

Automated Registration-Based Temporal Bone Computed Tomography Segmentation for Applications in Neurotologic Surgery

Andy S. Ding, MSE^{1,2} , Alexander Lu^{1,2}, Zhaoshuo Li, MSE³, Deepa Galaiya, MD¹, Jeffrey H. Siewerdse, PhD^{2,3}, Russell H. Taylor, PhD³ and Francis X. Creighton, MD¹

Otolaryngology—
Head and Neck Surgery
1–8
© American Academy of
Otolaryngology—Head and Neck
Surgery Foundation 2021
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/01945998211044982
<http://otojournal.org>


Abstract

Objective. This study investigates the accuracy of an automated method to rapidly segment relevant temporal bone anatomy from cone beam computed tomography (CT) images. Implementation of this segmentation pipeline has potential to improve surgical safety and decrease operative time by augmenting preoperative planning and interfacing with image-guided robotic surgical systems.

Study Design. Descriptive study of predicted segmentations.

Setting. Academic institution.

Methods. We have developed a computational pipeline based on the symmetric normalization registration method that predicts segmentations of anatomic structures in temporal bone CT scans using a labeled atlas. To evaluate accuracy, we created a data set by manually labeling relevant anatomic structures (eg, ossicles, labyrinth, facial nerve, external auditory canal, dura) for 16 deidentified high-resolution cone beam temporal bone CT images. Automated segmentations from this pipeline were compared against ground-truth manual segmentations by using modified Hausdorff distances and Dice scores. Runtimes were documented to determine the computational requirements of this method.

Results. Modified Hausdorff distances and Dice scores between predicted and ground-truth labels were as follows: malleus (0.100 ± 0.054 mm; Dice, 0.827 ± 0.068), incus (0.100 ± 0.033 mm; Dice, 0.837 ± 0.068), stapes (0.157 ± 0.048 mm; Dice, 0.358 ± 0.100), labyrinth (0.169 ± 0.100 mm; Dice, 0.838 ± 0.060), and facial nerve (0.522 ± 0.278 mm; Dice, 0.567 ± 0.130). A quad-core 16GB RAM workstation completed this segmentation pipeline in 10 minutes.

Conclusions. We demonstrated submillimeter accuracy for automated segmentation of temporal bone anatomy when compared against hand-segmented ground truth using our template registration pipeline. This method is not dependent on the training data volume that plagues many complex deep learning models. Favorable runtime and low computational requirements underscore this method's translational potential.

Keywords

temporal bone, automated segmentation, atlas, data set curation

Received July 19, 2021; accepted August 21, 2021.

Operating in the temporal bone is technically challenging due to the complex geometry of nerves, arteries, veins, and organs for hearing and balance within this region. Accessing the temporal bone requires drilling through varying densities of bone to identify surgical landmarks. Due to the limited visibility of the surgical field and narrow surgical corridors, temporal bone surgery poses a potential for accidental damage to surrounding anatomy. For example, semicircular canal injury from cholesteatoma removal can lead to severe sensorineural hearing loss,¹ while accidental contact with middle ear ossicles can cause conductive hearing loss from ossicular chain dislocation.² Even in patients with normal anatomy, an estimated 1.6% of patients reported permanent changes in taste after cochlear implantation at the end of their follow-up period.^{3–5} In more rare cases, patients are at risk for facial paralysis from accidental damage to the facial nerve or cerebrospinal fluid leakage from penetration of the surrounding dura.^{6,7}

¹Department of Otolaryngology—Head and Neck Surgery, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

²Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA

³Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, USA

This article was given as an oral presentation at the AAO-HNSF Annual Meeting & OTO Experience; October 4, 2021; Los Angeles, California.

Corresponding Author:

Andy S. Ding, Department of Otolaryngology—Head and Neck Surgery, School of Medicine, Johns Hopkins University, 733 N Broadway, Baltimore, MD 21205, USA.
Email: ading5@jhmi.edu

One possible approach for mitigating accidental damage to surrounding structures is the use of intraoperative image-guided robotic systems that can determine the location of robotically controlled instruments relative to patient imaging and enforce safety barriers around contacting critical anatomy.^{8–10} Such systems have been used extensively in other surgical specialties, such as orthopedics and neurosurgery,^{11,12} but given the complex bony anatomy and high degree of precision needed in neurotologic surgery, image-guided robotics has seen little implementation in our field. A key obstacle to utilizing the full potential of these technologies is the lack of accurate and efficient methods for labeling critical anatomy on patient computed tomography (CT) imaging. While manually segmenting surgically relevant landmarks on preoperative imaging can be performed, this is extremely time intensive and prone to interreader variability.¹³ To overcome these limitations, our group has focused on developing methods to automate this process. This study presents an efficient, accurate, and automated pipeline for segmenting structures in temporal bone CT scans that is not dependent on the use of a large number of training images.

Methods

This study was approved by the Johns Hopkins Medicine Institutional Review Board (IRB00279939). To build an automated segmentation pipeline for relevant anatomy in the temporal bone, we first curated a database of high-resolution temporal bone CT scans. These scans were manually segmented to serve as a validation set for this segmentation pipeline and were used to build a temporal bone template, whose segmentations were propagated to predict anatomic labels in target images.

Creation of Manual Temporal Bone Segmentation Data Sets

Deidentified and cropped cone beam temporal bone CT scans were obtained from the Johns Hopkins Department of Otolaryngology—Head and Neck Surgery. The resolution of scans used in this study was 0.1 mm per voxel length, with image dimensions of $512 \times 512 \times N$ voxels, where N refers to the number of axial CT slices. Scans with anatomy-altering pathology (eg, cholesteatoma, congenital deformities, trauma) or surgical history in the temporal bone were excluded from this study, leaving 16 data sets (8 right, 8 left) included for manual segmentation. A total of 16 anatomic structures (eg, ossicles, bony labyrinth, facial nerve, and chorda tympani) were labeled for each data set by using the open source software 3D Slicer (**Table I**). Scans were manually labeled by 2 medical trainees with experience in temporal bone anatomy and were verified by the senior author.

Multiple strategies were implemented to minimize variability in manual segmentations. The facial nerve was segmented from the labyrinthine segment to its exit from the stylomastoid foramen (**Figure 1A**). Since the chorda tympani was unable to be identified in the tympanic cavity, the nerve was segmented from its facial nerve branch point until it enters the middle ear. The petrotympanic fissure portion of

Table I. Relevant Anatomic Structures Hand Segmented in Each Temporal Bone.

ID	Structure
1	Bone
2	Malleus
3	Incus
4	Stapes
5	Bony labyrinth
6	Internal auditory canal
7	Superior vestibular nerve
8	Inferior vestibular nerve
9	Cochlear nerve
10	Facial nerve
11	Chorda tympani
12	Internal carotid artery
13	Sigmoid sinus + dura
14	Vestibular aqueduct
15	Mandible
16	External auditory canal

the chorda tympani was also segmented until its entrance into the infratemporal fossa (**Figure 1B**). The internal auditory canal was truncated medially by using the surrounding bone as reference. An oblique plane along the superior petrosal sinus groove delineated the medial extent of the internal auditory canal such that the segment lay flush with the surrounding petrous bone (**Figure 1C**). Due to similar CT density between the vertical segment of the internal carotid artery and surrounding soft tissue, only the petrous portion of the internal carotid artery was included (**Figure 1D**). Finally, the medial extent of the external auditory canal was segmented close to the tympanic membrane, while the lateral extent of the external auditory canal was delineated with a parasagittal plane at the spine of Henle (**Figure 1E**).

A joint-smoothing procedure, which preserves tight boundaries between segments, was performed in 3D Slicer with a factor of 0.30 units on all segments, except the chorda tympani, stapes, vestibular aqueduct, and inferior vestibular nerve. These latter 4 segments were joint smoothed with a factor of 0.15 units.

Average Temporal Bone Template Creation

A standard template-building technique was performed as previously described¹⁴ to generate an unbiased “average temporal bone,” hereafter called the average template image, whose corresponding labeled anatomy was used to predict the location of anatomy in other scans. One scan was designated as the template image, and the remaining scans were deformably registered to the template (**Figure 2A**). Scans on the contralateral side to the template image were flipped along the midsagittal plane. Deformable registration was accomplished by using the symmetric normalization registration method with Advanced Normalization Tools (ANTs) software.¹⁵ This method rigidly aligns a target image to the

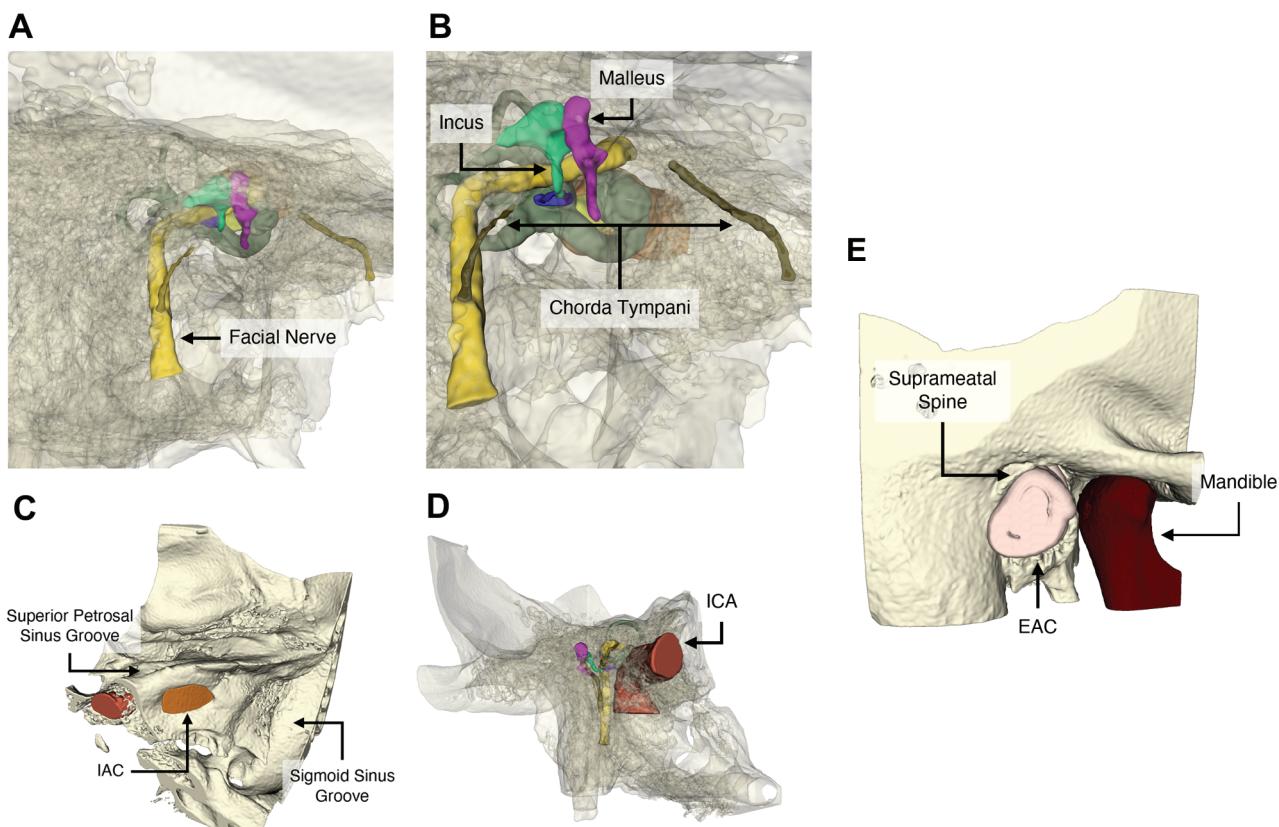


Figure 1. Three-dimensional rendering of sample segmentations. (A) Facial nerve. (B) Mastoid and petrotympanic fissure portions of the chorda tympani. (C) Internal auditory canal (IAC). (D) Internal carotid artery (ICA). (E) External auditory canal (EAC).

template image, iteratively morphs the target image to match the template image, and outputs a forward and an inverse deformation field. These deformation fields consist of 3-dimensional vectors that move each voxel of one image to the position of a corresponding voxel in the other image. The forward deformation fields in this case describe a voxel-by-voxel deformation of each target image to match the designated template image. The inverse deformation fields, however, contain information to morph the template image into each target image and can therefore be used to generate an average temporal bone from the template image. By taking a voxel-by-voxel average of the vectors within these inverse deformation fields, we calculated the average inverse deformation field from this cohort and applied it to the template image to generate the average template image. By applying the average inverse deformation field to the template image's segmentations, we generated high-quality labels corresponding to the average template image (**Figure 2, B and C**).

Automatic Segmentation Propagation Pipeline

To segment a new temporal bone scan, we deformably registered the average template image to the target scan using the symmetric normalization registration method with ANTs. Since the average template was registered to the target scan, the forward deformation field output by ANTs therefore contains voxel-by-voxel information to morph the average template to the target scan. The forward deformation field was

then applied to the average template's segmentations to map these segmentations onto the target image and produce predicted labels of target image anatomy.

Predicted Segmentation Accuracy Metrics

Predicted segmentations were evaluated against hand-segmented ground-truth labels via a leave-one-out cross-validation method. For each left-out data set, an average template was created with the remaining data sets. To minimize bias, 1 of the 16 data sets was designated as the template image for average template creation and was not included in accuracy analysis. We then applied our segmentation propagation pipeline on the segments of this average template to generate label predictions for the left-out data set. These predictions were then compared with ground-truth manual segmentations of the left-out data set. This process was repeated 15 times, once for each left-out data set, and the results were summarized statistically.

To evaluate the accuracy of predictions as compared with ground truth, Dice similarity scores were calculated,¹⁶ which measure the degree of overlap between volumes and range from 0 (no overlap) to 1 (perfect overlap). Modified Hausdorff distances (MHDs)¹⁷ were also calculated to capture the average error in millimeters between predicted and ground-truth segmentations. To this end, 3-dimensional point clouds were created for predicted and ground-truth segmentations. For each point from the predicted label, the corresponding

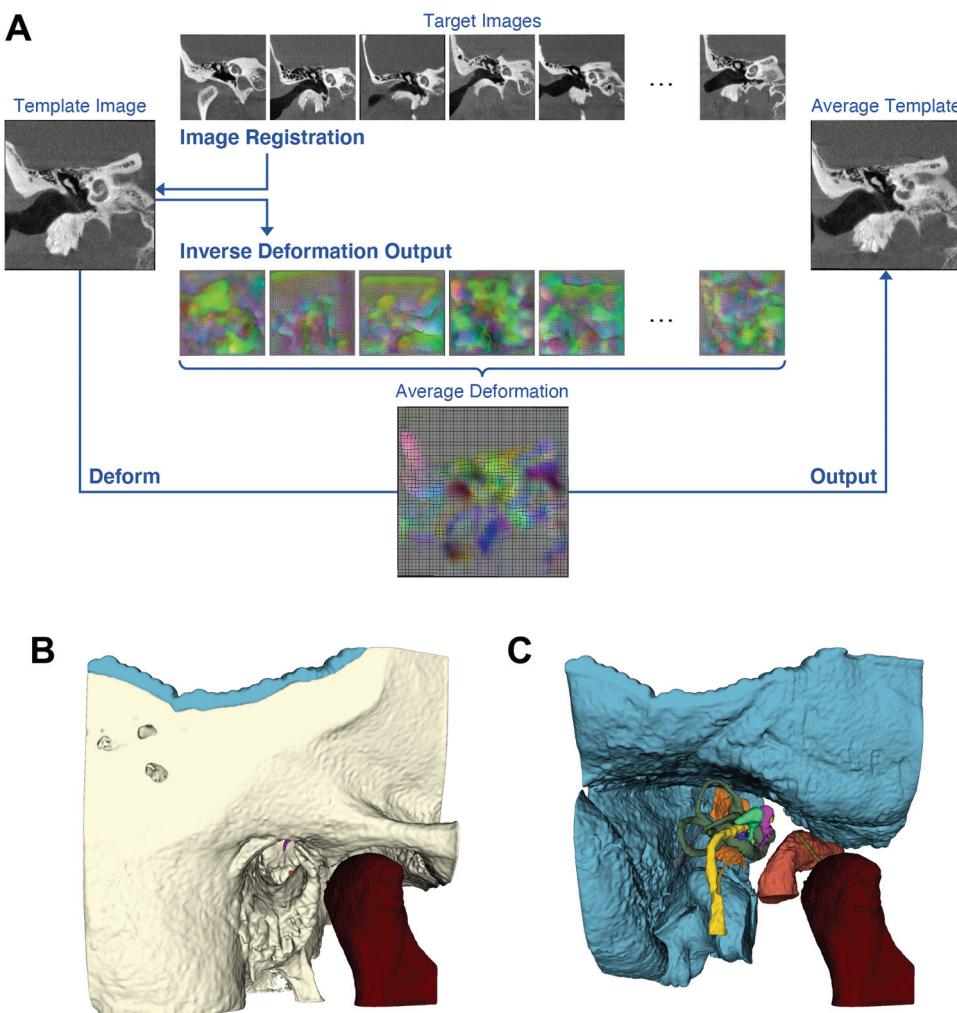


Figure 2. (A) Workflow for generating an average template image from multiple temporal bone computed tomography scans. (B, C) Three-dimensional rendering of the resultant average template with and without bone to reveal internal structures.

closest point from the ground-truth segmentation was determined with its associated distance. The average of these distances is defined as the MHD.

Constrained linear regressions were performed with Prism version 9.0 (GraphPad Software Inc) to examine the correlation between Dice scores and MHDs. Since an MHD of 0 mm corresponds to a Dice score of 1.0, linear regressions were constrained to a y-intercept of 1.0.

Pipeline Runtime Analysis

Critical steps within the segmentation propagation pipeline were time stamped and recorded for each target image segmentation. Difference between time stamps in seconds were calculated to determine the runtime of each pipeline step.

Results

Average MHDs and Dice scores for all segments are listed in **Table 2**. Predicted labels for all but 3 of the 16 structures segmented demonstrated MHDs <1 mm (**Figure 3A**). Automated segmentation of well-enclosed structures consistently achieved submillimeter MHDs: surrounding bone, 0.586 ± 0.219 mm; malleus, 0.100 ± 0.054 mm; incus, 0.100 ± 0.033

mm; stapes, 0.157 ± 0.048 mm; and bony labyrinth, 0.169 ± 0.100 mm. Predictions for the thinnest and smallest structures—namely, the stapes, superior and inferior vestibular nerves, chorda tympani, and vestibular aqueduct—exhibited lower Dice scores than other segments (**Figure 3B**). Apart from the stapes, labels for the middle ear ossicles resulted in high Dice scores: malleus, 0.827 ± 0.068 ; incus, 0.837 ± 0.045 ; and stapes, 0.358 ± 0.100 . Similarly, predicted bony labyrinth segmentations achieved an average Dice score of 0.838 ± 0.060 . MHDs for the facial nerve (0.522 ± 0.278 mm) were considerably smaller than for the chorda tympani (1.612 ± 1.010 mm), as reflected in their Dice scores (facial nerve, 0.567 ± 0.130 ; chorda tympani, 0.118 ± 0.120). Segmentation of smaller and/or thinner structures, such as the vestibular nerves and aqueduct, resulted in disproportionately lower Dice scores than larger and/or thicker structures (**Figure 4**), despite submillimeter MHDs. Constrained linear regression analyses showed that Dice scores for small or thin structures decreased significantly more ($P = .0017$) with respect to MHDs (slope, -0.7678 ; 95% CI, -1.046 to -0.4898) as compared with large or thick structures (slope, -0.3361 ; 95% CI, -0.4225 to -0.2498). Due to the oversensitivity of

Table 2. Modified Hausdorff Distances and Dice Scores Calculated Between Predicted Labels and Ground Truth for All Anatomic Structures Segmented.

Structure	Modified Hausdorff, mm		Dice score	
	Mean	SD	Mean	SD
Bone	0.586	0.219	0.712	0.037
Malleus	0.100	0.054	0.827	0.068
Incus	0.100	0.033	0.837	0.045
Stapes	0.157	0.048	0.358	0.100
Bony labyrinth	0.169	0.100	0.838	0.060
Internal auditory canal	1.106	0.623	0.681	0.129
Superior vestibular nerve	0.402	0.235	0.390	0.112
Inferior vestibular nerve	0.911	0.546	0.040	0.083
Cochlear nerve	0.490	0.254	0.524	0.161
Facial nerve	0.522	0.278	0.567	0.130
Chorda tympani	1.612	1.010	0.118	0.120
Internal carotid artery	0.605	0.250	0.768	0.092
Sigmoid sinus + dura	1.270	0.510	0.618	0.060
Vestibular aqueduct	0.949	0.533	0.220	0.130
Mandible	0.783	0.603	0.654	0.122
External auditory canal	0.695	0.276	0.821	0.057

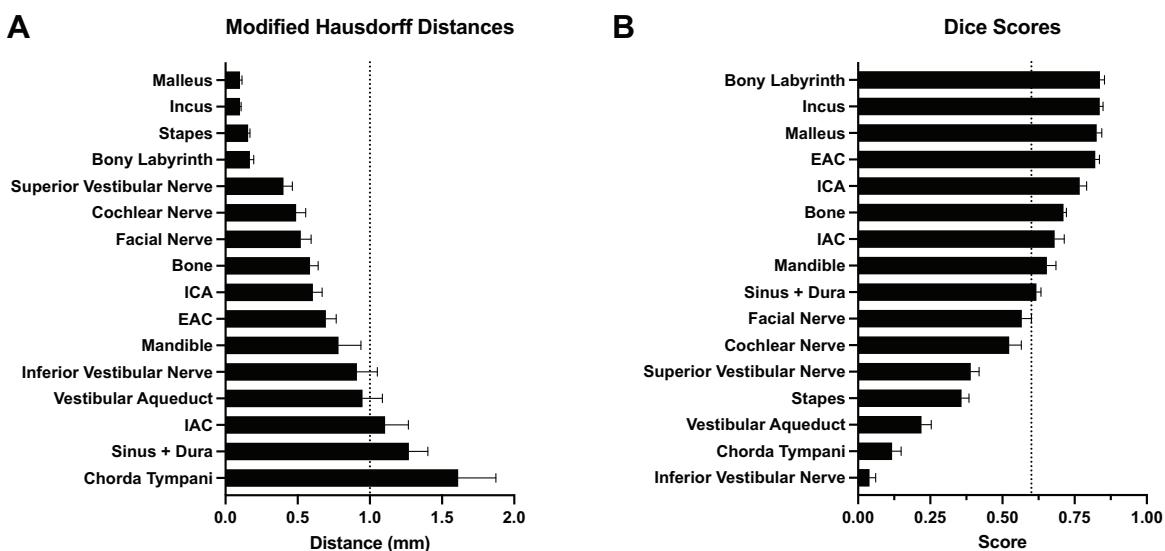


Figure 3. Accuracy metrics for propagated segments vs ground truth. (A) Modified Hausdorff distances between predictions and ground truth by increasing error. (B) Dice scores between predictions and ground truth by decreasing overlap. EAC, external auditory canal; IAC, internal auditory canal; ICA, internal carotid artery.

Dice scores to small translational perturbations for certain structures, MHDs provide a more objective accuracy metric in this study.

The image registration and segmentation propagation pipeline was performed on a quad-core machine with 16GB RAM and exhibited an average runtime of 601.4 seconds. Upon analyzing component steps within the pipeline, file reading took 21.7 seconds on average (3.60% total runtime); preprocessing, 23.5 seconds (3.91%); image registration, 451.0 seconds

(74.99%); label propagation, 62.7 seconds (10.42%); and file writing, 42.5 seconds (7.07%).

Discussion

Labeling of relevant anatomy in the temporal bone has the potential to significantly improve intraoperative safety in image-guided surgery,¹⁸ but high-quality manual segmentation of these structures is difficult and time-consuming. From our study experience, manual segmentation of all 16 relevant

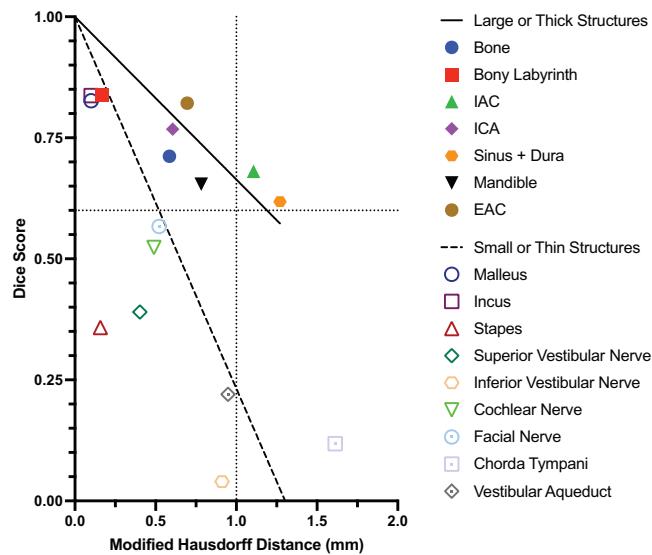


Figure 4. Accuracy comparison for smaller/thinner (unshaded) vs larger/thicker (shaded) structures. Linear regressions show Dice score sensitivity to translational errors for smaller/thinner structures. EAC, external auditory canal; IAC, internal auditory canal; ICA, internal carotid artery.

structures in a typical temporal bone data set took 6 to 10 hours in 3D Slicer, even with experienced labelers. While methods for automated segmentation within this space have been explored, these studies use deep learning or statistical shape model methods that require a significant number of manually segmented images for effective training and implementation.^{19–22} Furthermore, deep learning methods often require significant computing power and runtime for training, with some models requiring a dedicated graphics processing unit for adequate performance. In this study, we present an automated segmentation pipeline for the temporal bone that requires only 1 set of manually segmented labels. With favorable runtime on a commercially available personal computer, this pipeline requires minimal computing resources and user setup. Given that 75% of the runtime is attributed to the initial registration process, caching the resultant deformation fields to disk allows for a significantly reduced runtime of 144.4 seconds per target image. Moreover, other deformable registration implementations that use graphics processing units for computation are expected to further improve runtime performance.²³

An inherent disadvantage of statistical shape models and deep learning networks for automated segmentation is the difficulty in accommodating new anatomic labels once these models are generated.^{21,22,24,25} To do so, these models must be rebuilt after manually segmenting additional anatomy in all images used for training or statistical model generation. In contrast, segmentation propagation requires the user to manually segment the relevant structure on only the average template image. If deformation fields are saved after the initial registration process, the pipeline can update labels for rapid segmentation of new anatomy in all target images.

We have shown that segmentation propagation results in submillimeter accuracy for most of the structures in this

study. Generally, a Dice score of 0.7 represents adequate overlap, though this threshold was determined from anatomic structures in brain magnetic resonance imaging scans, which are larger and less geometrically complex than many of the structures labeled within our temporal bone CT data sets.¹⁶ Because of this difference in scale and complexity, we relaxed the Dice score cutoff to 0.6. Despite favorable results from MHD calculations, reported Dice coefficients for several structures, particularly nerves, were <0.6. We believe that this discrepancy is expected due to the known effect of small spatial changes on Dice scores between small and/or thin structures.¹⁶ The crura of the stapes, for example, have a thickness of 2 or 3 voxels (~0.2–0.3 mm). Therefore, a discrepancy of 0.2 mm between the predicted label and ground truth for the stapes potentially results in significantly small overlap between the structures. Our results corroborate this argument, as our pipeline reported an average MHD of 0.182 mm for the stapes but an average Dice score of 0.349. The vestibular nerves, chorda tympani, and vestibular aqueduct exhibited similar discrepancies between MHDs and Dice scores due to their thin geometry. Given the disadvantages of Dice scores for small and thin volumes, we believe that MHDs provide more reliable, robust, and outlier-resistant accuracy metrics for image guidance.^{17,26}

This segmentation pipeline generally performs with similar accuracy as other methods described in the literature.^{21,22,24,27} Segmentation of the malleus and incus with shape model-based or deep learning-based algorithms achieves Dice scores of 0.80 to 0.86, while our method achieves scores of 0.83 to 0.84. For the bony labyrinth, previous work has reported Dice scores of 0.82 to 0.91, while our method achieves an average score of 0.84. MHDs for these segments also lie <0.2 mm, which is consistent with distances currently reported in the literature.^{22,24} Based on these comparisons, the main advantages of this pipeline are its flexibility with incorporating new segmentations, ability to label without extensive training, and favorable runtime, while maintaining state-of-the-art segmentation accuracy.

There are several inherent limitations to this model. First, most anatomy evaluated in this study is not fully enclosed and was therefore subject to variable segmentation. While significant efforts were made to standardize segmentation of these structures, minor deviations in manual segmentations have the potential to decrease the validity of our segmentation accuracy analyses. Particularly for the dura, different segmentation techniques in 3D Slicer can result in different ground-truth labels between human labelers. Manually segmenting the dura requires labeling the floor of the temporal bone with a 3-dimensional brush. Discrepancies in brush size used for segmentation and the extent of the temporal bone floor segmented can cause notable differences in manual segmentations between images. Ultimately, variability in ground-truth segmentations for nonenclosed structures can lead to increased MHDs and decreased Dice scores when evaluating predicted labels. While we cannot fully eliminate this source of error, we believe that the strategies employed to standardize our segmentations, as described in the methods, minimized the effects of this limitation.

Second, cropping variation among temporal bone images from their CT scans can lead to inadequate segmentation of structures at image boundaries. After the initial registration step in this pipeline, some regions in the target image may be out of boundary in the template image and will therefore remain unlabeled. Because of this, anatomic structures at the boundary of the template image, such as the mandible and the internal carotid artery, will have artificially decreased accuracy metrics. Excluding these nonoverlapping areas from analysis may produce more representative accuracy metrics.

Finally, this segmentation pipeline struggles to segment anatomic variants consistently and accurately, since deformation fields can only regionally morph the template image but cannot change the underlying structural relationships between its segmentations. For example, the chorda tympani's branch point from the facial nerve varies widely in normal temporal bone scans. If the branch point in a target image lies on the medial aspect of the facial nerve but the branch point in the template image is on the lateral aspect, the propagated label will maintain the template image's lateral position of the branch point relative to the facial nerve. Similarly, while the superior and inferior vestibular nerves could be identified at the initial branching at the fundus, it was difficult to automate segmentation of these nerves, as they divided into the ampullary, singular, and other smaller branches due to large variation among temporal bones. The inconsistencies of nerve branching patterns with limitations of Dice score calculations on small and thin structures therefore contribute to decreased accuracy metrics for these structures.

Despite these limitations, this automated segmentation pipeline achieved submillimeter accuracy for most structures evaluated in this study, with efficient computational runtime, and required significantly lower computational power than existing automated segmentation methods.

Conclusion

This study presented an image registration-based pipeline that circumvents requirements of existing methods to segment anatomic structures in the temporal bone with submillimeter accuracy. Furthermore, we curated a database of high-quality comprehensive segmentations that include a set of critical temporal bone structures. Ultimately, automated segmentation methods for temporal bone CT scans such as our reported pipeline can improve existing image guidance technologies by informing surgeons of potential contact with critical anatomy in real time.

This study provides objective accuracy metrics of predictions from this pipeline with respect to ground-truth hand segmentations. Since each data set was labeled by 1 of 2 human labelers, we have yet to explore whether the reported accuracy of this pipeline is within the range of interlabeler error. Future work will recruit multiple human labelers and analyze their hand segmentations of data sets in this study against automated segmentations to investigate if this pipeline achieves similar labeling reliability to expert labelers.

Author Contributions

Andy S. Ding, conception, design, data acquisition, data analysis, manuscript drafting, data presentation, final approval; **Alexander Lu**, conception, design, data acquisition, manuscript drafting, final approval; **Zhaoshuo Li**, conception, design, manuscript drafting, final approval; **Deepa Galaiya**, conception, design, manuscript drafting, final approval; **Jeffrey H. Siewersden**, conception, design, manuscript drafting, final approval; **Russell H. Taylor**, conception, design, manuscript drafting, final approval; **Francis X. Creighton**, conception, design, data acquisition, manuscript drafting, final approval

Disclosures

Competing interests: Under a license agreement between Galen Robotics, Inc and the Johns Hopkins University, Russell H. Taylor and the university are entitled to royalty distributions on technology related to that described in the study discussed in this publication. Russell H. Taylor also is a paid consultant to and owns equity in Galen Robotics, Inc. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict-of-interest policies.

Sponsorships: None.

Funding source: Funding and equipment support were provided by a contract between Galen Robotics and the Johns Hopkins University.

ORCID iD

Andy S. Ding  <https://orcid.org/0000-0002-8490-057X>

References

- Brackmann D, Shelton C, Arriaga MA. *Otologic Surgery E-book*. Elsevier Health Sciences; 2015.
- Schick B, Dlugaczyk J. Surgery of the ear and the lateral skull base: pitfalls and complications. *GMS Curr Top Otorhinolaryngol Head Neck Surg*. 2013;12:Doc05. doi:10.3205/cto000097
- Linder TE, Lin F. Felsenbeinchirurgie. *HNO*. 2011;59(10):974-979. doi:10.1007/s00106-011-2359-z
- Michael P, Raut V. Chorda tympani injury: operative findings and postoperative symptoms. *Otolaryngol Head Neck Surg*. 2007;136(6):978-981. doi:10.1016/j.otohns.2006.12.022
- Jeppesen J, Faber CE. Surgical complications following cochlear implantation in adults based on a proposed reporting consensus. *Acta Otolaryngol*. 2013;133(10):1012-1021. doi:10.3109/00016489.2013.797604
- Fayad JN, Wanna GB, Micheletto JN, Parisier SC. Facial nerve paralysis following cochlear implant surgery. *Laryngoscope*. 2003; 113(8):1344-1346. doi:10.1097/00005537-200308000-00014
- Jeevan DS, Ormond DR, Kim AH, et al. Cerebrospinal fluid leaks and encephaloceles of temporal bone origin: nuances to diagnosis and management. *World Neurosurg*. 2015;83(4):560-566. doi:10.1016/j.wneu.2014.12.011
- Caversaccio M, Gavaghan K, Wimmer W, et al. Robotic cochlear implantation: surgical procedure and first clinical experience. *Acta Otolaryngol*. 2017;137(4):447-454. doi:10.1080/00016489.2017.1278573
- Feng AL, Razavi CR, Lakshminarayanan P, et al. The robotic ENT microsurgery system: a novel robotic platform for

- microvascular surgery. *Laryngoscope*. 2017;127(11):2495-2500. doi:10.1002/lary.26667
10. Razavi CR, Wilkening PR, Yin R, et al. Image-guided mastoidectomy with a cooperatively controlled ENT microsurgery robot. *Otolaryngol Head Neck Surg*. 2019;161(5):852-855. doi:10.1177/0194599819861526
 11. Ahmed AK, Zygourakis CC, Kalb S, et al. First spine surgery utilizing real-time image-guided robotic assistance. *Comput Assist Surg (Abingdon)*. 2019;24(1):13-17. doi:10.1080/24699322.2018.1542029
 12. Kochanski RB, Lombardi JM, Laratta JL, Lehman RA, O'Toole JE. Image-guided navigation and robotics in spine surgery. *Neurosurgery*. 2019;84(6):1179-1189. doi:10.1093/neuros/nyy630
 13. Joskowicz L, Cohen D, Caplan N, Sosna J. Automatic segmentation variability estimation with segmentation priors. *Med Image Anal*. 2018;50:54-64. doi:10.1016/j.media.2018.08.006
 14. Sinha A, Leonard S, Reiter A, Ishii M, Taylor RH, Hager GD. Automatic segmentation and statistical shape modeling of the paranasal sinuses to estimate natural variations. *Proc SPIE Int Soc Opt Eng*. 2016;9784:97840D.
 15. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011;54(3):2033-2044. doi:10.1016/j.neuroimage.2010.09.025
 16. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004;11(2):178-189. doi:10.1016/s1076-6332(03)00671-8
 17. Dubuisson M-P, Jain AK. A modified Hausdorff distance for object matching. In: *Proceedings of 12th International Conference on Pattern Recognition*. IEEE; 1994:566-568. doi:10.1109/ICPR.1994.576361
 18. Li Z, Gordon A, Looi T, Drake J, Forrest C, Taylor RH. Anatomical mesh-based virtual fixtures for surgical robots. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE; 2020. doi:10.1109/IROS45743.2020.9341590
 19. Fauser J, Stenin I, Bauer M, et al. Toward an automatic preoperative pipeline for image-guided temporal bone surgery. *Int J Comput Assist Radiol Surg*. 2019;14(6):967-976. doi:10.1007/s11548-019-01937-x
 20. Meike B, Matthias K, Georgios S. Segmentation of risk structures for otologic surgery using the probabilistic active shape model (PASM). In: *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*. Proceedings vol 9036. SPIE; 2014. doi:10.1117/12.2043411
 21. Neves CA, Tran ED, Kessler IM, Blevins NH. Fully automated preoperative segmentation of temporal bone structures from clinical CT scans. *Sci Rep*. 2021;11(1):116. doi:10.1038/s41598-020-80619-0
 22. Nikan S, Van Osch K, Bartling M, et al. PWD-3DNet: a deep learning-based fully-automated segmentation of multiple structures on temporal bone CT scans. *IEEE Trans Image Process*. 2021;30:739-753. doi:10.1109/tip.2020.3038363
 23. Muyan-Ozcelik P, Owens JD, Xia J, Samant SS. Fast deformable registration on the GPU: a CUDA implementation of demons. In: *2008 International Conference on Computational Sciences and Its Applications*. IEEE; 2008. doi:10.1109/ICCSA.2008.22
 24. Li X, Gong Z, Yin H, Zhang H, Wang Z, Zhuo L. A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images. *Neural Netw*. 2020;124:75-85. doi:10.1016/j.neunet.2020.01.005
 25. Van Osch K, Allen D, Gare B, Hudson TJ, Ladak H, Agrawal SK. Morphological analysis of sigmoid sinus anatomy: clinical applications to neurotological surgery. *J Otolaryngol Head Neck Surg*. 2019;48(1):2. doi:10.1186/s40463-019-0324-0
 26. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imag*. 2015;15(1):29. doi:10.1186/s12880-015-0068-x
 27. Powell KA, Liang T, Hittle B, Stredney D, Kerwin T, Wiet GJ. Atlas-based segmentation of temporal bone anatomy. *Int J Comput Assist Radiol Surg*. 2017;12(11):1937-1944. doi:10.1007/s11548-017-1658-6