Automated Segmentation of the Human Hippocampus Along Its Longitudinal Axis

Garikoitz Lerma-Usabiaga,^{1,*} Juan Eugenio Iglesias,¹ Ricardo Insausti,²
Douglas N. Greve,³ and Pedro M. Paz-Alonso¹

¹BCBL. Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain

²Human Neuroanatomy Laboratory, Department of Health Sciences and Regional Center for Biomedical Research, School of Medicine, University of Castilla-La Mancha, Albacete, Spain

³Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts, US

Abstract: The human hippocampal formation is a crucial brain structure for memory and cognitive function that is closely related to other subcortical and cortical brain regions. Recent neuroimaging studies have revealed differences along the hippocampal longitudinal axis in terms of structure, connectivity, and function, stressing the importance of improving the reliability of the available segmentation methods that are typically used to divide the hippocampus into its anterior and posterior parts. However, current segmentation conventions present two main sources of variability related to manual operations intended to correct in-scanner head position across subjects and the selection of dividing planes along the longitudinal axis. Here, our aim was twofold: (1) to characterize inter- and intra-rater variability associated with these manual operations and compare manual (landmark based) and automatic (percentage based) hippocampal anterior–posterior segmentation procedures; and (2) to propose and test automated rotation methods based on approximating the hippocampal longitudinal axis to a straight line (estimated with principal component analysis, PCA) or a quadratic Bézier curve (fitted with numerical methods); as well as an automated anterior–posterior hippocampal segmentation procedure based on the percentage-based method. Our results reveal that automated rotation and segmentation procedures, used in combination or independently, minimize inconsistencies generated by the accumulation of manual operations while providing higher statistical power to detect well-known effects. A Matlab-based implementation of these procedures is made publicly available to the research community. *Hum Brain Mapp* 37:3353–3367, 2016. © 2016 Wiley Periodicals, Inc.

Key words: hippocampus; magnetic resonance imaging; neuroanatomy; principal component analysis; reproducibility of results; segmentation

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: EMBO Short-Term Fellowship (158-2015) to G.L.-U.; Contract grant sponsor: Spanish Ministry of Economy and Competitiveness (MINECO; to J.E.I. and G.L.-U.); Contract grant number: TEC2014-51882-P; Contract grant sponsor: European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 654911 (project "THALAMODEL"; to J.E.I.); Contract grant sponsor: "Ramon y Cajal" Fellowship; Contract grant numbers: PSI2012-32093, PSI2015-65696, and RYC-2014-15440; Contract grant spon-

sor: MINECO (Severo Ochoa programme to P.M.P.-A.; SEV-2015-049)

*Correspondence to: Garikoitz Lerma-Usabiaga, Basque Center on Cognition Brain and Language, Paseo Mikeletegi 69, 2nd floor, 20009, Donostia-San Sebastián, Gipuzkoa, Spain. E-mail: g.lerma@bcbi.eu

Received for publication 26 January 2016; Revised 24 March 2016; Accepted 25 April 2016.

DOI: 10.1002/hbm.23245

Published online 9 May 2016 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

The human hippocampal formation has attracted the interest of researchers across different neuroscience fields, especially in the last decades when its critical involvement in the formation of new memories [Scoville and Milner, 2000], learning [Clark and Squire, 1998], and spatial navigation [Fyhn et al., 2004] become uncovered.

In line with prior anatomical research in rodents showing distinct input and output connections of different parts of the hippocampus [e.g., Swanson and Cowan, 1977], recent neuroimaging studies with humans have also evinced differences along the longitudinal axis of the hippocampus in terms of structure, connectivity, and function [Persson et al., 2014; Poppenk et al., 2013; Poppenk and Moscovitch, 2011; Robinson et al., 2015; Strange et al., 2014]. It is worth noticing that significant effort has been dedicated to the creation of automated hippocampal sub-field segmentation techniques, as those available in FreeSurfer [Iglesias et al., 2015; Van Leemput et al., 2009] or ASHS [Yushkevich et al., 2015]. Nevertheless, differential structural and functional contributions of anterior hippocampus (aHPC) and posterior hippocampus (pHPC) have been reported across several cognitive domains, such as episodic memory encoding and retrieval, spatial memory, emotion, and motivation [Poppenk et al., 2013]. Recent debates and theoretical accounts of hippocampal *y*-axis (longitudinal) specialization rely on neurobiological evidence from gene expression, anatomical, and electrophysiological recording studies in animals and humans that reveal multiple domains and a gradient organization along the hippocampal longitudinal axis [Strange et al., 2014]. Thus, improving the precision and reliability of the available segmentation methods typically used to divide aHPC and pHPC segments constitutes a critical step to further characterize hippocampal longitudinal axis specializations in *in vivo* magnetic resonance imaging (MRI) studies.

Current segmentation conventions are typically based on a two-step procedure: head position correction and choice for the dividing coronal plane. In the first step, the scanner head position is corrected by rotating the MRI scan to equate its orientation across subjects; this rotation determines the angle of the cutting plane used in the second step. The rotation, which includes a horizontal and a vertical component, is necessary because the head position differs substantially from subject to subject (see Fig. 1A). If no horizontal correction is performed, the coronal plane will cut the major axes of the left and right hippocampi at supplementary angles that could bias subsequent analyses (Fig. 1B). Likewise, vertical rotation is necessary to ensure that the cutting plane is perpendicular to the major axis of the hippocampus in all subjects. In the second step, a coronal plane is chosen to segment the hippocampus into anterior and posterior regions. The choice of a dividing coronal plane resolves the cutting point along the hippocampal longitudinal axis.

Figure 1.

Manual corrections of in-scanner head position before hippocampal longitudinal segmentation. **A)** In-scanner head position differs substantially from subject to subject. **B)** Inconsistencies in the correction of the in-scanner head position yields problems associated with the vertical and horizontal components of the cutting plane to be used in the posterior segmentation: the hippocampi will be divided with a different angle in the sagittal plane and, consequently, the left and the right hippocampi will be divided with a different angle. **C)** Current manual rotation procedures correct in-scanner head position aligning the brain with the interhemispheric fissure first and then either aligning the brain with the AC–PC line or horizontally positioning one of the hippocampi (usually the left).

This procedure presents two main sources of variability, related to how the in-scanner head position is corrected and equated across subjects, and how the segmenting plane along the longitudinal axis is chosen. In-scanner head position is commonly corrected by manually aligning the brain for roll, pitch, and yaw either along the interhemispheric fissure, anterior–posterior commissure (AC–PC) line orbits [Haller et al., 1997; Killiany et al., 1993; Malykhin et al., 2007; Pantel et al., 2000; Pruessner et al., 2000], or along the interhemispheric fissure, positioning one of the hippocampus (usually the left) in a horizontal position [Bartzokis et al., 1998; Convit et al., 1997; DeToledo-Morrell et al., 2004; Jack, 1994; Lehericy et al., 1994; Soininen et al., 1994; Watson et al., 1992] (see Fig. 1C). Regardless of the alignment criterion, manual rotation suffers from inter- and intra-rater variability, introducing inaccuracies and biases that carry over subsequent analysis of the data. Moreover, the variability in the manual rotations may negatively impact the reproducibility of the results

across laboratories and, subsequently, the validity of the observed results.

Regarding how the dividing plane along the hippocampal longitudinal axis is chosen, discrepancies across laboratories exist in regard to the number of segments and to the criteria used to choose the landmarks that define the position of the planes. In this vein, it is important to note that any aHPC/pHPC landmark-based separation is arbitrary, as there is not enough evidence of a dichotomous hippocampal boundary [Frankó et al., 2014]. Classical anatomical approaches segment the hippocampus into head, body, and tail [e.g., Bartzokis et al., 1998; Duvernoy, 2005; Hackert et al., 2002; Insausti and Amaral, 2012]. In contrast, based on the claim that there is not a clear anatomical landmark to unambiguously identify the point of transition between body and tail [Chupin et al., 2009; Frankó et al., 2014], other approaches divide the hippocampus only into aHPC and pHPC segments [Poppenk and Moscovitch, 2011].

There are two main methods to determine the landmark to divide the human hippocampus into aHPC and pHPC segments (see Fig. 2 and Poppenk et al., 2013 for a review): (1) Manual landmark-based segmentation, in which the last slice where the uncus apex (gyrus intralimbicus) is visible is manually chosen for aHPC/pHPC segmentation (Fig. 2A); and, (2) Percentage-based segmentation, where the total length of the hippocampal long axis is divided based on a percentage that approximates the manual landmark¹ (Fig. 2B). There is an additional method, known as Talairach/MNI coordinate-based segmentation in which brain images are spatially transformed into either of these stereotaxic spaces and then a specific coordinate along the longitudinal *y*-axis is selected as the division landmark (i.e., $y = -20$ mm for Talairach, $y = -21$ mm for MNI). Because this procedure is not affected by the main experimental manipulations of interest examined in the present work (i.e., manual rotations and manual landmark placement), it is not included in the analyses featured in the present report.²

The present study is aimed at quantifying the variability introduced by manual segmentation procedures, and proposes automated methods that ameliorate the problems associated to them. More specifically, the contribution of

the study is twofold. First, we quantify the inter- and intra-rater variability associated with manual interventions in the rotation and landmark selection procedures. To this end, two independent trained raters were asked to manually rotate and annotate the uncus apex in 100 samples at two different and randomized time points. Second, we present and evaluate automated methods to: (1) correct the in-scanner head position based on estimating the hippocampal longitudinal axis, and (2) segment the aHPC/pHPC based on the percentage-based segmentation procedure. Two automated rotation variants are presented: one that models the axis with a straight line (computed with PCA) and one that uses a quadratic Bézier curve (fitted with numerical methods). Due to its automated nature, these approaches can avoid the variability associated with manual rotations of the in-scanner brain position and the manual landmark selection. Here, the differential performance of the automated and manual methods is assessed through the ability of the segmentations to reproduce well-known effects like the left-right head asymmetry [Woolard and Heckers, 2012]. The implementation of these automated methods is made available to the research community as a set of Matlab scripts, available in Supporting Information and <https://github.com/garikowitz/hippovol>.

MATERIALS AND METHODS

Subjects

MRI data from 100 healthy young adults was used in this study. The subjects (age 26.12 ± 6.50 years; 55 females) were selected from the BCBL's internal database. All subjects were right-handed and had no history of psychiatric, neurological, attention, or learning disorders. They were recruited to participate in different previous studies, and all of them gave written informed consent in compliance with the ethical regulations established by the BCBL Ethics Committee and the guidelines of the Helsinki Declaration.

MRI Acquisition

The participants were scanned in a 3T Siemens TRIO whole-body MRI scanner (Siemens Medical Solutions, Erlangen, Germany), using a 32-channel head coil. Structural T1-weighted images were acquired with a MPRAGE sequence with TR = 2,300 ms, TE = 2.97 ms, flip angle = 9° , FOV = 256 mm, and voxel size = 1 mm^3 (isotropic).

Manual Operations

Two trained raters (A, B) performed two manual operations on the MRI scans. First, they manually rotated the structural scans aligning them vertically to the interhemispheric fissure in the coronal and axial planes to correct for yaw and roll, and aligning the longitudinal axis of the hippocampus horizontally separately for the left and right

¹Hackert et al. [2002] used a fixed 35% percentage of the hippocampal length to separate aHPC/pHPC. There is, however, less agreement about the to-be-followed practices to separate pHPC into body and tail. Based on histological examinations, Frankó et al. [2014] argue that there are no clear structural divisions to justify a distinct hippocampal body-tail segmentation landmark and that segmenting the pHPC at 50% of its length is as valid as any other criterion. Although here we focus on aHPC/pHPC segmentation, the software tool accompanying this manuscript allows to further segment the pHPC into body and tail.

²The Talairach/MNI coordinate-based segmentation procedure has been implemented in the Matlab tool that is released along the publication of this work for those researchers interested in it.

Figure 2.

Landmark- and percentage-based strategies for longitudinal hippocampal segmentation along its longitudinal axis [Poppenk et al., 2013]. **A)** In landmark-based segmentation, raters identify and annotate the last visible slice of the uncus apex to divide the hippocampal head or aHPC from pHPC. **B)** In percentage-based segmentation, a fixed percentage of the length of the total hip-

pocampus is chosen to separate the aHPC from the pHPC (e.g., Hackert et al., 2002; image B presents an image of a real hippocampal head over the original schema from Hackert et al.). Reproduced with permissions from Poppenk et al., *Trends Cogn Sci*, 2013, 17, 230–240 and Hackert et al., *Neuroimage*, 2002, 17, 1365–1372.

hippocampi in the sagittal plane for pitch correction (see Fig. 1) [Goncharova, 2001]. Second, using the coronal view of each rotated high-resolution scan, the raters registered the aHPC/pHPC division landmark, defined as the last slice where the uncus apex was visible when moving in the coronal plane from anterior to the pHPC in a rostro-caudal direction [Insausti and Amaral, 2012; Poppenk et al., 2013].

Raters were initially trained in these manual operations using a dataset of 65 scans. Then, right before they started performing these operations with the experimental sample of the present study, they were refreshed on these procedures under close supervision with another set of 10 scans.

With the 100 experimental subjects included in the present study, these operations (i.e., manually rotated the brain scans and registering the aHPC/pHPC division landmark) were performed twice, at time point I (Time I) and time point II (Time II), on each of the scans. To distribute rater's practice effects across these within-subjects conditions, the occurrence of Time I and Time II for each image was fully randomized, presenting both scans in an intermixed fashion. Raters were blind to this experimental manipulation and to the objectives of the study. Additionally, raters registered the aHPC/pHPC uncus apex division landmark in scans that were automatically rotated with the PCA-based method described in Automatic Extraction of Hippocampal Centerline section below. This operation was also performed twice per scan, in a fully randomized intermixed fashion (as for the rotations). These additional annotations relying on PCA-based rotations enable us to quantify the variability of the landmark placement in isolation, independently of the rotations.

Automatic Extraction of Hippocampal Centerline

Here, we propose two closely related methods to automatically extract the centerline of the hippocampus. The first one uses a linear approximation (i.e., a straight line estimated via PCA), and the second one a quadratic Bézier curve (see Fig. 3A1,B1). In both cases, we seek to minimize the sum of square distances of the voxels corresponding to the hippocampus to the centerline at hand. The linear approximation has the advantages that it directly corresponds to a rotation of the data (by aligning the centerline with the anterior–posterior axis) and that it has a closed-form solution—as we explain below. Conversely, the quadratic approximation has the advantage of better modeling the curved shape of the human hippocampus. This quadratic approximation requires additional numerical methods and higher computation demands relative to the linear approximation because the curve is fitted minimizing a cost function that does not correspond to a single rotation of the MRI scan (i.e., a curved line cannot be aligned to a straight axis).

In both methods, the first step is to compute binary hippocampal masks. Importantly, the algorithms proposed in the present study work on these masks (or on any hippocampal mask generated either by automated or manual procedures), and are not influenced per se by the nature of the input scans. Specifically, in this study, we used an automated program, *Freesurfer 5.3.0* [Fischl et al., 2002], to extract hippocampal masks. It has been previously shown that this software package produces hippocampal segmentations that are statistically as reliable as manual segmentations [e.g., Cherbuin et al., 2009; Morey et al., 2010, 2009;

Figure 3.

Linear PCA-based and quadratic Bézier-based approximations to automatically extract hippocampal centerline and place their corresponding aHPC/pHPC dividing planes (uncal apex). **A1)** PCA analysis over the tridimensional cloud of points of the hippocampus used to obtain the first principal component (red line), which approximately corresponds to the hippocampal longitudinal axis. This step is equivalent to automate the rotation of the brain for normalization before segmenting it. **A2)** Perpendicular dividing plane to the estimated PCA-based longitudinal axis used

to automatically segment aHPC/pHPC with the percentage-based segmentation method. **B1)** Quadratic approximation based on the Bézier curve (blue line), originated as the bending of the original PCA curve's three points. The blue planes are orthogonal to the principal axis and the green dots correspond to $\mathbf{x}_a$, $\mathbf{x}_c$, and $\mathbf{x}_p$, which can only move along the blue planes. **B2)** Perpendicular dividing plane to the Bézier curve used to automatically segment aHPC/pHPC with the percentage-based segmentation method.

Poppenk and Moscovitch, 2011; Tae et al., 2008]. Then, we extract the coordinates $\{\mathbf{r}_n\} = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$ of the N voxels inside the hippocampal mask. Next, we seek to minimize the function:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{n=1}^N d[\mathbf{r}_n, \mathbf{r}_c(t; \theta)]^2, \quad (1)$$

where d is the Euclidean distance and $\mathbf{r}_c(t; \theta)$ is the centerline, which is parameterized by a scalar t , and depends of a set of coefficients θ (whose optimal value is represented as $\hat{\theta}$).

For the linear PCA approximation, the centerline is given by:

$$\mathbf{r}_c(t; \theta) = \boldsymbol{\mu} + t' \mathbf{v}, \text{ with } t' \in [-\infty, \infty], \text{ and } \theta = (\boldsymbol{\mu}, \mathbf{v}), \quad (2)$$

where $\boldsymbol{\mu}$ is a point and $\mathbf{v}$ is a vector with norm one. In this case, the minimization problem in Eq. (1) has a closed-form solution given by PCA [Jolliffe, 2002]. More specifically, $\hat{\boldsymbol{\mu}}$ is given by the mean of the data:

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n=1}^N \mathbf{r}_n.$$

The vector $\hat{\mathbf{v}}$ is equal to the first principal component $\mathbf{v}_1$, which is the eigenvector corresponding to the leading eigenvalue of the covariance matrix of the data, given by:

$$\Sigma = \frac{1}{N} \sum_{n=1}^N (\mathbf{r}_n - \hat{\boldsymbol{\mu}})(\mathbf{r}_n - \hat{\boldsymbol{\mu}})^t.$$

An alternate parameterization uses two points $(\mathbf{x}_a, \mathbf{x}_p)$, which are obtained by projecting the data points onto the

major axis $\hat{v}$, and then taking the most anterior (x_a) and posterior (x_p) projections. In that case, the linear approximation can be rewritten as:

$$r_c(t) = x_a + t(x_p - x_a), \text{ with } t \in [0, 1], \quad (3)$$

where t now corresponds to the normalized arc length.

If we are approximating the centerline by a quadratic Bézier curve, a closed-form solution does not exist, so we resort to numerical methods. The general expression of a quadratic Bézier curve is:

$$r_c(t; \theta) = (1-t')[(1-t')x_0 + t'x_1] + t'[(1-t')x_1 + t'x_2], \quad (4)$$

with $t' \in [0, 1]$, and $\theta = (x_0, x_1, x_2)$, where (x_0, x_1, x_2) are three different points in space. Rather than using this expression directly, we reparameterize it to constrain x_0 , x_1 , and x_2 to remain within planes that cut the major axis of the hippocampus (as given by PCA) at its most anterior and posterior points, as well as its center (see Fig. 3B1). More specifically, we first compute the most anterior and posterior projections (x_a, x_p) as in the linear case, as well as the point right in between $x_c = \frac{1}{2}(x_a + x_p)$. Then, we use the parameterization:

$$x_0 = x_a + \lambda_a v_2 + \lambda'_a v_3, \quad x_1 = x_c + \lambda_c v_2 + \lambda'_c v_3, \quad x_2 = x_p + \lambda_p v_2 + \lambda'_p v_3,$$

where v_2 and v_3 are second and third principal components (and therefore perpendicular to v_1), and the new vector of coefficients is given by $\theta = \{\lambda_a, \lambda'_a, \lambda_c, \lambda'_c, \lambda_p, \lambda'_p\}$.

In addition to this reparameterization, we also discretize the curve by limiting the possible values of t' to multiples of 0.001, that is, $t' = \{0, 0.001, 0.002, \dots, 1\}$. The minimization of the cost function is carried out with numerical methods; in this study, we used the L-BFGS algorithm [Liu and Nocedal, 1989]. Finally, once the optimal coefficients θ have been computed, we reparameterize the curve once more such that t' is replaced by the normalized arc length $t \in [0, 1]$. In this new parameterization, t corresponds directly to the distance between the start and end points along the curve—which is the criterion we use to automatically divide the hippocampus along its axis.

Segmentation into aHPC/pHPC and Computation of Volumes

The aHPC and pHPC volumes are obtained by partitioning the hippocampal mask provided by FreeSurfer with a dividing plane, and subsequently to compute the volume of the mask on each side of the plane. In the manual case, the orientation (normal vector) of the plane is determined through a rotation, and the point at which cuts the hippocampus is given by the manually placed landmark corresponding to the uncus apex (see Manual Operations section). In the automated case, the plane is assumed to be perpendicular to the centerline extracted with the techniques proposed in Automatic Extraction of Hippocampal

Centerline section (see Fig. 3A2,B2). The point along the centerline through which the plane intersects the hippocampus is computed with a percentage-based method, in which the cutting point is given by setting t in Eq. (3) (linear) and Eq. (4) (quadratic Bézier) to a fixed percentage [Hackert et al., 2002].

In this study, we used $t = 41.3\%$ for the linear PCA approximation and $t = 42.1\%$ for the quadratic Bézier curve. For the linear PCA approximation, the percentage used was calculated by averaging the values of t corresponding to the manually placed landmarks (Manual Operations section). For the quadratic Bézier approximation, we draw a plane for all the manual landmarks selected from raters in the automated rotations and calculated the Bézier curve of the hippocampus. The intersection between the plane and the Bézier curve was considered as the dividing point between the aHPC and pHPC segments. Then, we calculated the final percentage as the average of all of them. For both the linear and the quadratic approximations, the total length of the axis was calculated drawing an orthogonal plane to the axis and selecting the last plane where all the structure was above/below it. It is important to note that the aHPC and pHPC percentages used for the linear PCA and the quadratic Bézier curve approximations yield different volumetric values because (1) the axes are different, with Bézier axis being longer and curved relative to the PCA axis (see Fig. 3A1,3B1), and (2) the dividing plane in the Bézier approximation is different, as it will be perpendicular to a curved axis and therefore will cut the hippocampus in an oblique manner (see Fig. 3A2,3B2).

In this study, we also consider two manual/automated hybrid methods, which are useful in our experiments to separate the effects of rotation and landmark placement on variability. First, a hybrid method in which the rotation is performed manually, but the landmark is chosen with percentages. And, second, a method in which the centerline is computed with PCA, but the landmark is placed manually; note that combining the Bézier centerline with manual landmark placing is not possible—since the orientation of the plane perpendicular to the centerline is not constant. These hybrid setups will enable us to specifically examine in isolation the effects of the variability of manual rotations on hippocampal aHPC/pHPC segmentation (using and automatized method), and the effects of the variability in the manual landmark selections (i.e., uncus apex) for hippocampal aHPC/pHPC segmentation when using an automatized procedure for the initial brain rotation.

Assessment of Variability in Rotation Angles

To examine inter-rater (A, B) and intra-rater (Time I, Time II) variability in manual rotations, it is necessary a metric to compare two different rotations. If two rotations are represented by matrices R_X and R_Y , we can compute

the difference between these two rotations as $R=R_X^{-1}R_Y$, and then compute the corresponding rotation angle as:

$$\alpha = \cos^{-1} \frac{R_{11} + R_{22} + R_{33} - 1}{2}$$

(note that R_{11} , R_{22} , and R_{33} correspond to the elements 11, 22, and 33 of the rotation matrix R , and that the result is the same if we define $R=R_Y^{-1}R_X$ instead). We can then use α as a measure of discrepancy between two manual rotations. Given a number of instances of α , we use a non-parametric statistical test [Wilcoxon signed rank, Wilcoxon, 1945] to assess whether the mean of this angle is significantly higher than a threshold α_t . We then find $\alpha_{0.05}$, the highest value of α for which $P < 0.05$, that is, the maximum value α at which the test reaches significance at the 0.05 level. We use $\alpha_{0.05}$ as the rotation variability metric in this study.

Assessment of Variability in Volume Estimates

To evaluate the reliability of volume estimates, we used the same measures utilized in Jeukens et al.'s [2009] study. Given two vectors $V_A = \{V_{A,1}, \dots, V_{A,M}\}$ and $V_B = \{V_{B,1}, \dots, V_{B,M}\}$ of volume estimates corresponding to two different measurements A and B (i.e., two raters, or the same rater at two different time points), the measures are the following:

- i. Intra-class correlation coefficient
 $ICC = \frac{\sigma_{bs}^2}{\sigma_{bs}^2 + \sigma_{ws}^2}$, where σ_{bs} is the between-subjects standard deviation (SD), and σ_{ws} is the within-subjects SD.
- ii. % Volume difference
 $VD (\%) = 200 \frac{|V_A - V_B|}{(V_A + V_B)}$, it is an indicator of the relative difference of the volumes obtained by A and B.

Assessment of Asymmetry

The asymmetry index (AI) is calculated separately for the total hippocampal volume, for the aHPC or head, and for the pHPC. The AI value is the percentage of the difference of the right volume versus the left, respect to the average $= \frac{200(V_r - V_l)}{(V_r + V_l)}$, where V_l and V_r are the volumes corresponding to the left and right sides, respectively.

Results

With the goal of quantifying the variability introduced by manual interventions, here we first present results characterizing the specific inter- and intra-rater variability in manual rotations and uncaval apex landmark selection (Variability in Manual Rotations and Uncaval Apex Landmark Placement section). Then, using the gold standard uncaval apex manual landmark selection for segmentation, we examine to what extent the use of manual versus the PCA

automated brain rotation influences aHPC/pHPC volume estimates (Variability in Hippocampal Volume Estimates as a Function of the Rotation Procedure Using the Manual Landmark-Based Segmentation Method section). In Variability in Hippocampal Volume Estimates and AIs as a Function of the Rotation Procedure Using the Percentage-Based Segmentation Method section, after assessing how the PCA and Bézier automated brain rotation solutions influence the location of the aHPC/pHPC-dividing plane obtained with the percentage-based automated segmentation, we examined to what extent the initial use of manual versus automated (PCA, Bézier) brain rotations influences aHPC/pHPC volume estimates obtained with this automated segmentation method. Finally, using the proposed automated brain rotation methods, we examine aHPC/pHPC volume differences (VDs) calculated with the gold standard manual landmark-based versus the automated percentage-based segmentations (Comparison of the Manual and Automated Segmentation Methods section).

Variability in Manual Rotations and Uncaval Apex Landmark Placement

Here, we examine inter- and intra-rater variability in the manual operations involved in current segmentation conventions intended to correct in-scanner head position across subjects and to visually select the aHPC/pHPC separating plane (i.e., uncaval apex) along the hippocampal longitudinal axis.

To characterize inter-rater (A, B) and intra-rater (Time I, Time II) variability in manual rotations, we calculated the vectors of the differences in rotation angles separately for left and right hippocampi. Inter-rater variability was examined by the difference in rotation angles made by rater A and B, obtained by averaging each rater rotations at time point Time I and Time II (i.e., A-B). Intra-rater variability was quantified by the difference in rotation angles of the annotations performed by each rater (A and B) at Time I relative to their own annotations at Time II (i.e., A1-A2, B1-B2). To test statistically significant differences in inter- and intra-rater variability, we computed *t*-tests against zero for each of the cases and used a Wilcoxon signed rank test with increasing values for the constant test starting from zero, we increased the value of the constant vector in increments of 0.1 rotation degrees. Figure 4A shows the Wilcoxon tests *P*-values for inter- and intra-rater comparisons as a function of the constant rotation angles. The maximum angles $\alpha_{0.05}$ at which statistical significance are achieved are summarized in Table I.

Conversely, to characterize the variability in the placement of the uncaval apex landmark, we first examined the inter- and intra-rater differences in the coronal slice selection in manually- and PCA-rotated brains. Mean distances and SDs in landmark placement in manually and PCA-rotated brains for rater A and B and left and right hippocampi are reported in Table II. The distribution of

A. Rotation. Inter- & intra-rater variability **B1. Landmark. Intra-rater variability** **B2. Landmark. Rotation x Rater interaction** **Figure 4.**

Inter- and intra-rater variability in manual interventions. **A)** Wilcoxon tests *P*-values for inter- and intra-rater comparisons as a function of rotation angles. A1A2 and B1B2 represent the Time I – Time II intra-rater variability. AB represents inter-rater variability (both A and B are calculated as Time I and Time II rotation averages). **B1)** Intra-rater differences (Time I – Time II) in the uncus apex slice selection for left and right hippocampi for manually- and PCA-rotated brains. **B2)** Rotation X Rater interaction, separated for left and right hippocampi, showing differences in the uncus apex slice selection in manually- and PCA-rotated brains for rater A and B.

distances between time points (Time I – Time II; intra-rater variability) for raters A and B are shown in Figure 4B1. Second, to examine the distance differences in the uncus apex slice selection as a function of raters and manual versus automated rotations, we conducted a mixed-model 2 (Rater: A, B) X (Rotation: manual, PCA) X 2 (Hemisphere: left, right) analysis of variance (ANOVA). This analysis revealed that the significant main effects of Rater, $F(1, 100) = 21.23$, $P < 0.0001$, and Rotation, $F(1, 100) = 12.20$, $P < 0.001$, were subsumed by a statistically significant Rater X Rotation interaction, $F(1, 100) = 13.34$, $P < 0.001$ (see Fig. 4B2). Tukey's honest post hoc analyses revealed that this interaction was due to a statistically significant higher distance in the landmark slice selection for rater A relative to rater B in the manual rotation procedure, $P < 0.0001$. This difference between raters was not observed in the PCA rotation, $P = 0.94$. No main or interactive significant effects emerged for the factor Hemisphere, $P > 0.05$.

Variability in Hippocampal Volume Estimates as a Function of the Rotation Procedure Using the Manual Landmark-Based Segmentation Method

The variability in rotation and landmark placement eventually carries over to volume estimates of aHPC and pHPC and AIs. A relevant question is to what extent aHPC and pHPC volumes segmented via the gold-standard manual landmark-based segmentation are affected by the use of manual versus PCA rotations.³ Table III and IV shows the intra- and inter-rater intra-class correlations (ICC) and VD for left and right aHPC and pHPC volume comparisons for manual and PCA-automated rotations.

As previous empirical evidence suggests that the hippocampal head drives the asymmetry right > left asymmetry observed in total hippocampal volume [e.g., Woolard and Heckers, 2012], we used the AI of the aHPC as a proxy for the quality of the landmark-based segmentation performed either over manually or PCA rotated brains. To this end, using aHPC volume AI as the dependent variable, we performed a 2 (Rater: A, B) X 2 (Rotation: Manual, PCA) mixed-model ANOVA with Time I and Time II averaged across each rater A and B. No main or interactive effects emerged in this analysis ($P > 0.05$) suggesting that an average value of the four values could be a good summary of the aHPC AI obtained using the manual landmark-based segmentation over manually and PCA rotated brains

³The manual and automated PCA rotations can be effectively compared since they both involved similar procedures to normalize the brain position before identifying a landmark (uncus apex) to segment aHPC/pHPC. However, it is not possible to combine the Bézier centerline with manual landmark placing since the orientation of the plane perpendicular to the centerline is not constant (see Segmentation into aHPC/pHPC and Computation of Volumes section for further details).

Figure 5.

aHPC averaged volume AI as a function of rotation (manual, PCA) and rater (A, B). Error bars show the standard error with a 0.95 confidence interval.

(see Fig. 5). Nevertheless, in line with our initial assumption of finding stronger inter-rater differences in aHPC volume AIs when using the manual landmark-based segmentation procedure over manually versus PCA rotated brains, simple-effect planned comparisons revealed that the inter-rater difference for aHPC volume AIs when using PCA was 1.01, $t(181) = 2.48$, $P = 0.014$, and when using manual rotation was almost the double [1.70 difference with $t(188) = 4.24$, $P = 0.00003$].

Variability in Hippocampal Volume Estimates and AIs as a Function of the Rotation Procedure Using the Percentage-Based Segmentation Method

Here, before examining to what extent the initial use of manual versus automated (PCA, Bézier) brain rotations

TABLE I. Inter- (A-B) and intra-rater (A1-A2, B1-B2) manual rotation differences by hemisphere and maximum rotation angle where the *P*-value of the Wilcoxon test is ≤ 0.05

Comparison	Left angle	Right angle
A1-A2	1.70	1.60
B1-B2	2.05	2.15
A-B	2.50	3.30

TABLE II. Absolute means and SDs for intra-rater manual landmark selection differences for Time I - Time II in left and right hippocampi, as a function of manual and automated (i.e., PCA) rotations

Rotation	Rater	Left Mean (SD)	Right Mean (SD)
Manual	A	0.580 (1.437)	0.260 (1.587)
PCA	A	0.140 (1.146)	0.040 (1.197)
Manual	B	0.260 (1.411)	0.140 (1.504)
PCA	B	0 (1.271)	0.050 (1.123)

influences aHPC/pHPC volume estimates obtained with the percentage-based automated segmentation method, we assessed first how PCA and Bézier automated brain rotation solutions influence the location of the aHPC/pHPC-dividing plane. To this end, for all the possible percentage length selections of the dividing plane, we calculate aHPC and pHPC mean volume AI comparisons and compute effect sizes (ES) for the hypothesis that the AI of the aHPC is different than zero.

Figure 6A1,B1 show the volume AI for the aHPC and pHPC for all possible values of $t \in [25,75]\%$, using the PCA-based and the Bézier-based rotations, respectively. Figure 6A2, 6B2 show the ES for the hypothesis that the AI of the aHPC is significantly different from zero, also for every possible values of $t \in [25,75]\%$, using the PCA-based and the Bézier-based rotations, respectively. Importantly, the ES figures show local maximums around the selected percentage values for which the plane intersects the hippocampus for the linear PCA (i.e., 41.3%) and the quadratic Bézier (i.e., 42.1%) approximations.

TABLE III. Intra-rater ICC and VD volume comparisons (Time I vs. Time II) as a function of hemisphere, hippocampal segment, rater, and rotation

Rater A	Left		Right	
	aHPC	pHPC	aHPC	pHPC
PCA rotation				
ICC	0.966	0.947	0.962	0.948
VD	2.624	3.150	3.421	4.018
Manual ROTATION				
ICC	0.963	0.930	0.952	0.930
VD	3.488	4.079	3.465	4.216

Rater B	Left		Right	
	aHPC	pHPC	aHPC	pHPC
PCA rotation				
ICC	0.933	0.914	0.945	0.930
VD	2.886	2.798	2.759	2.644
Manual rotation				
ICC	0.974	0.962	0.955	0.946
VD	2.868	2.908	3.245	3.376

TABLE IV. Inter-rater ICC and VD volume comparisons (A vs. B) as a function of hemisphere, hippocampal segment, rotation, and time

Rotation	Left aHPC	Left pHPC	Right aHPC	Right pHPC
PCA (Time I)				
ICC	0.975	0.963	0.968	0.959
VD	5.016	5.261	6.241	6.945
PCA (Time II)				
ICC	0.929	0.904	0.942	0.920
VD	6.305	6.734	6.683	7.230
Manual (Time I)				
ICC	0.967	0.945	0.938	0.920
VD	5.681	6.277	7.601	8.471
Manual (Time II)				
ICC	0.941	0.901	0.948	0.929
VD	5.726	6.139	6.949	7.890

To examine to what extent the manual versus automated (i.e., PCA, Bézier) rotation procedures also influence hippocampal segmentations obtained via the percentage-based automated segmentation method, we obtained ICC and VD values for left and right aHPC and pHPC hippocampal segments based on the three comparisons of interest: (1) manual versus PCA rotation (at 41.3%), (2) manual versus Bézier rotation (at 42.1%), and (3) PCA versus Bézier rotation (see Table V). We also calculated the volume AI of the aHPC obtained using the percentage-based segmentation method as a function of the Rotation Method (manual, PCA, Bézier) and submit them to a one-way ANOVA. This analysis revealed a statistically significant main effect of the Rotation Method, $F(2, 100) = 5.41$, $P < 0.005$. Simple-effect analysis showed that the mean volume AI of the aHPC were significantly different for the comparisons Manual versus PCA and Bézier, $P < 0.001$, and PCA versus Bézier, $P < 0.00001$ (see Fig. 7A). Additionally, we performed a power analysis ($\alpha = 0.05$, $\beta = 0.2$ for the detection of the aHPC volume AI as a function of the rotation method (see Fig. 7B). This analysis revealed that the PCA and Bézier automated rotation methods yielded moderate improvements (i.e., several tenths) in ES compared to the manual rotation procedure, as well as a reduction in the required sample size to detect the aHPC volume AI effect of 40% for PCA and of 58% for Bézier.

Comparison of the Manual and Automated Segmentation Methods

After examining the advantages of using the automated rotation procedures, here we compare aHPC and pHPC VDs for the gold-standard manual landmark-based segmentation performed on PCA rotated brains (average of 4 data points: rater A/B in Time I/Time II), with the auto-

ated percentage-based segmentation method performed on PCA and Bézier rotated brains. Table VI shows ICC and VD values for left and right aHPC and pHPC hippocampal segments for the two comparisons of interest: (1) manual landmark-based segmentation performed on PCA rotated brains (PCA-landmark) versus automated percentage-based segmentation performed on PCA rotated brains (PCA-percentage), and (2) manual landmark-based segmentation performed on PCA rotated brains (PCA-landmark) versus automated percentage-based segmentation performed on Bézier rotated brains (Bézier-percentage).

Similar to Variability in Hippocampal Volume Estimates and AIs as a Function of the Rotation Procedure Using the Percentage-Based Segmentation Method section, using the aHPC volume AI as the dependent variable, we performed a one-way ANOVA for the repeated measures factor Method (PCA-landmark, PCA-percentage, Bézier-percentage). No main effect of the factor Method emerged in this analysis [$F(2, 100) = 2.11$, $P = 0.12$]. Simple-effects planned comparisons revealed that the only statistically significant difference was the one already reported in Variability in Hippocampal Volume Estimates and AIs as a Function of the Rotation Procedure Using the Percentage-Based Segmentation Method section: PCA-percentage versus Bézier-percentage, $P < 0.00001$.

Thus, the use of the manual landmark-based or the automated percentage-based segmentation methods in combination with automated rotation procedures did not yield statistically significant differences in the aHPC volume AI effect. Nevertheless, when performing a power analysis for detecting the aHPC volume AI effect, there is reduction in the required sample sizes to detect this effect with automated methods relative to PCA-landmark: 27% less for PCA-percentage and 48% less for Bézier-percentage.

TABLE V. Brain rotations ICC and VD volume comparison as a function of hemisphere and hippocampal segment for the comparisons of interest

Methods	Left aHPC	Left pHPC	Right aHPC	Right pHPC
Manual versus PCA				
ICC	0.974	0.969	0.982	0.984
VD	4.603	4.867	3.401	3.611
Manual versus Bezier				
ICC	0.966	0.957	0.961	0.960
VD	3.078	3.367	3.918	4.337
PCA versus Bezier				
ICC	0.971	0.970	0.951	0.956
VD	4.093	4.278	5.831	6.278

Manual values are calculated as the averages of the manual rotations for both raters in both time points. In all cases, the segmentation was done with the automated percentage method.

PCA method

A1. aHPC and pHPC mean volume AI-s

A2. Effect size of the aHPC volume AI

Bèzier method

B1. aHPC and pHPC mean volume AI-s

B2. Effect size of the aHPC volume AI

Figure 6.

Hippocampal mean volume AI and ES for all the possible percentage length selections of the dividing plane for automated brain rotation methods. **A1)** Mean volume AI for aHPC and pHPC obtained using the PCA-based rotation method. The horizontal dot-dashed line shows the volume AI for the whole hippocampus. **A2)** ES for the hypothesis that the volume AI of the aHPC obtained using the PCA-based rotation method is differ-

ent from zero. The horizontal dashed line represents the ES for the total hippocampus. The horizontal solid line shows the ES for the average manual landmark values. **B1)** and **B2)** represent mean volume AIs for aHPC and pHPC and the ES for the hypothesis that the AI of the aHPC is different from zero, respectively, obtained using the Bèzier-based rotation method.

DISCUSSION

The present work was aimed at introducing and examining automated methods to segment the human hippocampus along its longitudinal axis with a series of experiments intended to characterize the variability of manual operations and compare results obtained with standard manual rotation and landmark placement, hybrid automated rotation and manual landmark placement, and fully automated methods.

Our results highlight the inherent variability present in manual procedures. Manual rotations yielded substantial inter- and intra-rater variability relative to the PCA-based automated rotation (Variability in Manual Rotations and Uncal Apex Landmark Placement section, Fig. 4A), which is not subjected to this form of undesired variance. Importantly, the PCA-based rotation is intended to mimic the manual horizontal rotation using its first principal component, which approximately corresponds to the hippocampal longitudinal axis. Similarly, our result revealed that

manual landmark selection for aHPC/pHPC segmentation is subjected to the same problems as manual rotation, being susceptible to unwanted inter- and intra-rater variability (Fig. 4B). These results provide a useful characterization of the variability of raters' manual rotations and their subsequent effects on manual landmark placement, and underline the advantages of using automated rotation methods: they are (1) tailored to each hippocampus, and (2) consistent across time points and raters.

We further evaluated manual versus automated rotation carryover effects on the manual landmark and percentage-based aHPC/pHPC segmentations. On the one hand, whereas manual rotations had a clear negative impact on manual landmark positioning, automated rotation solutions led to more reliable manual landmark positioning at the inter- and intra-rater levels (Variability in Hippocampal Volume Estimates as a Function of the Rotation Procedure Using the Manual Landmark-Based Segmentation Method section). This finding strongly advises for the use of automated tools to consistently rotate the hippocampus prior

Figure 7.

aHPC volume AI effect detection (using the percentage-based segmentation method) as a function of the rotation procedure. **A)** Volume AI per rotation procedure. Error bars show the standard error with a 0.95 confidence interval. **B)** Power analysis with $\alpha = 0.05$, $\beta = 0.2$ (ES of the aHPC volume AI effect) and number of subjects needed to find a statistically significant AI effect per rotation procedure.

to the application of manual landmark positioning. On the other hand, carryover effects of manual rotations on the percentage-based segmentation method revealed VDs within acceptable margins, according to the criteria proposed by Jeukens et al. [2009]: $ICC \geq 0.85$ and $VD \leq 15\%$. However, as shown in Variability in Hippocampal Volume Estimates and AIs as a Function of the Rotation Procedure Using the Percentage-Based Segmentation Method section, the manual versus automated rotations using the automated percentage-based segmentation method yielded significant differences on the detection of well-know right > left aHPC asymmetry effect [Woolard and Heckers, 2012] and, consequently, on the statistical power to detect

this effect (Fig. 7). When using an automated rotation procedure, we observed a reduction in the required sample size to detect aHPC asymmetry effects of about 50%. This gain in statistical power can be entirely attributable to the lack of variability on the original rotation when using an automated rotation procedure, given that the segmentation procedure is equivalent in all cases.

Our results also revealed that if the initial variability introduced by manual rotations is removed using automated rotation methods (i.e., PCA, Bézier), the hippocampal segmentation results obtained with the gold-standard manual landmark selection can be satisfactorily consistent (Comparison of the Manual and Automated Segmentation Methods section). When using any of the proposed automated rotation solutions, there were no statistically significant differences in the aHPC/pHPC volume estimates (i.e., ICC, VD, AI) obtained with either the manual landmark selection segmentation procedure or the percentage-based automated segmentation method. This absence of significant differences suggests that either the manual landmark-based or automated percentage-based segmentation procedures can be reliably used on automatically rotated hippocampi, but it is worth noticing that the automated methods come with better reproducibility and a considerable gain in ES/reduction in the required sample sizes required to detect well-known effects. These results also highlight that the automated procedures suggested in the present work can be applied flexibly, being possible to use automated brain rotation to remove the undesired variability only associated with manual brain rotations and continue with hippocampal segmentation based on the visual selection of a given anatomical landmark, or to proceed with both automated rotation and segmentation methods to eliminate inter- and intra-rater variability associated with these manual operations.

Given the benefits of PCA- and Bézier-based automated hippocampal rotation solutions in reducing undesired variability in the rotation itself and, critically, in the subsequent hippocampal segmentation, it is important to consider the conceptual implications of using the PCA- or Bézier-based mathematical rotations and to compare their effects on the results obtained in the subsequent hippocampal segmentation. The PCA-based rotation yields a straight parallel axis that goes along the sagittal view of the hippocampus. Thus, in the PCA rotated brain the hippocampus looks flat aligned with the axial planes and, accordingly, the perpendicular dividing plane looks like a coronal plane in the rotated brain. Conversely, the dividing plane based on the Bézier curve is slightly tilted, going below the uncus apex, due to the curve's direction. The use of the PCA- versus Bézier-based rotation has implications for the results, as shown in Variability in Hippocampal Volume Estimates and AIs as a Function of the Rotation Procedure Using the Percentage-Based Segmentation Method section, where the same automated

TABLE VI. Manual versus automated ICC and VD volume comparisons as a function of hemisphere and hippocampal segment for PCA-landmark versus PCA-percentage and PCA-landmark versus Bézier-percentage

Methods	Left aHPC	Left pHPC	Right aHPC	Right pHPC
PCA-landmark versus PCA-percentage				
ICC	0.916	0.897	0.886	0.883
VD	5.710	5.975	6.164	6.715
PCA-landmark versus Bézier-percentage				
ICC	0.945	0.929	0.931	0.916
VD	4.311	4.549	4.694	5.302

Manual values are calculated as the averages of the manual landmark registrations for both raters at both time points. In all cases, the rotation was automated.

percentage-based segmentation was applied to different rotations solutions. Our data showed a slightly better statistical power to detect aHPC asymmetry effects when using the Bézier- versus PCA-based rotation solution.

Therefore, depending on the biological assumptions of the researcher, our tool will give the flexibility to test different hypotheses, either using the PCA- or Bézier-based rotation methods, or selecting a different percentage to test different hypothesis. For instance, in line with findings from neuroanatomical research on hippocampal postnatal development showing that hippocampus growth in size over development is more evident at the body and tail [Insausti et al., 2010], DeMaster et al. [2014] argued that the uncus apex is a too posterior landmark for aHPC/pHPC hippocampal division when dealing with developing human subjects. So, to separate the aHPC from the pHPC, DeMaster et al. used the slice in which digitations on the dorsal edge of the hippocampus were no longer apparent and where the hippocampus begins to round into a teardrop shape. Our tool can be adapted to a desired aHPC/pHPC segmenting landmark (i.e., uncus apex, digitation), offering the researchers the possibility to use the percentage that better corresponds to their main biological assumptions for hippocampal segmentation along its longitudinal axis.

It should be also noted that to better adjust the percentage at which the hippocampus is segmented, our tool also allows to visually inspecting the resulting segmentations at the individual level in a three-dimensional rendering. Furthermore, as the tool is released with all the code required to replicate the analyses reported in this article (available as Matlab scripts in Supporting Information and <https://github.com/garikoitz/hippovol>) and can be applied to any datasets, the interested researcher could even normalize the rotation of all brains automatically according to PCA- or Bézier-based solutions, select the landmarks manually, and then run the script again to obtain the final longitudinal segmentations. Similarly, for the researchers interested on segmenting the pHPC into body and tail subsections, our script automatically provides these measurements separately (the head corresponds to the aHPC and addition of the body and the tail to the pHPC). Given that Frankó et al. [2014] argue that there are no clear biological divisions to justify a precise hippocampal body-tail segmentation landmark and that, if a segmentation between these posterior regions is required, segmenting the pHPC at 50% of its length is as valid as any other criterion, the automated segmentation of the pHPC in two length halves to obtain the body, and tail is performed in our tool at the 50% of the pHPC length. Nevertheless, for the researcher interested in exploring other landmarks to separate the tail from the body, the script also allows to load coordinates as specific landmarks to select the precise desired dividing location to segment the pHPC into body and tail.

CONCLUSION

In the present work, we propose automated tools for hippocampal segmentation along its longitudinal axis. We have shown that these procedures can successfully replicate well-known effects obtained with gold-standard manual methods, with the advantage of (1) reducing the variability associated to manual operations, (2) increasing the statistical power to detect well-known effects, and (3) potentially facilitate the standardization of the procedures to inter- and intra-lab data sharing and reproducibility. Importantly, here we departed from a widely used whole-hippocampus segmenting technique, but our tool is designed to work with any of the available whole-hippocampus segmentation techniques or procedures.

Despite of the theoretical and practical relevance of having a reliable method to segment the hippocampus along its longitudinal axis, currently there are still a wide variety of idiosyncratic segmentation procedures for aHPC/pHPC segmentation that still relies on trained raters performing manual operations. Here, we proposed and tested automated methods for hippocampal segmentation along its longitudinal axis in relation to gold standard manual procedures, demonstrating that these methods can be useful to automate and to improve reproducibility to this end.

It is worth noticing that here we propose two automated solutions that may have implications for the results. The Bézier-based solution, compared to the PCA-based alternative, shows a slightly better statistical power to detect aHPC asymmetry effects and segments the hippocampus following its curved shape; so it can be chosen in studies with smaller samples interested in asymmetry effects and when researchers' assumptions fit with the positioning of the human hippocampus that this solution offers. In contrast, compared to the Bézier-based solution, the PCA-based alternative better imitates the hippocampal positioning obtained with current manual segmentation procedures; so it can be chosen in studies that care about reproducing previous findings and when researchers' assumptions fit better with this approach. Our main goal here is to offer researchers useful and flexible automated tools that best fit their segmenting needs. These tools pave the way for new research avenues where only the sections of the hippocampus that contribute most to any effect—whether it could be AI, developmental changes, volumetric correlations with behavioral indexes, or via the creation of seed regions for structural or functional MRI analysis—can be selected and examined in line with specific research questions of interest.

In sum, exploring the functional and structural contribution of the different hippocampal segments along its longitudinal axis constitutes an active and relevant field of research in cognitive neuroscience. We characterized the variability associated to the manual operations typically applied in existing protocols, providing flexible automated tools to segment the hippocampus along its longitudinal axis, and validated it through a comparison with the gold

standards. This tool is made publicly available, allowing researchers exploring the contributions of the different hippocampal sections, and to facilitate data sharing, reproducibility, and comparisons across laboratories.

ACKNOWLEDGMENT

We thank Larraitz Lopez and Oihana Vadillo for their assistance with data collection and David Carcedo and Itzal Uranga for data generation.

REFERENCES

- Bartzokis G, Altshuler LL, Greider T, Curran J, Keen B, Dixon W (1998): Reliability of medial temporal lobe volume measurements using reformatted 3D images. *Psychiatry Res Neuroimaging* 82:11–24.
- Cherbuin N, Anstey KJ, Réglade-Meslin C, Sachdev PS (2009): In vivo hippocampal measurement and memory: A comparison of manual tracing and automated segmentation in a large community-based sample. *PLoS One* 4:e5265.
- Chupin M, Gerardin E, Cuingent R, Boutet C, Lemieux L, Lehericy S, Benali H, Garnero L, Colliot O (2009): Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* 19:579–587.
- Clark RE, Squire LE (1998): Classical conditioning and brain systems: The role of awareness. *Science* 280:77–81.
- Convit A, De Leon MJ, Tarshish C, De Santi S, Tsui W, Rusinek H, George A (1997): Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiol Aging* 18:131–138.
- DeMaster D, Pathman T, Lee JK, Ghetti S (2014): Structural development of the hippocampus and episodic memory: Developmental differences along the anterior/posterior axis. *Cereb Cortex* 24:3036–3045.
- DeToledo-Morrell L, Stoub TR, Bulgakova M, Wilson RS, Bennett DA, Leurgans S, Wu J, Turner DA (2004): MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD. *Neurobiol Aging* 25:1197–1203.
- Duvernoy HM (2005): *The Human Hippocampus: Functional Anatomy, Vascularization and Serial Sections with MRI*, 3rd ed. Springer, Berlin.
- 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 labeling of neuroanatomical structures in the human brain. *Neuron* 33:341–355.
- Frankó E, Insausti AM, Artacho-Pérula E, Insausti R, Chavoix C (2014): Identification of the human medial temporal lobe regions on magnetic resonance images. *Hum Brain Mapp* 35:248–256.
- Fyhn M, Molden S, Witter MP, Moser EI, Moser MB (2004): Spatial representation in the entorhinal cortex. *Science* 305:1258–1264.
- Goncharova I, Dickerson BC, Stoub TR, DeToledo-Morrell L (2001): MRI of human entorhinal cortex: A reliable protocol for volumetric measurement. *Neurobiol Aging* 22:737–745.
- Hackert VH, den Heijer T, Oudkerk M, Koudstaal PJ, Hofman A, Breteler MMB (2002): Hippocampal head size associated with verbal memory performance in nondemented elderly. *Neuroimage* 17:1365–1372.
- Haller JW, Banerjee A, Christensen GE, Gado M, Joshi S, Miller MI, Sheline Y, Vannier MW, Csernansky JG (1997): Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. *Radiology* 202:504–510.
- Iglesias JE, Augustinack JC, Nguyen K, Player CM, Player A, Wright M, Roy N, Frosch MP, McKee AC, Wald LL, Fischl B, Van Leemput K (2015): A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *Neuroimage* 115:117–137.
- Insausti R, Amaral DG (2012): Hippocampal formation. In: Mai JK, Paxinos G, editors. *The Human Nervous System*, 3rd ed. Elsevier, Amsterdam.
- Insausti R, Cebada-Sánchez S, Marcos P (2010): Postnatal Development of the Human Hippocampal Formation. In: *Advances in Anatomy, Embryology and Cell Biology*. Springer, Heidelberg.
- Jack CR (1994): MRI-based hippocampal volume measurements in epilepsy. *Epilepsia* 35:S21–S29.
- Jeukens CR, Vlooswijk MC, Marian Majoie HJ, de Krom MC, Aldenkamp AP, Hofman PA, Jansen JF, Backes WH (2009): Hippocampal MRI volumetry at 3 Tesla: Reliability and practical guidance. *Invest Radiol* 44:509–517.
- Jolliffe IT (2002): *Principal Component Analysis*, 2nd ed. Springer, New York.
- Killiany RJ, Moss MB, Albert MS, Sandor T, Tieman J, Jolesz F (1993): Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. *Arch Neurol* 50:949–954.
- Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B (2009): Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* 19:549–557.
- Lehericy S, Baulac M, Chiras C, Martin N, Pierot L, Pillon B, Deweer B, Dubois B, Marsault C (1994): Amygdalo-hippocampal MR volumetric measurements in the early stages of Alzheimer's disease. *Am J Neuroradiol* 15:927–937.
- Liu DC, Nocedal J (1989): On the limited memory BFGS method for large scale optimization. *Math Program* 45:503–528.
- Mai JK, Paxinos G (2011): *The human nervous system*, 3rd ed. Elsevier, Amsterdam.
- Malykhin NV, Bouchard TP, Ogilvie CJ, Coupland NJ, Seres P, Camicioli R (2007): Three-dimensional volumetric analysis and reconstruction of amygdala and hippocampal head, body and tail. *Psychiatry Res* 155:155–165.
- Morey RA, Petty CM, Xu Y, Hayes JP, Wagner HR, Lewis DV, LaBar KS, Styner M, McCarthy G (2009): A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* 45:855–866.
- Morey RA, Selgrade ES, Wagner HR, Huettel SA, Wang L, McCarthy G (2010): Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Hum Brain Mapp* 31:1751–1762.
- Pantel J, O'Leary DS, Cretsinger K, Bockholt HJ, Keefe H, Magnotta VA, Andreasen NC (2000): A new method for the in vivo volumetric measurement of the human hippocampus with high neuroanatomical accuracy. *Hippocampus* 10:752–758.
- Persson J, Spreng RN, Turner G, Herlitz A, Morell A, Stening E, Wahlund L-O, Wikström J, Söderlund H (2014): Sex differences in volume and structural covariance of the anterior and posterior hippocampus. *Neuroimage* 99C:215–225.

- Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013): Long-axis specialization of the human hippocampus. *Trends Cogn Sci* 17:230–240.
- Poppenk J, Moscovitch M (2011): A hippocampal marker of recollection memory ability among healthy young adults: Contributions of posterior and anterior segments. *Neuron* 72:931–937.
- Pruessner JC, Li LM, Serles W, Pruessner M, Collins DL, Kabani N, Lupien S, Evans AC (2000): Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. *Cereb Cortex* 10:433–442.
- Robinson JL, Barron DS, Kirby LAJ, Bottenhorn KL, Hill AC, Murphy JE, Katz JS, Salibi N, Eickhoff SB, Fox PT (2015): Neurofunctional topography of the human hippocampus. *Hum Brain Mapp* 36:5018–5037.
- Scoville WB, Milner B (2000): Loss of recent memory after bilateral hippocampal lesions. 1957. *J Neuropsychiatry Clin Neurosci* 12:103–113.
- Soininen HS, Partanen K, Pitkänen A, Vainio P, Hanninen T, Hallikainen M, Koivisto K, Riekkinen PJ (1994): Volumetric MRI analysis of the amygdala and the hippocampus in subjects with age-associated memory impairment: Correlation to visual and verbal memory. *Neurology* 44:1660–1660.
- Strange BA, Witter MP, Lein ES, Moser EI (2014): Functional organization of the hippocampal longitudinal axis. *Nat Rev Neurosci* 15:655–669.
- Swanson L, Cowan W (1977): An autoradiographic study of the organization of the efferent connections of the hippocampal formation in the rat. *J Comp Neurol* 172:49–84.
- Tae WS, Kim SS, Lee KU, Nam E-C, Kim KW (2008): Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology* 50:569–581.
- Watson C, Andermann F, Gloor P, Jones-Gotman M, Peters T, Evans A, Olivier A, Melanson D, Leroux G (1992): Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. *Neurology* 42:1743–1750.
- Wilcoxon F (1945): Individual comparisons by ranking methods. *Biometrics Bull* 1:80–83.
- Woolard A, Heckers S (2012): Anatomical and functional correlates of human hippocampal volume asymmetry. *Psychiatry Res* 201:48–53.
- Yushkevich PA, Pluta JB, Wang H, Xie L, Ding S-L, Gertje EC, Mancuso L, Kliot D, Das SR, Wolk DA (2015): Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum Brain Mapp* 36:258–287.