Data in Brief 000 (2020) 106043

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Towards Optimising MRI Characterisation of Tissue (TOMCAT) Dataset including all Longitudinal Automatic Segmentation of Hippocampal Subfields (LASHiS) data

Thomas B Shaw[a,*], Ashley York[b], Markus Barth[a,c,d,†], Steffen Bollmann[a,d,†,*]

[a] Centre for Advanced Imaging, The University of Queensland, Brisbane, Australia
[b] School of Psychology, The University of Queensland, Brisbane, Australia
[c] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia
[d] ARC Training Centre for Innovation in Biomedical Imaging Technology, The University of Queensland, Brisbane, QLD, Australia

ARTICLE INFO

Article history:
Received 16 April 2020
Revised 30 May 2020
Accepted 14 July 2020
Available online 20 July 2020

Keywords:
Hippocampus
Longitudinal studies
Segmentation
Magnetic resonance imaging
Image processing
Computer-assisted

ABSTRACT

Seven healthy participants were scanned using a Siemens Magnetom 7 Tesla (T) whole-body research MRI scanner (Siemens Healthcare, Erlangen, Germany). The first scan session was acquired in 2016 (time point one), the second and third session in 2019 (time point two and three, respectively) with the third session acquired 45 min following the second as a scan-rescan condition. The following scans were acquired for all time points: structural T1 weighted (T1w) MP2RAGE, high in-plane resolution Turbo-Spin Echo (TSE) dedicated for hippocampus subfield segmentation. The data were used in three projects to date, for more insight see: 1) Non-linear realignment for Turbo-Spin Echo retrospective motion correction and hippocampus segmentation improvement [1] 2) Longitudinal Automatic Segmentation of Hippocampal Subfields (LASHiS) using multi-contrast MRI [2]. 3) The challenge of bias-free coil combination for quantitative susceptibility mapping at ultra-high field [3]. Data were

DOI of original article: 10.1016/j.neuroimage.2020.116798
* Corresponding authors.
E-mail addresses: t.shaw@uq.edu.au (T.B. Shaw), steffen.bollmann@cai.uq.edu.au (S. Bollmann).
† Shared Senior Author

https://doi.org/10.1016/j.dib.2020.106043

converted from DICOM to nifti format following the Brain Imaging Data Structure (BIDS) [4]. Data were analysed for the accompanying manuscript "Longitudinal Automatic Segmentation of Hippocampal Subfields (LASHiS) using multi-contrast MRI" including test-retest reliability and longitudinal Bayesian Linear Mixed Effects (LME) modelling.

Specifications Table

Subject	Neuroscience
Specific subject area	Neuroimaging and Magnetic Resonance Imaging Data
Type of data	MRI data (Nifti organized in BIDS standard) and associated data describing relationship between MRI data.
How data were acquired	Magnetic Resonance Imaging using a Siemens 7 T Magnetom whole-body research scanner (Siemens Healthcare, Erlangen, Germany), with maximum gradient strength of 70 mT/m and a slew rate of 200 mT/m/s and a 7 T Tx/32 channel Rx head array (Nova Medical, Wilmington, MA, USA).
Data format	Raw Nifti (.nii.gz) format. Analysed data is in R code and related plots and raw statistical data are in PDFs and CSV files.
Parameters for data collection	2D TSE sequence (Siemens WIP tse_UHF_WIP729C, variant: tse2d1_9), TR: 10300 ms, TE: 102 ms, FA: 132°, FoV: 220 mm, voxel size of $0.4 \times 0.4 \times 0.8$ mm^3 Turbo factor of 9; iPAT (GRAPPA) factor 2, acquisition time (TA) 4 min 12 s, number of slices = 72, repeated thrice over a slab aligned orthogonally to the hippocampus formation. Whole-brain T1w prototype MP2RAGE sequence (WIP 900; [5, 6]) at 0.75 mm isotropic voxel size (TR/TE/TIs = 4300 ms / 2.5 ms / 840 ms, 2370 ms, TA = 6:54, number of slices = 256)
Description of data collection	Participants were scanned in the Centre For Advanced Imaging's 7 T MRI scanner at The University of Queensland http://hdl.handle.net/102.100.100/62799. The research was approved by the university human ethics committee and written informed consent was obtained from the participants.
Data source location	Institution: The University of Queensland City/Town/Region: Brisbane Country: Australia Latitude and longitude (and GPS coordinates) for collected samples/data: -27.49974175182734,153.01199407746012] Scanner unique identifier: http://hdl.handle.net/102.100.100/62799
Data accessibility	Repository name: Open Science Foundation (OSF) Data identification number: bt4ez Direct URL to data: https://osf.io/bt4ez
Related research article	1) Shaw, T. B., York, A., Ziaei, M., Barth, M., Bollmann, S. (2020). Longitudinal Automatic Segmentation of Hippocampal Subfields (LASHiS) using multi-contrast MRI. *NeuroImage*, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2020.116798

Value of the Data

These data are unique as they contain a longitudinal, high-resolution MR image dataset acquired at ultra-high field (7 Tesla) focussing on the hippocampus with a test-retest condition. This is especially important as a test set for longitudinal hippocampus segmentation methods.

This dataset benefits researchers interested in hippocampus segmentation methods and interested in validating their own methods, as well as researchers interested in high-quality data for developing image processing methods.

These data can be used for further refinement/development of experiments in the direction of hippocampal segmentation and shape analysis/morphology generally as the fine details of the hippocampus can be explored and analysed with this dataset. There also exists the possibility for incorporating this data into machine learning algorithms and atlas creation for multi-atlas fusion methods.The additional value of this data is its longitudinal and test-retest acquisition, as well as being acquired at 7 T. These data are difficult and expensive to acquire and therefore would be of high benefit to the scientific community.

1. Data Description

The data are split into two sections: The Open Science Framework (OSF) repository, which contains the raw MRI nifti files in BIDS format. The Github repository associated with this work contains the figures, tables, raw data, and processing scripts associated with this work. Please see the sections below for explanations of the data naming.

Within the OSF repository:

- /subject_identifier
  - /session_identifier
    - anat
      - <subject_identifier>_<session-identifier>_IV1_defaced.nii.gz
        The first inversion of the MP2RAGE sequence
      - <subject_identifier>_<session-identifier>_IV2_defaced.nii.gz
        The second inversion of the MP2RAGE sequence
      - <subject_identifier>_<session-identifier>_T1Map_defaced.nii.gz
        T1 Maps computed through the MP2RAGE sequence
      - <subject_identifier>_<session-identifier>_T1w_defaced.nii.gz
        Uniform Denoised (UNIDEN) T1w image from the MP2RAGE sequence (used in all processing)
      - <subject_identifier>_<session-identifier>_UNI_defaced.nii.gz
        Uniform image from the MP2RAGE sequence
      - <subject_identifier>_<session-identifier>_T2w_run-1_tse.nii.gz
        First run of the TSE sequence
      - <subject_identifier>_<session-identifier>_T2w_run-2_tse.nii.gz
        Second run of the TSE sequence
      - <subject_identifier>_<session-identifier>_T2w_run-3_tse.nii.gz
        Third run of the TSE sequence

Within the LASHiS Github directory:

- /Experiment_files_for_LASHiS

  /Data
  - Raw data and .csv files for LASHiS experiment one and two
  - The stan model files (.txt and .stan) are the models fed into the stan program defined in Scripts/Analysis/stan_plotResultsADNI.R and Scripts/Analysis/stan_plotResultsTOMCAT.R.
  - The subfields.csv file contains the names of the subfields that are read in by Scripts/Analysis/stan_plotResultsTOMCAT.R and Scripts/Analysis/stan_plotResultsADNI.R
  - /ADNI
    - Unreconciled data are the original data that were created by LASHiS, Freesurfer, and ASHS. These data were concatenated by the common subfields as described in the LASHiS paper. This directory also includes demographic details of the participants and both image and subject IDs.

(*continued on next page*)

- /Experiment_files_for_LASHiS
  - "Fully" reconciled data are augmented by the script in Scripts/Analysis/reconcileDataADNI.R and are read into the STAN model.
  - /TOMCAT
    - Unreconciled data are the original data that were created by LASHiS, Freesurfer, and ASHS. These data were concatenated by the common subfields as described in the LASHiS paper. This directory also includes demographic details of the participants and both image and subject IDs.
  - /test_retest_data
    - LASHiS_experiment_one_DICE.omv contains the wide format Jamovi data per participant for each of the subfields and each of the methods (Freesurfer, ASHS, LASHiS, etc.) for DICE scores in the test-retest experiment. These can be read and analysed using the open source program Jamovi.
    - LASHiS_experiment_one_volumeSim.omv - as above though for volume similarity.
- /Figures
  - Figures from the LASHiS manuscript created from the analysis for the LASHiS experiments including supplementary materials
- /Scripts
  - /1_preprocessing
    - Preprocessing scripts for LASHiS data (TOMCAT and ADNI) that ingest the raw MP2RAGE and TSE data and pre-process the data including denoising, N4 bias correction, interpolation, non-linear realignment
  - /2_ADNI_experiment
    - Scripts for running LASHiS and Freesurfer methods for hippocampus subfield segmentation for the LASHiS manuscript
  - /Analysis
    - R code for the Bayesian LME experiment including code to concatenate raw results, analyse the and plotting of all results including the STAN models.
    - Hippo_scans_10_27_2019_reconciled.csv contains a list of scans and demographics from ADNI
    - reconcileDataADNI.R is code that ensures that all data are present for all sessions for all pipelines (Freesurfer, LASHiS, etc.) for each participant.
- the stanPlotResults files are the main analysis scripts for the LME for each of the datasets (TOMCAT and ADNI)
  - /LASHiS_experiment_one
    R code for the test-retest experiment for the LASHiS manuscript.
- pairesT_Bayes.R is the script for creating the figures from the Bayes T-tests
  The script for conducting the Bayesian *t*-tests is in /Data//test_retest_data/LASHiS experiment_one_DICE.omv as the Jamovi file contains both Data and Analysis. Matrix_file_code_02092019.R is the code to generate the supplementary figures for the significance matrices
  - /bidscoin
    - All associated code for converting ADNI data to the Brain Imaging Data Standard using bidscoin
  - /Singularity
    - Code for creating the Docker/Singularity image that contains the software for the entire reproducible analysis pipeline
    - The docker/singularity image is hosted at https://hub.docker.com/r/caid/adni_lashis_simg

2. Experimental Design, Materials, and Methods

Seven healthy participants (age: M = 26.29, SD = 3.35, sex: 3 female, 4 male) were scanned using a 7 T whole-body research scanner (Siemens Healthcare, Erlangen, Germany), with maximum gradient strength of 70 mT/m and a slew rate of 200 mT/m/s and a 7 T Tx/32 channel Rx head array (Nova Medical, Wilmington, MA, USA) in three sessions with three years between session one and two, and 45 minutes between two and three for a scan-rescan condition.

The study was approved by the university human ethics committee and written informed consent was obtained from the participants. Participants agreed to share their de-identified data. Participants were instructed to remain as still as possible for the duration of the scan.

For the first time point, the participants were recruited as part of a larger study [3], though only the relevant sequences are detailed here. For the second and third time points, the participants were scanned at 7 T as part of a larger study attempting to optimize MRI contrast for

characterizing tissue at Ultra-High Field (UHF). The second and third session were separated by 45 min, during which the participant was continuously scanned with other MR sequences.

For the TOMCAT dataset described here, participants were scanned using a 2D TSE sequence (Siemens WIP tse_UHF_WIP729C, variant: tse2d1_9), TR: 10300 ms, TE: 102 ms, FA: 132°, FoV: 220 mm, voxel size of 0.4 × 0.4 × 0.8 mm^3 Turbo factor of 9; iPAT (GRAPPA) factor 2, acquisition time (TA) 4 min 12 s. The scan was repeated thrice over a slab aligned orthogonally to the hippocampus formation. An anatomical whole-brain T1w scan using a prototype MP2RAGE sequence (WIP 900; [5,6] at 0.75 mm isotropic voxel size was also acquired (TR/TE/TIs = 4300 ms / 2.5 ms / 840 ms, 2370 ms, TA = 6:54). The TSE sequence was planned on the MP2RAGE by an experienced MR operator.

Data were converted to the BIDS [4] format using BIDSCoin (https://github.com/Donders-Institute/bidscoin) and all data were processed using containerized software utilizing Singularity https://sylabs.io/singularity/ and open-source software including ANTs [7,8], FSL [9], and R [10].

At the first time point, the nominal resolution of the MP2RAGE was 0.5 mm isotropic with the same parameters. This resulted in a smaller FOV for these scans, with none of the seven participants having brain tissue excluded from the scan. For all subsequent processing, all MP2RAGE images for the first time point were resampled to 0.75 mm isotropic using b-spline interpolation.

3. Pre-processing and cross-sectional processing

For pre-processing of all data, we included modified pre-processing steps based on the ANTs cortical thickness pipeline [11] and our previous work. The scripts for all preprocessing are located in the accompanying Github directory under /LASHiS/Experiment_files_for_LASHiS/Scripts/**1_preprocessing**. These scripts contain calls to the following programs, and were incorporated to ensure consistent segmentation results across participants. Details of the pre-processing can be found in [2]

4. Longitudinal Assessment of Hippocampal Subfields (LASHiS)

The Longitudinal Automatic Segmentation of Hippocampus Subfields (LASHiS) pipeline consists of the following steps with the input of any number of T1w and T2w individual time points per participant:

a pre-processing of both T1w and T2w scans described above.

1) Offline construction of a sample-specific atlas for LASHiS. Optionally, the ASHS pipeline can be optimised through the incorporation of a group-specific atlas. Similarly, creation of a group-specific atlas is a benefit to our proposed method. This atlas is comprised of a representative pool of subjects (approximately 20-30 participants), manually labelled, and passed through the ASHS_train pipeline described in Yushkevich et al. [12]. Essentially, the manual segmentations are used as inputs (priors) for the joint-label fusion algorithm in subsequent segmentations, and to train classifiers for the ASHS cross-sectional pipeline. Creating a group-specific atlas (of 20-30 subjects) would be beneficial for large longitudinal studies, as segmentation training would be performed on group-specific characteristics. However, having a group specific atlas is generally not necessary for robust performance of ASHS [13].
2) Labelling of individual time points of each subject using ASHS and a representative atlas (or an atlas created in Step 1) to yield hippocampus subfield estimates.
3) Construction of a multi-contrast Single Subject Template using antsMultivariateTemplateConstruction.sh.

4) Joint-label fusion of each of the individual time point labels to the Single Subject Template using both contrasts and individual hippocampus subfield labels to produce a labelled Single Subject Template using an implementation of the Joint Label Fusion algorithm from Wang et al. [14].
5) Application of the inverse subject-to-Single-Subject-Template transformations to Single Subject Template labels using C3D [15].
6) Measurement and calculation of subfield labels in subject-space using C3D.

The entire LASHiS script is an automatic pipeline in a single shell script located in github.com/caisr/lashis named LASHiS.sh. This script can be executed on Linux systems by cloning the git directory and executing the script, or by using the corresponding Docker/Singularity images located in dockerhub/caid using the commands located in the README of the LASHiS GitHub repository. This can be executed in a variety of environments, for example in Windows using the windows subsystem for Linux 2.0, docker for Windows, or on Mac using Docker or Vagrant https://singularity.lbl.gov/install-mac#setup. The run command for LASHiS using Docker/Singularity is located within the readme of the LASHiS Github page.

5. Statistical Evaluation and the LASHiS pipeline

The analyzed data in the Github respository associated with this dataset includes the results and raw CSV files for each of the participants' raw scores for the following five methods for hippocampus subfields segmentation. Examples of the output of each segmentation strategy are given in Fig. 2 of the Github directory. Details of the statistical evaluation can be found in [2]

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

The authors acknowledge the facilities and scientific and technical assistance of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at the Centre for Advanced Imaging, The University of Queensland. MB acknowledges funding from Australian Research Council (ARC) Future Fellowship grant FT140100865, ARC Training Centre grant IC170100035, ARC Discovery Project grant DP200103386, NHMRC-NIH BRAIN Initiative Collaborative Research Grant APP1117020, NIH grant 1R01MH111419. This research was undertaken with the assistance of resources and services from the Queensland Cyber Infrastructure Foundation (QCIF). The authors would like to acknowledge Siemens Healthineers for the support of our project through the provision of the WIP sequence used to acquire the tse_UHF_WIP729C (variant: tse2d1_9) and MP2RAGE sequence (WIP 900) data in this publication.

References

[1] T.B. Shaw, et al., Non-linear realignment improves hippocampus subfield segmentation reliability, Neuroimage 203 (Dec. 2019) 116206, doi:10.1016/j.neuroimage.2019.116206.
[2] T. Shaw, A. York, M. Ziaei, M. Barth, S. Bollmann, Longitudinal Automatic Segmentation of Hippocampal Subfields (LASHiS) using Multi-Contrast MRI, Neuroimage (Apr. 2020) 116798, doi:10.1016/j.neuroimage.2020.116798.
[3] S. Bollmann, et al., The challenge of bias-free coil combination for quantitative susceptibility mapping at ultra-high field, Magn. Reson. Med. 79 (1) (2018) 97–107, doi:10.1002/mrm.26644.
[4] K.J. Gorgolewski, et al., The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments, Sci. Data 3 (Jun. 2016) 160044, doi:10.1038/sdata.2016.44.

[5] J.P. Marques, T. Kober, G. Krueger, W. van der Zwaag, P.-F. Van de Moortele, R. Gruetter, MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field, Neuroimage 49 (2) (Jan. 2010) 1271–1281, doi:10.1016/j.neuroimage.2009.10.002.
[6] J.T. O'Brien, A. Lloyd, I. McKeith, A. Gholkar, N. Ferrier, A longitudinal study of hippocampal volume, cortisol levels, and cognition in older depressed subjects, Am. J. Psychiatry 161 (11) (2004) 2081–2090, doi:10.1176/appi.ajp.161.11.2081.
[7] N.J. Tustison, et al., N4ITK: Improved N3 Bias Correction, IEEE Trans. Med. Imaging 29 (6) (Jun. 2010) 1310–1320, doi:10.1109/TMI.2010.2046908.
[8] S.G. Avants, B Tustison, N., Adv. Norm. Tools, ANTS 1.9 (2013).
[9] UK. Analysis Group FMRIB, Oxford, *FSL - FMRIB Software Library v6.0*. 2019.
[10] R Core Team, R: A Lang. Environ. Stat. Comput. (2019).
[11] N.J. Tustison, et al., Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements, Neuroimage 99 (Oct. 2014) 166–179, doi:10.1016/j.neuroimage.2014.05.044.
[12] P.A. Yushkevich, et al., Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment, Hum. Brain. Mapp. 36 (1) (Jan. 2015) 258–287, doi:10.1002/hbm.22627.
[13] L. Xie, et al., Automated Multi-Atlas Segmentation of Hippocampal and Extrahippocampal Subregions in Alzheimer's Disease at 3T and 7T: What Atlas Composition Works Best? J. Alzheimers Dis. 63 (1) (2018) 217–225, doi:10.3233/JAD-170932.
[14] H. Wang, J.W. Suh, S.R. Das, J.B. Pluta, C. Craige, P.A. Yushkevich, Multi-Atlas Segmentation with Joint Label Fusion, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (Mar. 2013) 611–623, doi:10.1109/TPAMI.2012.143.
[15] P.A. Yushkevich, et al., User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability, Neuroimage 31 (3) (Jul. 2006) 1116–1128, doi:10.1016/j.neuroimage.2006.01.015.