3: automatically computed volumes corresponding to the classification experiments: (a) CN (×) vs AD, (b) CN (×) vs MCI, (c) CN at 18 months (×) vs MCI converting at 18 months, (d) MCI not converting (×) vs MCI converting. {x,

●} for correct segmentations, {×, ■} for acceptable segmentations, and {–, ▲} for unsatisfactory segmentations. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

TABLE 6.

Experiment 3: Results

	AD vs. CN	MCI vs. CN	MCIc vs. CN	MCIc vs. MCIc
MMS	-20%**	-8%**	-9%**	-3%*
Segmentation QC	2.7 vs. 3.1**	2.9 vs. 3.1**	2.7 vs. 3.1**	2.7 vs. 2.9
Hc volume (cm <sup>3</sup> )	1.83 vs. 2.43	2.14 vs. 2.43	1.95 vs. 2.43	1.95 vs. 2.28
Mean vol. reduction	-25%**	-12%**	-20%**	-14%**
Class. rate	76%	61%	71%	64%
Sensitivity	75%	61%	67%	60%
Specificity	77%	61%	72%	65%
Threshold (cm <sup>3</sup> )	2.13	2.29	2.19	2.11

Top four rows: group comparisons. Bottom four rows: classification rate, sensitivity, specificity, and threshold for classification (see text for details).

\*P < 0.01.

\*\*P < 0.001.

whole cohort. Segmentation quality differences were reduced, but average volume reductions were preserved. Note that the populations are not matched for age, sex, and scanner. Classification results all appear better than on the complete cohort; accuracy was 80% for AD versus CN, 74% for MCI converters versus CN, and 67% for MCI converters versus MCI nonconverters.

Among the subjects we used, SNT volumes were available for 122 AD and 128 CN, and 186 MCI with 18 months follow-up (65 converters and 121 nonconverters). Classification

TABLE 7.

Experiment 3: Results for Age Group (70–80)

	AD vs. CN	MCI vs. CN	MCIc vs. CN	MCIc vs. MCIc
MMS	-20%**	-7%**	-10%**	-4%*
Segmentation QC	2.7 vs. 3.1**	2.9 vs. 3.1	2.9 vs. 3.1**	2.9 vs. 3.0
Hc volume (cm <sup>3</sup> )	1.80 vs. 2.46	2.16 vs. 2.46	1.95 vs. 2.47	1.95 vs. 2.28
Mean vol. reduction	-27%**	-12%**	-21%**	-14%*
Class. rate	80%	63%	74%	67%
Sensitivity	80%	63%	75%	65%
Specificity	79%	63%	74%	68%
Threshold (cm <sup>3</sup> )	2.13	2.31	2.22	2.12

Top four rows: group comparisons. Bottom four rows: classification rate, sensitivity, specificity, and threshold for classification (see text for details).

\*P < 0.01.

\*\*P < 0.001.

TABLE 8.

Experiment 3: Comparison with ADNI volumes

	SACHA volumes		SNT volumes	
	AD vs. CN	MCIc vs. MCIc	AD vs. CN	MCIc vs. MCIc
Hc volume (cm <sup>3</sup> )	1.85 vs. 2.43	1.91 vs. 2.28	1.59 vs. 2.12	1.69 vs. 1.88
Mean vol. reduction	-24%**	-16%**	-25%**	-10%**
Class. rate	76%	65%	80%	65%
Sensitivity	74%	63%	79%	67%
Specificity	78%	67%	81%	64%
Threshold (cm <sup>3</sup> )	2.14	2.10	1.85	1.80

Top two rows: group comparisons. Bottom four rows: classification rate, sensitivity, specificity, and threshold for classification) for our volumes (SACHA) and the volumes given by ADNI (SNT) (see text for details).

\*\*P < 0.001.

results on this population for volumes derived from our method and SNT volumes are given in Table 8. Group analysis is similar for AD versus CN, whereas our volumes tend to show a larger difference between converter and nonconverter MCI. Regarding classification, results are similar for converter versus nonconverter MCI, whereas, for AD versus CN, the classification is less good for our volumes (76% compared to 80%).

**DISCUSSION**

We have demonstrated in this article that the fully automatic hippocampus segmentation method presented here is accurate for data coming from patients and normal subjects acquired on a variety of MRI platforms, with a systematic qualitative evaluation process (the segmentation proved correct in 63%, acceptable in 31%, and not satisfactory in 6% of the cases). It has also proven its usefulness in discriminating between CN subjects, patients with MCI, and patients with AD in a setting which corresponds better to clinical routine. This study confirms the results that were shown in Chupin et al. (2008), while being applied to more realistic datasets. Furthermore, the segmentation process is fast (15 min, including 10 for the registration and five for bilateral segmentation) and is implemented as the SACHA module in a user-friendly environment (<http://brainvisa.info>).

Most importantly, no atlas modification was necessary, compared to Chupin et al. (2008). The hybrid anatomical and probabilistic priors make the segmentation more robust to pathology and acquisition parameters than the semiautomatic method (Chupin et al., 2007). Furthermore, the partial integra-

tion of probabilistic maps as a constraint in the deformation process makes it more robust to pathology than methods that rely more strongly on a single atlas. In fact, it was previously demonstrated that segmentation based on the registration of a single subject atlas does not perform satisfactorily when the atlas does not belong to the same disease category as the subject (Carmichael et al., 2005).

Validation studies on the segmentation of the hippocampus in patients with AD are limited and difficult to compare because of different patient samples and evaluation strategies (Crum et al., 2001; Hsu et al., 2002). Recently, a method based on the registration and segmentation module of SPM5 (Firbank et al., 2008) was evaluated on nine elderly controls with an RV of 5% and a DO of 74%; and nine patients with AD, with an RV of 15% and a DO of 67%. Another method, based on finding the best match among a library of templates (Barnes et al., 2008), with a refinement step based on intensity, was evaluated on 19 elderly controls, with a DO of 82%, and 36 patients with AD, with a DO of 84%. Finally, a method based on statistically learned image features was evaluated on 21 subjects (seven controls, seven patients with MCI, and seven patients with AD) from the ADNI database (Morra et al., 2008), and a DO value of 83% were reported.

Using fully automatic volumetry of the hippocampus, we were able to discriminate patients with AD from controls with 76–80% accuracy, in this study. This remains in line with previous results based on manual segmentation which report accuracy between 82 and 90% for AD, e.g., Frisoni et al. (1999), Xu et al. (2000). As for automatic methods, very few studies investigated the classification of individual patients. Fischl et al. (2002) detected significant group differences in hippocampal volume but did not investigate classification of individual participants. Using both volume and shape features, Csernansky et al. (2000) reported a sensitivity of 83% and a specificity of 78%. The accuracy that we report for MCI (61–63% for the whole group and 71–74% for the patients converting to AD before 18 months) is also comparable to that obtained using manual segmentation (between 60 and 74%, e.g., Xu et al. (2000) and Pennanen et al. (2004)). Compared to our results reported in Chupin et al. (2008) and Colliot et al. (2008) (87% for AD vs. controls, 74% for MCI vs. controls), the classification accuracies obtained here on the ADNI database are slightly lower. This can be explained by several factors. First, ADNI is a multicenter database (41 centers, different voxel sizes and acquisition parameters) whereas the data in our previous study came from a single scanner. Moreover, the population includes a large number of subjects with vascular lesions, thus being closer to real life datasets. The systematic quality control procedure allowed establishing that the cases which were not consistent for the classification did not always correspond to cases which were not satisfactory for the segmentation. Classification results without unsatisfactory segmentations did not prove any better, which is likely to be due to the intrinsic variability of the hippocampal volume among the study population.

We compared the classification accuracy derived from our method and derived from SNT volumes. Accuracy was similar for converter versus nonconverter MCI, while SNT volumes were slightly more discriminative for AD versus CN (80–76%). Note that our approach is fully automatic and fast, while the SNT approach requires the placement of more than 22 landmarks per hippocampus, and the unsatisfactory segmentations were manually edited.

In the second experiment, we have also shown that the preprocessing steps that we kept from those available in ADNI have an effect on the segmentation and/or the classification results. These correction steps appear useful in the present study. Furthermore, we have shown that the normalization by total intracranial volume does not improve classification results; this may be due to errors in the TIV values (due to CSF segmentation), or to the absence of linear relationship between the volume of the hippocampus and TIV.

Groups are not matched for age and gender, and the database includes more controls than patients older than 80-yr-old; we have shown that age impacts on the classification results. Furthermore, controls over 80 will correspond to a far more variable population than younger controls, as most of them will have cerebral atrophy and ventricular enlargement, and are likely to have incipient dementias. Age groups should be taken into account when devising a diagnostic tool; careful considerations should be given to the age range.

Finally, results when comparing MCI converters and nonconverters show that the hippocampus conveys useful information for designing prognostic tools. Nevertheless, hippocampal volume is not yet sufficient for a complete discrimination of the two populations. Shape analysis and/or classification methods using both local and global information may give complementary information and improve classification reliability.

## Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and through generous contributions from the following: Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck & Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories, and the Institute for the Study of Aging, with participation from the U.S. Food and Drug Administration. Industry partnerships are coordinated through the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles.

## REFERENCES

- Ashburner J, Friston KJ. 2005. Unified segmentation. *Neuroimage* 26: 839–851.
- Barnes J, Foster J, Boyes RG, Pepple T, Moore EK, Schott JM, Frost C. 2008. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *Neuroimage* 40: 1655–1671.
- Bloch I, Colliot O, Camara O, Géraud T. 2005. Fusion of spatial relationships for guiding recognition, example of brain structure recognition in 3D MRI. *Pattern Recognit Lett* 26:449–457.
- Braak H, Braak E. 1995. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging* 16:271–278.
- Carmichael OT, Aizenstein HA, Davis SW, Becker JT, Thompson PM, Meltzer C, Liu Y. 2005. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *Neuroimage* 27:979–990.
- Chupin M, Mukuna-Bantumbakulu AR, Hasboun D, Bardinet E, Baillet S, Kinkingnéhun S, Lemieux L, Dubois B, Garnero L. 2007. Automated segmentation of the hippocampus and the amygdala driven by competition and anatomical priors: Method and validation on healthy subjects and patients with Alzheimer's disease. *Neuroimage* 34:996–1019.
- Chupin M, Chételat G, Lemieux L, Dubois B, Garnero L, Benali H, Eustache F, Lehericy S, Desgranges B, Colliot O. 2008. Fully automatic hippocampus segmentation discriminates between early Alzheimer's disease and normal aging. In *Proceedings of IEEE ISBI*, 14–17 May, Paris, pp 97–100.
- Chupin M, Hammers A, Bardinet E, Colliot O, Liu RSN, Duncan JS, Garnero L, Lemieux L. 2009. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: Method and validation. *Neuroimage* 46:749–761.
- Colliot O, Chételat G, Chupin M, Desgranges B, Magnin B, Benali H, Dubois B, Garnero L, Eustache F, Lehericy S. 2008. Discrimination of Alzheimer's disease, mild cognitive impairment and normal aging using automated segmentation of the hippocampus. *Radiology* 248:194–201.
- Crum WR, Scahill RI, Fox NC. 2001. Automated hippocampal segmentation by regional fluid registration of serial MRI: Validation and application in Alzheimer's disease. *Neuroimage* 13: 847–855.
- Csernansky JG, Wang L, Joshi S, Miller JP, Gado M, Kido D, McKeel D, Morris JC, Miller MI. 2000. Early DAT is distinguished from aging by highdimensional mapping of the hippocampus. *Neurology* 55:1636–1643.
- Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P. 2007. Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* 6:734–746.
- Duchesne S, Pruessner JC, Collins DL. 2002. Appearance-based segmentation of medial temporal lobe structures. *Neuroimage* 17: 515–531.
- Firbank MJ, Barber R, Burton EJ, O'Brien JT. 2008. Validation of a fully automated hippocampal segmentation method on patients with dementia. *Hum Brain Mapp* 29:1442–1449.
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. 2002. Whole brain segmentation: Automated labelling of neuroanatomical structures in the human brain. *Neuron* 33:341–355.
- Frisoni GB, Laakso MP, 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.
- Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* 33:115–126.
- Hsu YY, Schuff N, Du AT, Mark K, Zhu X, Hardin D, Weiner MW. 2002. Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. *J Magn Reson Imag* 16:305–310.
- Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW, Study A. 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imag* 27:685–691.
- Kelemen A, Szekely G, Gerig G. 1999. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE Trans Med Imag* 18: 828–839.
- Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR, Schuff N, Weiner MW, Thompson PM. 2008. Mapping hippocampal degeneration in 400 subjects with a novel automated segmentation approach. In *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 14–17 May, Paris, pp 336–339.
- Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hänninen T, Laakso MP, Hallikainen M, Vanhanen M, Nissinen A, Helkala EL, Vainio P, Vanninen R, Partanen K, Soininen H. 2004. Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol Aging* 25:303–310.
- Shen D, Moffat S, Resnick SM, Davatzikos C. 2002. Measuring size and shape of the hippocampus in MR images using a deformable shape model. *Neuroimage* 15:422–434.
- Xu Y, Jack CR, O'Brien PC, Kokmen E, Smith GE, Ivnik RJ, Boeven BF, Tangalos RG, Petersen RC. 2000. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. *Neurology* 54: 1760–1767.