Title

Assessing hippocampal development and language in early childhood: Evidence from a new application of the Automatic Segmentation Adapter Tool

Journal

Human Brain Mapping, 36(11)

ISSN

1065-9471

Authors

Lee, Joshua K
Nordahl, Christine W
Amaral, David G
et al.

Publication Date

2015-11-01

DOI

10.1002/hbm.22931

Peer reviewed

◆ Human Brain Mapping 36:4483–4496 (2015) ◆

Assessing Hippocampal Development and Language in Early Childhood: Evidence From a New Application of the Automatic Segmentation Adapter Tool

Joshua K. Lee,[1,2]* Christine W. Nordahl,[3] David G. Amaral,[3] Aaron Lee,[3] Marjorie Solomon,[3] and Simona Ghetti[1,2]*

[1]*Department of Psychology, University of California, Davis, California*
[2]*Center for Mind and Brain, University of California, Davis, California*
[3]*MIND Institute, University of California, Davis, California*

Abstract: Volumetric assessments of the hippocampus and other brain structures during childhood provide useful indices of brain development and correlates of cognitive functioning in typically and atypically developing children. Automated methods such as FreeSurfer promise efficient and replicable segmentation, but may include errors which are avoided by trained manual tracers. A recently devised automated correction tool that uses a machine learning algorithm to remove systematic errors, the Automatic Segmentation Adapter Tool (ASAT), was capable of substantially improving the accuracy of FreeSurfer segmentations in an adult sample [Wang et al., 2011], but the utility of ASAT has not been examined in pediatric samples. In Study 1, the validity of FreeSurfer and ASAT corrected hippocampal segmentations were examined in 20 typically developing children and 20 children with autism spectrum disorder aged 2 and 3 years. We showed that while neither FreeSurfer nor ASAT accuracy differed by disorder or age, the accuracy of ASAT corrected segmentations were substantially better than FreeSurfer segmentations in every case, using as few as 10 training examples. In Study 2, we applied ASAT to 89 typically developing children aged 2 to 4 years to examine relations between hippocampal volume, age, sex, and expressive language. Girls had smaller hippocampi overall, and in left hippocampus this difference was larger in older than younger girls. Expressive language ability was greater in older children, and this difference was larger in those with larger hippocampi, bilaterally. Overall, this research shows that ASAT is highly reliable and useful to examinations relating behavior to hippocampal structure. *Hum Brain Mapp 36:4483–4496, 2015.*

Key words: segmentation; freesurfer; autism; development; hippocampus; language; segmentation adapter; methods

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

Contract grant sponsor: National Institute of Mental Health; Contract grant numbers: R01 MH089626, U24 MH081810, R00 MH085099, R01 NS16980; R01MH091109; Contract grant sponsor: UC Davis MIND Institute.

***Correspondence to**: Joshua K. Lee, 202 Cousteau Place, Davis, CA 95618. E-mail: jkilee@ucdavis.edu or Simona Ghetti, 202 Cousteau Place, Davis, CA 95618. E-mail: sghetti@ucdavis.edu

Received for publication 6 February 2015; Revised 23 July 2015; Accepted 28 July 2015.

DOI: 10.1002/hbm.22931
Published online 17 August 2015 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Recent years have witnessed an increase in studies examining developmental differences in subcortical structures, such as the hippocampus and the amygdala [Wierenga et al., 2014] to inform hypotheses about typical [Koolschijn and Crone, 2013; Lee et al., 2014] and atypical neurocognitive development [Schumann et al., 2004]. Traditionally, these volumetric assessments have been obtained by manually segmenting structural magnetic resonance images by expert raters [Schumann et al., 2004]. Manual methods remain the gold-standard both in terms of reliability and validity. However, these methods are time-consuming and require considerable personnel training. Furthermore, in developmental research with longitudinal designs in which data are collected and analyzed over a number of years, the use of manual raters might introduce unique sources of error, such as rater drift or unintended alterations in how a rating protocol is implemented because of changes in research personnel. Addressing these problems in longitudinal designs requires periodic re-assessment against older datasets [Nugent et al., 2007; Warshaw et al,, 2001].

For these reasons, over the last several decades, automated segmentation methods have been used in a variety of studies and populations [Everaerd et al., 2012; Kumfor et al., 2014; Weiss et al., 2005]. Although automated segmentation promises quicker outcome, it may also be prone to error, as segmentations may result in inclusion of irrelevant tissue when structural boundaries are not clear because of imaging protocol characteristics or motion; if boundaries are not clear, errors may range from sizeable to catastrophic. For example, FreeSurfer [Fischl, 2012], a widely-used software package for cortical thickness and segmentation of subcortical volumes, frequently mislabels voxels from the lateral ventricles and amygdala as portions of the hippocampal formation [Dewey et al., 2010; Shen et al., 2009; Tae et al., 2008; Wang et al., 2011]. This systematic error produces substantially larger hippocampal volume estimates than produced by manual rating. These kinds of errors may be more frequent or robust when the automated method is used with imaging protocols [Han et al., 2006; Jovicich et al., 2009] or study populations [e.g., Evans et al., 2012] that differ from those originally employed to validate the method. For example, FreeSurfer atlases were optimized for analyzing adult brain structure [Fischl et al., 2002], and there is the risk that its accuracy with use of very young children may be reduced because of group differences in brain morphology and tissue contrast, resulting in suboptimal registration and segmentation of subcortical brain structures. This problem may be particularly evident in studies of young children, a population in which both brain morphology and tissue contrast may alter the reliability and validity of FreeSurfer segmentation. While some data suggest that FreeSurfer segmentations are reliable assessments of hippocampal volume compared with manual tracing with older children [ages 8 to 11; DeMaster et al., 2014], there has not been a published validation of hippocampal segmentation with FreeSurfer in younger children.

These risks may be alleviated through the use of semi-automated correction methods that have been recently introduced in the literature. Specifically, the Automatic Segmentation Adapter Tool (ASAT is freely available software (http://www.nitrc.org/projects/segadapter), which can identify and correct segmentation errors produced by automated methods in adults [Wang et al., 2011]. The ASAT is conceptually based on the idea that automated segmentation methods, such as FreeSurfer, commit both random and systematic errors, and that systematic errors can be identified and removed using machine learning [i.e. Adaboost classifiers; Freund and Schapire, 1995]. In the ASAT, systematic errors are those segmentation differences between a host (e.g. FreeSurfer) and a ground truth (e.g. training set manually defined by an expert rater) which are predictable on the basis of the intensity, spatial, and contextual characteristics of mislabeled voxels in the brain image [Wang et al., 2011]. Consequently, using manual segmentations as a training set, the ASAT was designed to learn in which particular way an automated method like FreeSurfer commits segmentation errors, and then remove those errors to improve the accuracy of the segmentations. Excellent performance by the ASAT was demonstrated in Wang et al., [2011]. However, the ASAT has not been validated in pediatric samples, and there is a possibility that it may fail to improve segmentation adequately, or introduce additional biases related to young age.

The current research sought to examine whether the ASAT could correct FreeSurfer hippocampal segmentations in a sample of preschoolers, which included children with typical development as well as children with autism spectrum disorder (ASD). Structural brain abnormality and variability are present in developmental disorders such as ASD [Hua et al., 2013], and these may lead to poor performance by automated methods. The inclusion of diverse populations of typically and atypically developing children is important for a comprehensive test of segmentation methods. To achieve these goals, we conducted two studies. In Study 1, we conducted a cross-validation study of FreeSurfer and ASAT-corrected FreeSurfer segmentations of the hippocampus in a sample of typically and atypically developing ASD preschoolers. First, manual hippocampal segmentations were carried out by an expert rater in the entire sample. We then asked whether ASAT-

Abbreviations

ASAT	automatic segmentation adapter tool
ASD	autism spectrum disorder
DSCs	dice similarity coefficients
ICV	intracranial volume
ROIs	regions of interest

corrected segmentations produced significant improvements over FreeSurfer if trained using a relatively large or small set of training atlases, namely 36 versus 10. A demonstration of the effectiveness of ASAT with a small training set should make the method more accessible to a wider range of investigations. In Study 2, we used ASAT-corrected FreeSurfer segmentations in a larger sample of typically developing children and used the resulting hippocampal volumes to test for predicted associations with age, sex and expressive language.

STUDY 1

The goal of Study 1 was to establish the efficacy of the ASAT at correcting FreeSurfer generated hippocampal segmentations and maintaining the high levels of accuracy achieved with manual tracing, while incorporating the practical advantages of automated segmentation. To achieve this goal, we compared FreeSurfer-generated hippocampal segmentations that were either uncorrected or corrected with the Automatic Segmentation Adapter Tool [ASAT; Wang et al., 2011] to manual segmentations of the same sample in typically developing and atypically developing preschoolers, namely children with ASD.

The ASAT is trained to correct hippocampal segmentations from a set of training examples in which the hippocampus is segmented manually. In adults, the performance of ASAT segmentation improves as a function of the number of examples used to train the ASAT [Wang et al., 2011]. However, this improvement is not linear, and the marginal gain in performance with each additional training example becomes increasingly smaller [Wang et al., 2011]. Thus, we sought to compare the efficacy of the ASAT in error correction with a smaller (i.e., 10) versus a larger (i.e., 36) set of training examples. These two training set sizes were chosen to respectively represent moderate versus large training size; prior research in adult samples suggests that each training set size is sufficient for good performance [Wang et al., 2011], but this has not been established in young children. Evidence that a relatively small training size is sufficient for good performance in this population would be helpful to guide future research.

In addition to examining accuracy of FreeSurfer-generated and ASAT-corrected segmentations of the hippocampus as a whole, we were also interested in segmentation accuracy as a function of location along the anterior-posterior hippocampal axis. The examination of functional differences along this hippocampal axis has recently gained momentum [Poppenk et al., 2013; Strange et al., 2014], and there is initial evidence that structural development follows different trajectories in anterior versus posterior hippocampal regions [DeMaster et al., 2014; Gogtay et al., 2006]. Of importance, previous research highlights that FreeSurfer errors are particularly evident in the anterior portion of the hippocampal formation where the boundary with the amygdala is not always clear [Dewey et al., 2010]. Thus, it seemed imperative to establish whether the accuracy of segmentation and correction methods differs along this hippocampal axis.

Participants

Participants included 20 (16 males/4 females) typically developing children ($M = 2.98$, SD = 0.36, range: 2.27–3.60 years) and 20 (16 males/4 females) children with ASD ($M = 2.98$, SD = 0.48, Range: 2.26–3.70 years). This sample was pseudo-randomly selected from a larger cohort of participants assessed as a part of the Autism Phenome Project, conducted at the MIND Institute at the University of California, Davis (UC Davis), with the restriction that typically and atypically developing children were matched for sex and age. Typically developing children were excluded from participation for positive diagnosis of neurological or developmental delays, language impairments, or behavioral problems. ASD participants met the criteria established by the Collaborative Programs of Excellence in Autism, exceeding cutoffs for ASD diagnosis using ADOS-G [Autism Diagnostic Observation Schedule-Generic; Lord et al., 2000] and the Autism Diagnostic Interview–Revised [Lord et al., 1994], as assessed by a licensed clinical psychologist with expertise in Autism research. Informed consent was provided by the legal guardian(s) of the child. This research was approved by the UC Davis institutional review board.

Methods

Image acquisition

T1-weighted magnetization-prepared rapid gradient echo (MPRAGE) images were acquired at the UC Davis Imaging Research Center in a 3T Siemens Tim Trio scanner with an 8-channel head coil. (TR = 2,170 ms; TE = 4.86 ms; matrix: 256 × 256; slice thickness 1 mm, voxel: 1 mm isotropic). Motion artifacts were minimized by acquiring scans at night while children were asleep [Nordahl et al., 2008]. Post-acquisition, images were rigidly aligned to MNI standard-space using FLIRT from the FMRIB Software Library [FSL; Smith et al., 2004].

Manual tracing and atlas creation

Delineation of the boundaries of the left and right hippocampal formation was performed manually. We note that while we use the term hippocampus throughout this paper, the region that was segmented included the following cytoarchitectonic fields: dentate gyrus, hippocampus proper (CA1, CA2, CA3), subiculum, presubiculum, and parasubiculum. Segmentation of the hippocampus was conducted by an experienced rater (A.L.) using an established protocol [Schumann et al., 2004]. Left and right hippocampi were segmented separately, resulting in separate binary images for left and right hippocampus.

FreeSurfer host pipeline

Initial segmentation of hippocampal volumes were estimated for all participants using FreeSurfer [http://surfer.nmr.mgh.harvard.edu/; Version 5.1.0; Fischl, 2012] software suite of image analysis tools. The standard processing pipeline was employed [Fischl, 2012; Fischl et al., 2002]. Briefly, this involved motion correction, removal of non-brain tissue, bias field correction, affine and nonlinear registration with FreeSurfer's Talairach atlas, and propagation of labels to each voxel in reference to the prior probabilities specified in FreeSurfer's subcortical atlas [Fischl et al., 2002]. After segmentation and inspection for catastrophic registration errors, left and right hippocampal segmentations were extracted from the subcortical image and converted into binary NIfTI formatted images.

Automated segmentation adapter tool

The segmentation adapter command-line tool is an open-source, and freely available software [http://www.nitrc.org/projects/segadapter/; Wang et al., 2011]. The tool requires a set of structural brain images, reference segmentations (i.e. ground-truth), and host segmentations (the segmentations one wishes to correct).

Training parameters. To train the ASAT, several parameters must be specified. These parameters include: (1) the feature radius, which specifies the spatial neighborhood containing the features ASAT uses to identify segmentation errors; (2) the sampling rate, which specifies the proportion of voxels used to train the ASAT; (3) the number of training iterations, during which the ASAT is allowed to learn; and (4) the dilation radius, which specifies the radius that the host segmentation (e.g., FreeSurfer segmentation) is dilated to create working regions of interest (ROIs) for the ASAT training. In general, increasing the feature radius, sampling rate, or number of training examples only marginally improves ASAT performance [Wang et al., 2011], but can greatly increase the computational demands of the procedure. On the basis of Wang et al., [2011] and considering our goal of choosing parameters which would yield excellent performance while being within the capabilities of a typical analysis workstation, the sampling radius was set to $4 \times 4 \times 4$ voxels, the sampling rate was set at 50%, the number of training iterations was 500. The last parameter, the *dilation radius* is perhaps the most critical parameter to set appropriately. The purpose of this parameter is to grow the host segmentation (e.g. FreeSurfer) sufficiently to completely cover the manual/reference segmentation. ASAT is designed to identify and correct commission errors, but does not correct omission errors. Therefore, the dilation parameter will be most helpful if it is set liberally to ensure that all or the overwhelming majority (e.g., $>99\%$) of reference voxels are within the dilated host segmentation. In Study 1, we found that a dilation radius of 2 voxels was sufficient to ensure that dilated FreeSurfer segmentations encompassed all reference voxels. Further detail on the significance of these parameters and alternative settings are examined in Wang et al., [2011].

Cross-validation procedures

Cross validation procedures involve using manual segmentations as the ground truth to train the ASAT and correct the segmentations of the host (i.e., Free Surfer) in the remaining subsample. For each of our trainings sets (36 or 10), we conducted 10 cross-validation studies consistent with the literature [McLachlan et al., 2005]. Each training set was balanced for number of typically and atypically developing children and sex.

In the ASAT-36, each of the 10 cross-validation studies used segmentations for 36 of the participants' hippocampi to train and then correct the FreeSurfer-generated segmentations of the remaining four participants' hippocampi (i.e., this procedure was followed separately for left and right hippocampus). Thus, no hippocampus used for training was then used in the test sample within each of the cross-validation studies.

In the ASAT-10, each of the 10 cross-validation studies used segmentations for 10 of the participants' hippocampi to train the ASAT and then to correct the FreeSurfer-generated segmentations of four participants. We note that we maintained the set of to-be corrected segmentations equal to 4 on each of these cross-validation studies (even though in ASAT-10 a larger test sample would be possible) to be able to formally compare the results of ASAT-10 to ASAT-36. Again, no hippocampus used for training was then used in its corresponding test sample. These procedures produced a total of 40 corrected segmentations from 10 separately trained ASAT-36 and ASAT-10 algorithms per hemisphere.

Results and Discussion

Reliability of segmentations with the manual protocol was primarily assessed using dice similarity coefficients (DSCs; Dice, 1945). When applied to image segmentations, DSCs measure the spatial overlap of two independently traced segmentation and is computed as a ratio. Namely, $DSC=(2|A \cap B|)/(|A + B|)$, where A and B are the number of voxels in each individual segmentation, and $|A \cap B|$ is the number of voxels included in both segmentations. DSCs range between 0, indicating no agreement (i.e., no overlap), and 1, indicating perfect agreement (i.e., perfect overlap). DSC scores were computed using the Convert3d utility available at http://www.itksnap.org. Since it is important to measure potential systematic biases of each method with age and diagnosis, DSC scores obtained from FreeSurfer and ASAT methods using manually-segmented volumes as the criterion were compared. Descriptively (Fig. 1), it is clear that higher DSC scores were observed for each individual participant in

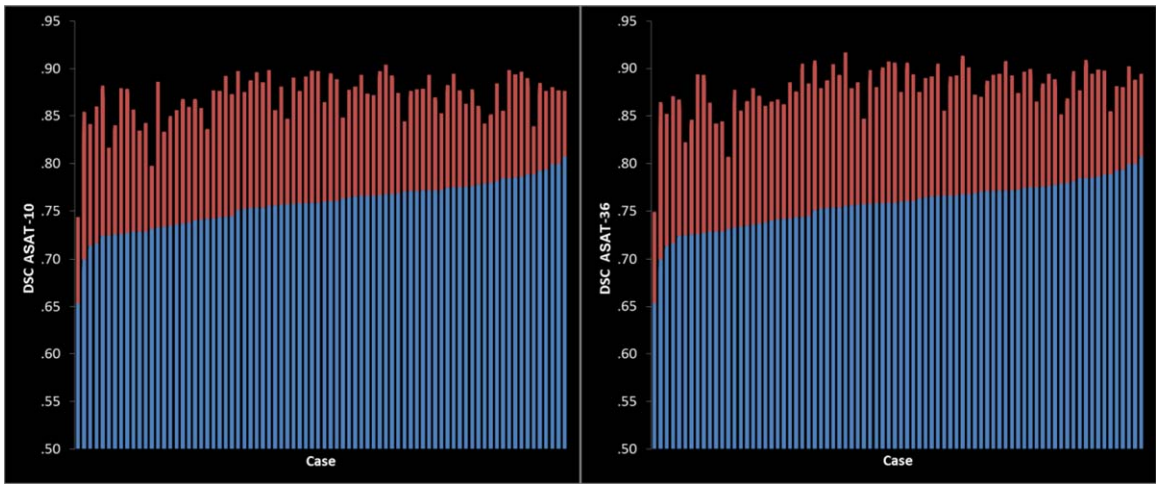

Figure 1.

Left: spatial overlap (DSC) of FreeSurfer (blue) and ASAT-10 (red) with manual tracings. Right: spatial overlap of FreeSurfer (blue) and ASAT-36 (red) with manual. Results for each case are sorted by FreeSurfer DSC. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

ASAT segmentations compared with FreeSurfer with an improvement exceeding 10% of voxel overlap as confirmed in the analyses below. Example slices of ASAT-10, FreeSurfer, and manual segmentations are illustrated in Figure 2. As seen in Figure 3, ASAT-10 performed well in each of its cross-validation runs, suggesting that despite

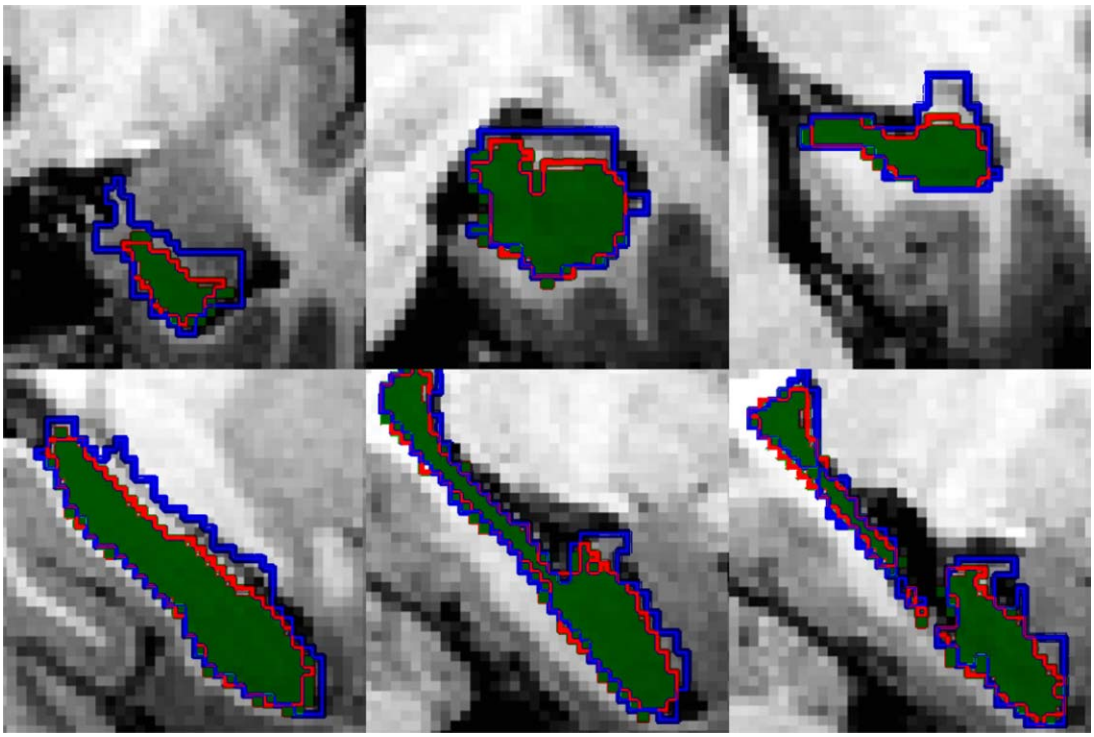

Figure 2.

Example coronal (top) and sagittal (bottom) views of left hippocampal segmentations for FreeSurfer (blue outline), ASAT-10 (red outline), and manual tracings (green). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

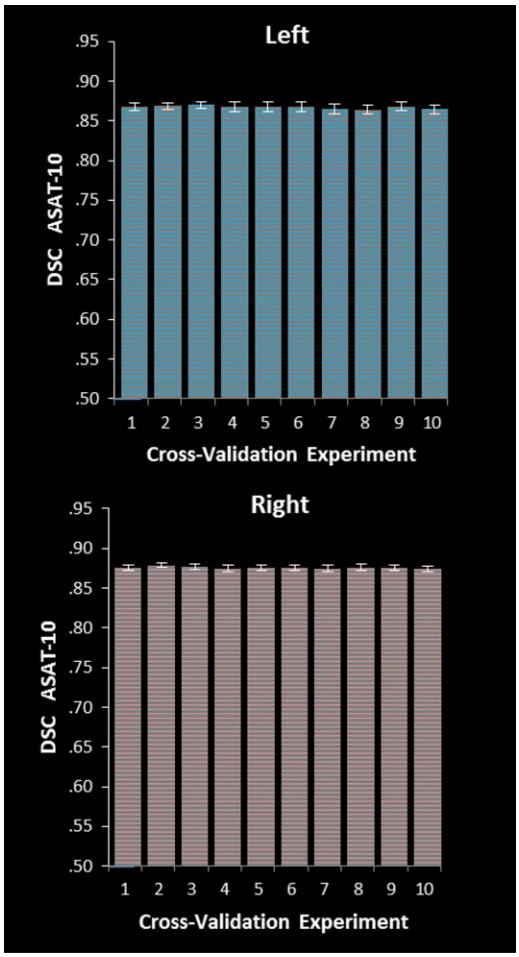

Figure 3.

Mean and standard errors for DSC of ASAT-10 on each cross-validation run in left and right hippocampus. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

ASAT-10 using only 10 training examples, performance was not sensitive to the particular set of examples used.

Hippocampal size may differ as a function of age, sex, hemisphere and participant group, and it is important to establish the extent to which segmentation techniques are biased because of any of these variables of interest. For a formal examination of these potential differences, we conducted a 2 (age: 2-year-olds versus 3-year-olds) × (participant group: ASD, Typical) × 2 (hemisphere: left, right hippocampus) × 3 (method: FreeSurfer, ASAT-10, ASAT-36) multivariate analysis of variance (repeated-measures MANOVA). Age-groups were constructed by dividing the sample into 2-year-olds ($M = 2.65$, SD = .29, range = 2.6–2.99 years, $n = 20$), and 3-year-olds ($n = 20$, $M = 3.31$, SD = .23, range = 3.01–3.70, $n = 20$). Table I reports DSC means and standard deviations for each method by diagnosis and age group. Follow-up multiple comparisons of main effects were examined using Bonferroni corrected *P*-values.

Results revealed a significant main effect of method, $F(2, 35) = 860.66$, $P < 0.001$, Wilks' $\lambda = 0.02$, $\eta_p^2 = 0.98$. Both ASAT-36 and ASAT-10 achieved higher DSC scores than FreeSurfer (ASAT-36, Mean difference = 0.12, $P \leq .001$, 95% CI [0.12, 0.13] and ASAT-10, mean difference = 0.11, $P \leq 0.001$, 95% CI [0.11, 0.12]). The ASAT-36 resulted in reliably higher DSCs than ASAT-10, difference = 0.01, $P \leq 0.001$, 95% CI [0.006, 0.010]. However, it is worth noting that this difference is a magnitude smaller than the differences between ASAT and FreeSurfer. There also was a significant effect of hemisphere, $F(1, 36) = 5.60$, $P = 0.024$, Wilks' $\lambda = 0.87$, $\eta_p^2 = 0.14$. All segmentation methods achieved higher DSCs in the right hippocampus than in the left, difference = 0.01, $P = 0.02$, 95% CI [0.001, 0.015]. There was no main or interactive significant effect of age and participant group, $Fs \leq 2.6$, $ps \geq 0.09$, $\eta_p^2 s \leq 0.13$.

Although segmentation accuracy did not differ as an effect of age-group in the prior analysis, we further examined whether age is related to segmentation accuracy when considered as a continuous variable to ensure that the age group subdivision did not obscure overall associations with age. Conducting Pearson's correlations between age and FreeSurfer, ASAT-10, and ASAT-36 left and right segmentations did not reveal a significant relations, $|r|s \leq 0.10$, uncorrected $ps \geq 0.56$. This result confirms the result from the MANOVA analysis.

We next examined whether performance of FreeSurfer and ASAT differed along the anterior to posterior axis of the hippocampus. For this analysis, we focused on performance of FreeSurfer and ASAT-10, given its similarity to ASAT-36 performance and the potential of this type of correction to be helpful even in small-scale studies. For brevity, we report the analysis of the right hippocampus, but the degree of overlap between FreeSurfer and ASAT segmentations along the axis is virtually identical in the left hippocampus.

Coronal slices from FreeSurfer, ASAT-10, and manually traced volumes were contrasted (Fig. 4). To compare slices across participants, volumes were aligned to the most anterior extent of the hippocampal head. Since the anterior-posterior length of the hippocampus is different in different people, only slices for which all participants contributed estimated volumes were evaluated. We computed difference scores for each slice by subtracting the manually segmented volume from the FreeSurfer and ASAT-10 slice volumes and examined these differences in a 2 (Method: FreeSurfer, ASAT-10) × 22 (Slice: anterior #1 to posterior #22 hippocampus) MANOVA. Results revealed significant main effects of method and slice, $Fs \geq 7.39$, $ps \leq 0.001$, Wilks' $\lambda \leq 0.11$, $\eta_p^2 \geq 0.89$, which were qualified by a significant method × slice interaction, $F(21,19) = 13.44$, $P < 0.001$, Wilks' $\lambda = 0.06$, $\eta_p^2 > 0.94$.

A significant linear contrast with slice × method was observed, $F(1,39) = 98.04$, $P < 0.001$, $\eta_p^2 = 0.72$. Following up, a significant linear contrast with slice was observed for the differences between FreeSurfer and manual slice volumes,

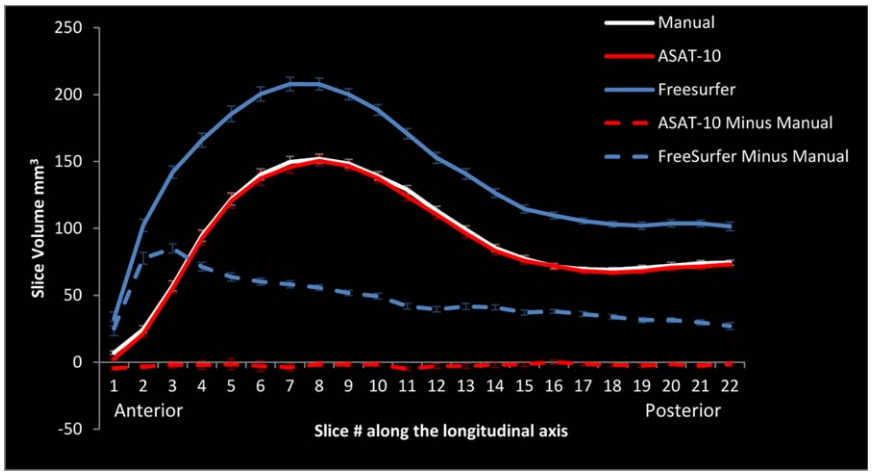

Figure 4.

Average slice volumes and standard errors in right hippocampus of FreeSurfer (solid blue), ASAT-10 (solid red), and manual segmentations (solid white) along the longitudinal axis of the hippocampus. Differences between automated and manual slice volumes are depicted with dashed lines. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$F(1,39) = 96.92$, $P \leq 0.001$, $\eta_P^2 = 0.7140$, with the biggest differences occurring in regions of anterior hippocampus. Notably, the biggest differences between FreeSurfer and manual tracings were evident in the portions corresponding to the hippocampal head where erroneous inclusion of voxels from the amygdala and ventricles has been previously reported in adults [Dewey et al., 2010]. However, the linear contrast did not reach statistical significance for ASAT-10, $F(1,39) = 0.21$, $P = 0.65$, $\eta_P^2 = 0.005$. As evident from Figure 4, the difference between ASAT-10 and manual tracing was minimal and unbiased along the anterior-posterior axis, and did not differ from zero. As a last analysis, we confirmed ASAT-36 would perform similarly as ASAT-10 did along the anterior to posterior axis. As expected, all results replicated those reported for ASAT-10; we again observed a slice by method interaction, $P < 0.001$, a significant slice $\times$ method linear contrast, $P < 0.001$, such that the linear contrast was evident in FreeSurfer, $P < 0.001$, but was not in ASAT-36 slices, $P = 0.44$.

Segmentations produced by ASAT in Study 1 were neither biased nor negatively affected by age, participant group, or hemisphere. Further, comparison of ASAT segmentations with their manual reference segmentations across 1 mm slices suggested that ASAT performance was consistent across the entire longitudinal axis of the hippocampus. In contrast, the differences in volume between FreeSurfer and the manual reference segmentations were greatest in regions of the anterior hippocampus; this finding is consistent with prior research [Dewey et al., 2010] in which FreeSurfer commonly included ventricle space and amygdala into the hippocampal head region. Thus, ASAT seems capable of consistently removing these errors. We acknowledge that some of the differences between manual and FreeSurfer segmentations are because of differences in segmentation protocol (e.g. inclusion of alveus, a submillimeter layer of white matter). However, these minor protocol differences cannot account for the substantial differences in volume which are typically in the range of several thousand cubic millimeters. Overall these results demonstrate the high validity and reliability of ASAT to correct FreeSurfer hippocampal segmentation with as few as 10 training examples and which do not appear to be obviously biased by age, participant status, or hemisphere. Given these results, we conducted Study 2 as a first application of ASAT-10-corrected volumes to investigate early brain development as well as one important domain of cognition, namely language.

TABLE I. DSC by method, hemisphere, status, and age-group

	Left hippocampus				Right hippocampus			
	Typical		ASD		Typical		ASD	
DSC	2 year-olds	3 year-olds	2 year-olds	3 year-olds	2 year-olds	3 year-olds	2 year-olds	3 year-olds
FreeSurfer	0.76 (0.03)	0.76 (0.02)	0.75 (0.03)	0.77 (0.02)	0.76 (0.03)	0.76 (0.02)	0.74 (0.03)	0.75 (0.02)
ASAT-10	0.87 (0.03)	0.88 (0.02)	0.87 (0.03)	0.88 (0.02)	0.87 (0.03)	0.88 (0.02)	0.86 (0.03)	0.87 (0.02)
ASAT-36	0.87 (0.03)	0.88 (0.02)	0.88 (0.03)	0.89 (0.02)	0.88 (0.03)	0.89 (0.02)	0.87 (0.03)	0.88 (0.02)

The values denote the mean and values within parenthesis are standard deviation values.

STUDY 2

The goal of Study 2 was to provide an initial application of ASAT to investigations of early brain and cognitive development. We chose to apply ASAT using only 10 training examples because of the high performance of ASAT-10 in Study 1 and because this modestly sized training set may find application in even relatively small scale neuroimaging studies. Specifically, we examined whether these volumes could be used to yield predicted associations between these and age, sex, and expressive language.

The available data suggest age and sex related differences in hippocampal volume during the first years of life, but direct evidence is relatively scant. For example, Uematsu et al., [2012] reported age-related increases in hippocampal volume over the first 5 years of life; furthermore, they found larger hippocampi in males compared to females. While a strength of Uematsu et al. [2012] is that hippocampal segmentation was conducted manually, these relations were reported both using volumes uncorrected for overall intracranial volume (ICV) or corrected using a procedure which corrected for individual variation of ICV within each age group, but not across ages, The latter approach was motivated by the fact that there are substantial age-differences in ICV which may derive from different factors at different ages and that a uniform correction across a wide range may not be appropriate. Although this is a valid concern, the proposed solution may effectively reduce the scope of the correction and limit our current ability to assess whether age differences in hippocampal volume survive after accounting for overall differences in ICV. Nevertheless, on the basis of this limited evidence we predicted that we would also observe larger uncorrected hippocampal volumes in older compared with younger children and in males compared with females. We further predicted age-related differences would be observed in hippocampal volume after accounting for ICV and potential interactions with age.

We also applied ASAT-10 corrected volumes to the investigation of how hippocampal structure might support cognition in early development. While the importance of the hippocampus in supporting memory and spatial navigation in adults and children is well recognized [Burgess et al., 2002; Ghetti and Lee, 2011], the hippocampus is hypothesized to act in the service of other forms of cognition including certain aspects of language [Cohen, 2015; Duff and Brown-Schmidt, 2012]. [TQ1]In adults, hippocampal activation has been associated with learning novel vocabulary [Breitenstein et al., 2005; Jung et al., 2014]. Hippocampal amnesia in adults has been associated with less cohesive verbal communication skills [Kurczek and Duff, 2011], including an inability to relate various details within even short verbal reports [Kurczek et al., 2013]. Further, tasks assessing semantic verbal fluency have been associated with differential hippocampal activation [Glikmann-Johnston et al., 2015]. Learning of new vocabulary has been associated with longitudinal increase in hippocampal volume [Mårtensson et al., 2012]. However, to date little research has examined whether the hippocampus is associated with verbal ability during the preschool years. There are few exceptions. For example, using whole brain voxel-based morphometry in a small sample of infants, Deniz Can et al. [2013] reported a positive relation between right hippocampal volume at 7 months of age and later expressive language production at 12 months of age, as measured by the Mullen Scales of Early Development (MSEL). The availability of the MSEL language assessments in the present sample allowed us to examine whether an association between hippocampal structure and expressive language ability persists after infancy during the preschool years. There is currently a real paucity of research examining the associations between hippocampal structure and cognitive development during the preschool years, and thus the present application of the ASAT method begins to contribute to this literature.

Participants

Eighty-nine (59 male/30 female) typically developing children aged 2.23–4.73 years, $M = 3.12$, SD = 0.54, participated in Study 2 as part of the Autism Phenome Project. Typically developing children participating in Study 1 also participated in Study 2. Inclusion and exclusion criteria were the same as in Study 1. Participants came from a diverse racial and ethnic background: 5.8% identified themselves as African-American, 10.5% Asian or Pacific Islander, 55.8% non-Hispanic Caucasian, 16.3% Hispanic Caucasian, 3.5% of mixed race, while 8.1% of participants declined to identify. Three male participants failed to complete the language development assessment; all participants successfully completed the MRI portion of the study.

Methods

Behavioral measures of language development

Expressive and receptive language abilities were assessed by the Mullen Scales of Early Development [MSEL; Mullen, 1995] using raw scores from the expressive and receptive language subscales. MSEL developmental quotients for overall (DQ), verbal (VDQ), and nonverbal (NVDQ) intellectual ability were also computed.

Image acquisition and processing

Acquisition of images, preprocessing, and FreeSurfer analysis were performed as described in Study 1. ASAT-10 algorithms for each hemisphere based on manual tracing from Study 1 were used to correct FreeSurfer segmentations in Study 2.

Intracranial volume

Volumes of each hippocampus was adjusted by an estimated intracranial volume (ICV). ICV estimates were

TABLE II. Age and sex differences in left and right hippocampal volume

	ASAT-10			
Left hippocampus	β	SE B	t	P
Constant		29.98	86.420	<0.001
Age (z-score)	0.09	29.55	0.92	0.36
Sex (M = −1, F = 1)	−0.30	31.54	−2.90	0.005
ICV (z-score)	0.36	32.52	3.23	0.002
Age × sex	−0.15	32.74	−1.34	0.19
ICV × age	−0.18	32.69	−1.62	0.11

Model $F(81, 5) = 8.76$, $P < 0.001$, adjusted-$R^2 = 0.31$

	ASAT-10			
Right hippocampus	β	*SE B*	t	P
Constant		28.45	80.01	<0.001
Age (z-score)	−0.13	28.05	−1.30	0.20
Sex (M = −1, F = 1)	−0.23	29.93	−2.27	0.03
ICV (z-score)	0.43	30.87	3.84	<0.001
Age × sex	−0.23	31.08	−2.07	0.04
ICV × age	−0.21	31.03	−1.90	0.06

Model $F(81,5) = 8.37$, $P = 0.001$, adjusted-$R^2 = 0.30$

obtained from the structural images using the procedure outlined in Lee, et al. [2014]. Briefly, bias-corrected brain images were skull-stripped using BET, linearly registered to a standard template, and the inverse of the determinant of the resulting affine matrix was computed, giving an estimate of ICV.

Results and Discussion

In the preliminary analyses, we confirmed that ASAT-10 continued to have excellent specific agreement with ASAT-36 in this expanded sample as indicated by high intra-class correlation coefficients for absolute agreement of a single measurement in left, ICC = 0.97, and right hippocampus, ICC = 0.98. Preliminary analyses also revealed that verbal and nonverbal intellectual abilities of the sample as measured by the MSEL Developmental Quotient (DQ) were within a fairly typical range (MSEL DQ: $M = 107$, SD = 11.5, range = 82–132; verbal DQ: $M = 108$, SD = 12.3, range = 80–139; nonverbal DQ: $M = 106$, SD = 14.5, range = 73–139).

We began by examining age and sex-related differences in hippocampal volume using multiple regressions without accounting for the contribution of ICV. Left and right hippocampal volumes were separately regressed on age, sex (male = −1, female = 1), and age × sex interaction term. Consistent with Uematsu et al. [2012] larger hippocampal volumes were observed in males compared with females, βs ≤ −0.39, SEs ≥ 28.85, ts ≤ −3.93, ps < 0.001, and larger hippocampal volumes were observed in older compared with younger children in left, $\beta = 0.20$, SE = 28.15, $t = 2.05$, $P = 0.04$, but not right hippocampus, $\beta = 0.00$, $P = 0.99$, the latter of which was potentially qualified by a marginal age x sex interaction, $\beta = -0.18$, SE = 28.47, $t = -1.76$, $P = 0.08$.

We then examined age and sex related differences in volume while including overall intracranial volume in the regression models. Left and right hippocampal volumes were separately regressed on age, sex (male = −1, female = 1], ICV, age × sex, and ICV × age interaction terms. The age × ICV interaction term was included to account for the possibility that ICV may affect hippocampal volume differently as a function of age as suggested by Uematsu et al. [2012]. Results for left and right hippocampus for these regressions are reported in Table II. Consistent with the prior result, larger hippocampal volumes were observed in males compared with females even when ICV was included in the model. Potentially inconsistent with Uematsu et al. [2012], age alone was not a reliable predictor of left hippocampal volume above the effect of ICV. However, age was qualified by sex as a predictor of right hippocampal volume such that right hippocampal volumes were smaller in older than in younger female children, but not in male children when ICV was included in the model (Fig. 5). Further ICV and age marginally interacted in right ($P = 0.06$) and left ($P = 0.11$) hippocampus (and the interaction was statistically significant at $P < 0.05$ with left and right hippocampal volumes averaged), such that ICV was more positively related to hippocampal volume in younger than older children. Consistent with Uematsu et al., [2012] this suggests that a linear correction of hippocampal volumes by ICV may not be appropriate for all populations, especially those undergoing rapid brain development. Taken together however,

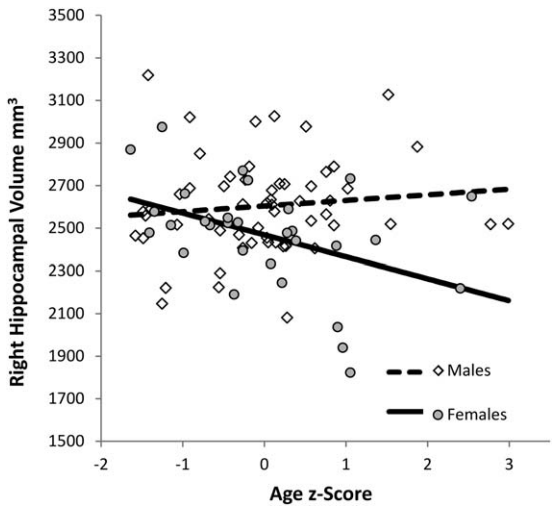

Figure 5.
Partial residual plots of the interaction between age (z-score) and sex (male = −1, female = 1) predicting right hippocampal volume as evaluated at median ICV (z = 0.04).

TABLE III. Volume left hippocampus and expressive language

	ASAT-10			
Left hippocampus	β	SE B	t	P
Constant		0.78	50.54	<0.001
Age (z-score)	0.80	0.69	11.36	<0.001
Sex (M = −1, F =1)	0.14	0.79	1.75	0.09
Volume (z-score)	0.13	0.81	1.50	0.12
ICV (z-score)	0.11	0.78	1.31	0.20
Sex × age	−0.04	0.76	−0.48	0.63
Age × volume	0.24	0.86	2.75	0.007
Volume × sex	0.12	0.83	1.56	0.12
Age × ICV	−0.24	0.91	−2.57	0.01
Volume × ICV	0.05	0.83	0.63	0.53

Model $F(76,9) = 20.44$, $P < 0.001$, adjusted-$R^2 = 0.67$.

these data suggest that within the range of two to four years of age, subtle age-and sex- related differences in hippocampal volume may be observed.

The failure to observe clear age-related differences in the right hippocampus after accounting for ICV seems inconsistent with Uematsu et al. [2012]. However, as noted previously the method of correction for ICV by Uematsu et al. [2012] was unusual in that it only accounted for individual variability within age group and not across ages. More generally, differences in age-related trajectories in the right and left hippocampus at different points in development are not unprecedented [Gogtay et al., 2006]. It is conceivable that discrepant trajectories may be seen in early childhood, especially when examining a relatively narrow developmental periods.

We then examined the relation between hippocampal volume and early language development with multiple linear regression, regressing expressive or receptive language scores from the MSEL on age (z-score), sex (male = −1, female = 1), left or right hippocampal volumes (z-score), ICV (z-score), sex × age, age × volume, volume × sex, age × ICV, and volume × ICV. These interaction terms were included because findings from the previous regressions indicated that hippocampal volume was best predicted when interaction terms among the predictors were included. Thus, we took a similar approach here and ensured that these potential interactions were accounted for in our regressions predicting language. We first examined early expressive language ability. Results for left and right hippocampus are reported in Tables III and IV, respectively. Expressive language ability was greater in older than younger children, and this age-related difference was greater in children with larger hippocampi (Fig. 6). Alternatively this interaction could be interpreted such that the relation between hippocampal volume and expressive language increases with age. We note that in the regression with left hippocampal volume, the relation between ICV and expressive language was moderated by age such that ICV was a more positive predictor of expressive language ability in younger than in older children. We next examined early receptive language development. Results for regression on left hippocampus are presented in Table V, which revealed that the effect of volume was moderated by sex, such that left hippocampal volume positively predicted receptive language in females, but not in males (Fig. 7). Regressions of receptive language on right hippocampal volume revealed positive effects of age, $\beta s \geq 0.86$, $ps < 0.001$, but no main, or interactive relations with volume, $|\beta s| \leq 0.11$ $ps \geq 0.15$. Finally, the availability of a nonverbal DQ score from the MSEL battery allowed us to additionally test whether the relations we observed with language would extend to a non-verbal index of intellectual ability. Using the same analytical approach as before, regressions on left or right hippocampus, revealed significant positive effects of age, $\beta s \geq 0.31$, $ps \leq 0.01$, and sex, $\beta s \geq 0.27$, $ps \leq 0.04$; however these analyses revealed no main, or interactive effects with hippocampal volume, $|\beta s| \leq 0.18$, $ps \geq 0.16$, or ICV, $|\beta s| \leq 0.19$, $ps \geq 0.14$.

Overall these results provide further evidence of a connection between hippocampal structure and early language ability. Deniz Can et al. [2013] found a correlation between right hippocampal structure in infancy and later expressive language at 12 months. However, unlike Deniz Can et al. [2013], we also found hippocampal brain structure moderated age-related improvements in expressive language scores, such that age-related improvements were greater in children with larger hippocampi. Finally, and in contrast to Deniz Can et al. [2013], we found a relation between receptive language and right hippocampal structure.

It is difficult to establish the reasons underlying the Deniz Can et al. [2013] null finding in left hippocampus as our study differs from theirs on a number of dimensions. Differences in results may be due to differences in the nature of the relations examined (i.e., concurrent relations in our study and longitudinal relations in Deniz Can et al.

TABLE IV. Volume right hippocampus and expressive language

	ASAT-10			
Right hippocampus	β	SE B	t	P
Constant		0.79	50.39	<0.001
Age (z-score)	0.78	0.72	10.66	<0.001
Sex (M = −1, F = 1)	0.12	0.79	1.56	0.12
Volume (z-score)	0.02	0.78	0.21	0.83
ICV (z-score)	0.14	0.82	1.69	0.09
Sex × age	−0.05	0.78	−0.57	0.57
Age × volume	0.18	0.73	2.38	0.02
Volume × sex	0.03	0.83	0.34	0.74
Age × ICV	−0.13	0.80	−1.60	0.11
Volume × ICV	−0.09	0.77	−1.15	0.25

Model $F(76,9) = 19.35$, $P < 0.001$, adjusted-$R^2 = 0.66$.

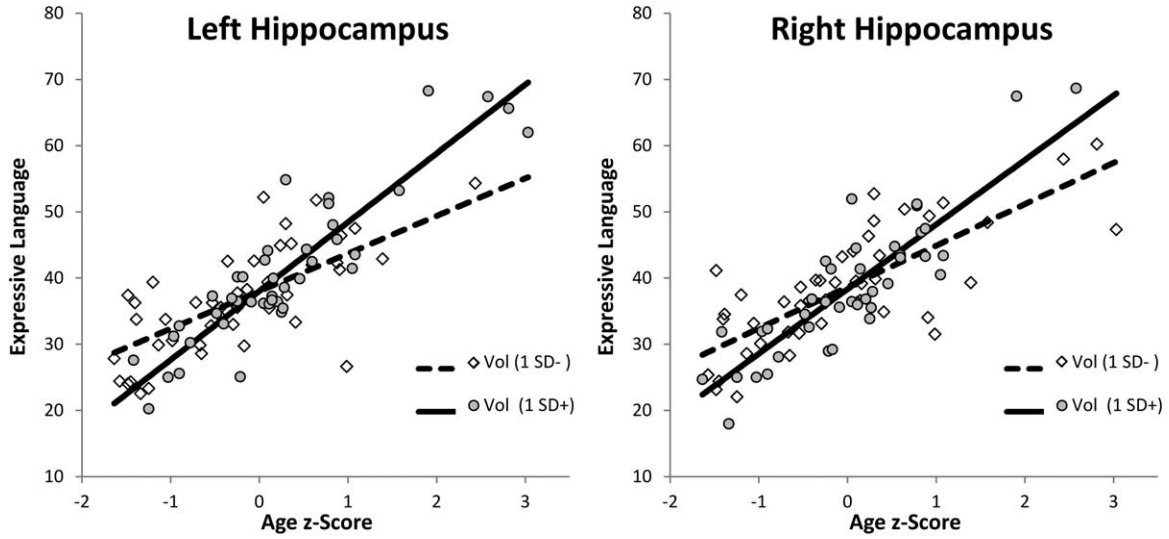

Figure 6.

Partial residual plots of the interactions between age (z-score) and volume (z-scores) predicting expressive language ability for left hippocampus (on left) and right hippocampus (on right). Interactions are plotted at ± 1 SD of hippocampal volume and. evaluated at median ICV (z = 0.054) and sex (male = −1).

[2013], differences in assessment tools (i.e., volumetric assessments in the present study and whole brain voxel-based morphometry approach in Deniz Can et al. 2013], or differences in analytical approaches (i.e., inclusion of interaction terms in our study, exclusion of these terms in Deniz Can et al. 2013]. Despite these differences, overall our findings were consistent with our prediction of a relation between hippocampal structure and early language ability.

While this research examined relation between hippocampal structure and language ability, we cannot determine from these data what underlies these statistical relations or how the hippocampus specifically contributes to early language ability or development. We attempted to address this potential limitation by examining a nonverbal measure of intellectual ability, and failed to detect any reliable relations with hippocampal volume. Although this result suggests that hippocampal structure may contribute to language development beyond a general contribution to intellectual ability, the null findings should not be construed as evidence that hippocampal structure during early development does not contribute to early nonverbal forms of cognition (e.g. deferred imitation of sequences; Adlam et al., 2005], and future work is necessary to further investigate the role of the hippocampus to cognitive development.

TABLE V. Volume left hippocampus and receptive language

	ASAT-10			
Left hippocampus	*B*	SE B	*t*	*P*
Constant		0.63	58.73	<0.000
Age z-score	0.84	0.55	12.64	<0.001
Sex (M = −1, F =1)	0.10	0.64	1.39	0.17
Volume (z-score)	0.03	0.65	0.38	0.71
ICV (z-score)	0.13	0.63	1.66	0.10
Sex × age	0.04	0.62	0.48	0.64
Age × volume	−0.03	0.69	−0.35	0.73
Volume × sex	0.15	0.68	2.05	0.044
Age × ICV	−0.13	0.74	−1.50	0.14
Volume × ICV	0.15	0.67	1.96	0.054

Model $F(76,9) = 23.86$, $P < 0.001$, adjusted-$R^2 = 0.71$.

We note that same analyses conducted In Study 2 were also conducted using FreeSurfer segmentations and manually corrected ASAT volumes produced by rater AL. The details of these supplemental analyses are included as Supporting Information in a separate document. Overall, results using FreeSurfer and manually corrected ASAT volumes were analogous to those reported in the main manuscript (Supporting Information Tables I–VI). These results might suggest that FreeSurfer segmentations may function similarly as do more accurate segmentations methods (e.g., ASAT-10). However, we also included additional analyses that suggested that the magnitude of segmentation errors committed by FreeSurfer differed as a function of size of the hippocampus (see Supporting Information, p. 3), such that the discrepancy between FreeSurfer and manual segmentations was greater with larger manually corrected hippocampi. No reliable relation was detected for the much smaller errors committed by ASAT-10.

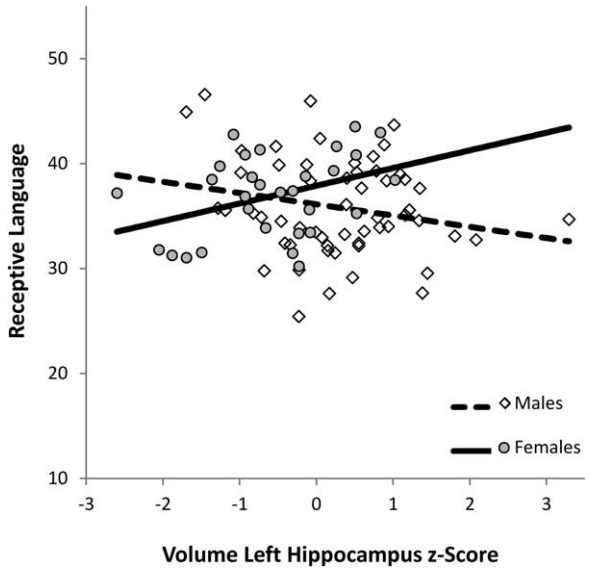

Figure 7.

Partial residual plots of the interaction between left hippocampal volume (z-score) and sex (male = −1, female = 1) predicting receptive language ability, as evaluated at the median of ICV (z = 0.04) and age (z = 0.00).

These results warrant caution in using FreeSurfer hippocampal segmentation for two reasons. First, given the many factors that have been found to predict the size of the hippocampus [e.g., age in typical development, Wierenga et al., 2014; autism spectrum disorder, Schumann, et al., 2004, hypoxic-ischemic injury, Gadian et al., 2000; premature birth, Nosarti et al., 2002], the magnitude of error FreeSurfer segmentations may differ depending on the kind of population observed and the factors examined. Second, in analyses similar to ours, Wenger et al. [2014] reported the opposite finding in a sample of older adults, such that FreeSurfer errors were negatively correlated to the volumes of manual segmentations in older adults, while in a sample of middle-aged adults no reliable relation was found between errors and manual volumes. These results taken together suggest that segmentation errors by FreeSurfer are not committed consistently across age-groups. Given the relative magnitude of FreeSurfer error in comparison to the volume of manual segmentations (i.e. ∼49%), the heterogeneity in the extent and direction of error biases by FreeSurfer present serious threats to the validity of results using FreeSurfer volumes, at least if contrasting different age-groups.

Overall results of Study 2 suggest that ASAT-10 corrected volumes may be useful in detecting relations with age and cognitive development, suggesting that just 10 training examples may be sufficient to obtain high reliable and valid segmentation that can be used fruitfully for investigations of associations between hippocampal volume and cognitive functioning.

GENERAL DISCUSSION

These studies validated and applied a recently devised tool, the Automatic Segmentation Adapter Tool [ASAT; Wang et al., 2011] to volumetric analyses of the hippocampus in young children. In Study 1, ASAT was shown to substantially improve and adapt hippocampal segmentations produced by FreeSurfer in a sample of ASD and TD of 2- and 3-year-old children. The method requires as few as 10 training examples to achieve excellent performance. Training the ASAT's Adaboost classifiers took less than a day on our analysis computer, and applying each correction took one or two seconds to perform. Moreover, the algorithm can be used repeatedly in the future with new data sets with similar participant demographics. Improvements in the DSC overlap were both substantial and exceedingly reliable; improvements were seen in 100% of the cross-validation cases in both left and right hippocampus. We note that DSC spatial overlaps that are achieved by ASAT closely approach those achieved by the gold-standard method with expert manual raters [Wang and Yushkevich, 2012].

Overall, these results show that while FreeSurfer's standard hippocampal segmentations were not biased with respect to age, participant status, or hemisphere, they did include a substantial number of mislabeled voxels, and that these errors were particularly strong in anterior hippocampal regions. However, with as few as 10 training examples ASAT was capable of substantially improving segmentation accuracy without introducing biases based on age, diagnosis status, or location along the longitudinal axis of the hippocampus. ASAT accuracy improvements were strong and highly reliable such that improvement was seen in every case across ASAT-10 and ASAT-36 cross-validation studies, in both left and right hippocampal segmentations. Finally, although we examined ASAT performance only in the context of correcting hippocampal segmentations produced by FreeSurfer, there is no reason to expect that ASAT cannot be trained to correct systematic segmentation errors in other brain regions (e.g. amygdala or cerebellum), or by another automated method (e.g., FSL FIRST).

Study 2 applied ASAT-10 corrected segmentations to examine brain-behavior associations in a sample of 2-, 3-, and 4-year-old children. Specifically, we examined relations between hippocampal structure and age, sex, and expressive/receptive language development. Uematsu et al. [2012] reported age-related increases in raw hippocampal volume in the first 4 to 5 years of life and overall larger hippocampal volumes in male participants. Here, we investigated age and sex-related differences in hippocampal volume in a narrower age-range of 2-, 3-, and 4-year-old children, and in part replicated these findings. Age-related increases were observed in raw volume of left hippocampus, and males had larger left and right raw hippocampal volumes than females. Analyses that accounted for the contribution of ICV however qualified relations

with age by sex, and potentially by ICV. Taken together, these data suggest that within the range of two to four years of age, subtle age-and sex- related differences in hippocampal volume may be observed. Data from the adult literature have documented relations between language and hippocampal structure and function [e.g. Cohen, 2015; Duff and Brown-Schmidt, 2012). Initial evidence using voxel-based morphometry in infants [Deniz Can et al., 2013] suggested these relations also exist between early language development and hippocampal structure in infancy. Here we also found relations between language ability and hippocampal structure in early development, suggesting that the hippocampus may be important to language development after infancy. Overall, these results provide initial evidence that ASAT might be useful to investigations of early brain development.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Autism Phenome Project staff for helping with the logistics of family visits and data collection. They especially thank all of the families and children who participated in the study. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors declare no competing financial interests.

REFERENCES

Adlam ALR, Vargha-Khadem F, Mishkin M, De Haan M (2005): Deferred imitation of action sequences in developmental amnesia. J Cognitive Neurosci 17:240–248.

Breitenstein C, Jansen A, Deppe M, Foerster AF, Sommer J, Wolbers T, Knecht S (2005): Hippocampus activity differentiates good from poor learners of a novel lexicon. Neuroimage 25:958–968.

Burgess N, Maguire EA, O'Keefe J (2002): The human hippocampus and spatial and episodic memory. Neuron 35:625–641.

Cohen NJ (2015): Navigating Life. Hippocampus 25:704–708.

DeMaster D, Pathman T, Lee JK, Ghetti S (2014): Structural development of the hippocampus and episodic memory: Developmental differences along the anterior/posterior axis. Cerebral Cortex 24:3036–3045.

Deniz Can D, Richards T, Kuhl PK (2013): Early gray-matter and white-matter concentration in infancy predict later language skills: A whole brain voxel-based morphometry study. Brain Lang 124:34–44.

Dewey J, Hana G, Russell T, Price J, McCaffrey D, Harezlak J, Tate DF (2010): Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. Neuroimage 51:1334–1344.

Duff MC, Brown-Schmidt S (2012): The hippocampus and the flexible use and processing of language. Front Hum Neurosci 6:69

Evans AC, Janke AL, Collins DL, Baillet S (2012): Brain templates and atlases. Neuroimage 62:911–922.

Everaerd D, Gerritsen L, Rijpkema M, Frodl T, van Oostrom I, Franke B, Tendolkar I (2012): Sex modulates the interactive effect of the serotonin transporter gene polymorphism and childhood adversity on hippocampal volume. Neuropsychopharmacology 37:1848–1855.

Fischl B (2012): FreeSurfer. Neuroimage 62:774–781.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Dale AM (2002): Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33:341–355.

Freund Y, Schapire RE (1995): A desicion-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119–139.

Gadian DG, Aicardi J, Watkins KE, Porter DA, Mishkin M, Vargha- KF (2000): Developmental amnesia associated with early hypoxic–ischaemic injury. Brain 123:499–507.

Ghetti S, Lee J (2011): Children's episodic memory. Wiley Interdiscip Rev Cogn Sci 2:365–373.

Glikmann-Johnston Y, Oren N, Hendler T, Shapira-Lichter I (2015): Distinct functional connectivity of the hippocampus during semantic and phonemic fluency. Neuropsychologia 69: 39–49.

Gogtay N, Nugent TF, Herman DH, Ordonez A, Greenstein D, Hayashi KM, Thompson PM (2006): Dynamic mapping of normal human hippocampal development. Hippocampus 16:664–672.

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Fischl B (2006): Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. Neuroimage 32:180–194.

Hua X, Thompson PM, Leow AD, Madsen SK, Caplan R, Alger JR, Levitt JG (2013): Brain growth rate abnormalities visualized in adolescents with autism. Hum Brain Mapp 34:425–436.

Jovicich J, Czanner S, Han Salat D, van der Kouwe A, Quinn B Fischl B (2009): MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 46:177–192.

Jung RE, Ryman SG, Vakhtin AA, Carrasco J, Wertz C, Flores RA (2014): Subcortical correlates of individual differences in aptitude. PLoS One 9:e89425.

Koolschijn PCM, Crone EA (2013): Sex differences and structural brain maturation from childhood to early adulthood. Dev Cogn Neurosci 5:106–118.

Kumfor F, Sapey-Triomphe LA, Leyton CE, Burrell JR, Hodges JR, Piguet O (2014): Degradation of emotion processing ability in corticobasal syndrome and Alzheimer's disease. Brain 137: 3061–3072.

Kurczek J, Duff MC (2011): Cohesion, coherence, and declarative memory: Discourse patterns in individuals with hippocampal amnesia. Aphasiology 25:700–712.

Kurczek J, Brown-Schmidt S, Duff M (2013): Hippocampal contributions to language: Evidence of referential processing deficits in amnesia. J Exp Psychol Gen 142:1346.

Lee JK, Ekstrom AD, Ghetti S (2014): Volume of hippocampal subfields and episodic memory in childhood and adolescence. NeuroImage 94:162–171.

Lord C, Rutter M, Le Couteur A (1994): Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24:659–685.

Lord C, Risi S, Lambrecht L, Cook EH, Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M (2000): The autism diagnostic

observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord 30:205–223.

Mårtensson J, Eriksson J, Bodammer NC, Lindgren M, Johansson M, Nyberg L, Lövdén M (2012): Growth of language-related brain areas after foreign language learning. Neuroimage 63:240–244.

McLachlan G, Do KA, Ambroise C (2005): Analyzing Microarray Gene Expression Data, Vol. 422. New York: Wiley.

Mullen EM (1995): Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service.

Nordahl CW, Simon TJ, Zierhut C, Solomon M, Rogers SJ, Amaral DG (2008): Brief report: methods for acquiring structural MRI data in very young children with autism without the use of sedation. J Autism Dev Disord 38:1581–1590.

Nosarti C, Al-Asady MH, Frangou S, Stewart AL, Rifkin L Murray RM (2002): Adolescents who were born very preterm have decreased brain volumes. Brain 125:1616–1623.

Nugent IIITF, Herman DH, Ordonez A, Greenstein D, Hayashi KM, Lenane M, Gogtay N (2007): Dynamic mapping of hippocampal development in childhood onset schizophrenia. Schizophr Res 90:62–70.

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013): Long-axis specialization of the human hippocampus. Trends Cogn Sci 17:230–240.

Schumann CM, Hamstra J, Goodlin-Jones BL, Lotspeich LJ, Kwon H, Buonocore MH, Amaral DG (2004): The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. J Neurosci 24:6392–6401.

Shen L, Firpi HA, Saykin AJ, West JD (2009): Parametric surface modeling and registration for comparison of manual and automated segmentation of the hippocampus. Hippocampus 19: 588–595.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Matthews PM (2004): Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23:208–219.

Strange BA, Witter MP, Lein ES, Moser EI (2014): Functional organization of the hippocampal longitudinal axis. Nature Rev Neurosci 15:655–669.

Tae WS, Kim SS, Lee KU, Nam EC, 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.

Uematsu A, Matsui M, Tanaka C, Takahashi T, Noguchi K, Suzuki M, Nishijo H (2012): Developmental trajectories of amygdala and hippocampus from infancy to early adulthood in healthy individuals. PLoS One 7:e46970.

Wang H, Yushkevich PA (2012): Guiding automatic segmentation with multiple manual segmentations. In: Ayache N, Delingette H, Golland P, Mori K, editor. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012. Berlin: Springer. p 429–436.

Wang H, Das SR, Suh JW, Altinay M, Pluta J, Craige C, Yushkevich PA (2011): A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. NeuroImage 55:968–985.

Warshaw MG, Dyck I, Allsworth J, Stout RL, Keller MB (2001): Maintaining reliability in a long-term psychiatric study: An ongoing inter-rater reliability monitoring program using the longitudinal interval follow-up evaluation. J Psychiatr Res 35:297–305.

Weiss AP, DeWitt I, Goff D, Ditman T, Heckers S (2005): Anterior and posterior hippocampal volumes in schizophrenia. Schizophr Res 73:103–112.

Wenger E, Mårtensson J, Noack H, Bodammer NC, Kühn S, Schaefer S, Lövdén M (2014): Comparing manual and automatic segmentation of hippocampal volumes: Reliability and validity issues in younger and older brains. Hum Brain Map 35:4236–4248.

Wierenga L, Langen M, Ambrosino S, van Dijk S, Oranje B, Durston S (2014): Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. NeuroImage 96:67–72.