**Full citation:** Andersen SAW, Bergman M, Keith JP, Powell KA, Hittle B, Malhotra P, Wiet GJ. Segmentation of Temporal Bone Anatomy for Patient-Specific Virtual Reality Simulation. Ann Otol Rhinol Laryngol. 2021 Jul;130(7):724-730.
**DOI:** 10.1177/0003489420970217

# Segmentation of Temporal Bone Anatomy for Patient-Specific Virtual Reality Simulation

**Authors:** Steven Arild Wuyts Andersen, MD, PhD (1,2,3); Maxwell Bergman, MD (2); Jason P. Keith (4); Kimerly A. Powell, PhD (4); Brad Hittle, BSc (4); Prashant Malhotra, MD (1,2); Gregory J. Wiet, MD, FACS, FAAP (1,2)

**Affiliations:**
1. Dept. of Otolaryngology, Nationwide Children's Hospital, Columbus, OH, USA.
2. Dept. of Otolaryngology–Head and Neck Surgery, Ohio State University, Columbus, OH, USA.
3. Dept. of Otorhinolaryngology–Head & Neck Surgery, Rigshospitalet, Copenhagen, Denmark.
4. Dept. of Biomedical Informatics, Ohio State University, Columbus, OH, USA.

**Correspondence:** Steven Andersen, MD, PhD. Department of Otorhinolaryngology—Head & Neck Surgery, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark. E-mail: stevenarild@gmail.com.

**Conflicts of interests:** The OTOPLAN software was provided on loan by MED-EL. MED-EL did not play a role in the conception, execution, interpretation, or publication of the study. The authors have no conflict of interest to declare.
**Funding**: This research was supported by Nationwide Children's Hospital Research Institute Intramural funds and Dr. Andersen by the Independent Research Fund Denmark (8026-00003B).
**Financial disclosures:** None.
**Presentations:** This paper was presented as a poster at the Triological Society Combined Sections Meeting, 23–25 January 2020, Coronado, CA, USA.

**Objectives:** Virtual reality (VR) simulation for patient-specific pre-surgical planning and rehearsal requires accurate segmentation of key surgical landmark structures such as the facial nerve, ossicles, and cochlea. The aim of this study was to explore different approaches to segmentation of temporal bone surgical anatomy for patient-specific VR simulation.

**Methods:** De-identified, clinical CT imaging of nine pediatric patients aged three months to 12 years were obtained retrospectively. The patients represented normal anatomy and key structures were manually segmented using open source software. The OTOPLAN (CAScination AG, Bern, Switzerland) otological planning software was used for guided segmentation. An atlas-based algorithm was used for computerized, automated segmentation. Experience with the different approaches as well as time and resulting models were compared.

**Results:** Manual segmentation was time consuming but also the most flexible. The OTOPLAN software is not designed specifically for our purpose and therefore the number of structures that can be segmented is limited, there was some user-to-user variation as well as volume differences compared with manual segmentation. The atlas-based automated segmentation potentially allows a full range of structures to be segmented and produces segmentations comparable to those of manual segmentation with a processing time that is acceptable because of the minimal user interaction.

**Conclusion:** Segmentation is fundamental for patient-specific VR simulation for pre-surgical planning and rehearsal in temporal bone surgery. The automated segmentation algorithm currently offers the most flexible and feasible approach and should be implemented. Further research is needed in relation to cases of abnormal anatomy.

**Level of evidence:** 4.

**Key words:** temporal bone anatomy; segmentation; virtual reality surgical simulation; patient-specific rehearsal; pre-surgical planning; pediatric otology

## Introduction

Patient-specific virtual reality surgical simulation (VRSS) has been described as the "holy grail" in personalized surgical care as this would allow individualized pre-operative case rehearsal/planning with potential benefits such as increased surgeon preparedness, reduced operating time, and fewer surgical risks and complications.[1] In temporal bone surgery, patient-specific VRSS training is perceived to be useful and found to positively impact surgeon confidence in performing mastoid dissection.[2] High-fidelity VRSS is dependent on accurate representation of key anatomical structures and surgical landmarks such as the facial nerve, cochlea, chorda tympani etc. Current limitations to patient-specific temporal bone simulation relates to the processing of the clinical imaging data and the quality of the visualization of the temporal bone soft structures in the VR simulator.[3,4]

Delineation (i.e. segmentation) of the key anatomical structures and landmarks are necessary to build the corresponding 3D models and to have an accurate visual representation in the simulation environment. Segmentation of clinical imaging studies traditionally requires considerable effort by trained technicians or clinicians: either by time-consuming, manual slice-by-slice identification and outlining of the relevant anatomy in computer software or by substantial tweaking of segmentations resulting from semi-automated routines currently integrated in commercially available temporal bone simulators.[4]

Recent developments include an atlas-based approach for *automated* segmentation with minimal user interaction that has demonstrated a high degree of precision for temporal bone structures compared with manual segmentations.[5,6] However, this image processing currently also needs to be done outside of the simulation environment. A new commercially available tool—the OTOPLAN® otological planning software—uses a semi-automated and *guided* approach for 3D visualization and measurement of key landmark structures of the temporal bone for otosurgical cochlear implant planning and is targeted at clinician end-users.[7,8] How this guided approach performs in relation to segmentation of temporal bone anatomy remains largely uninvestigated.

A feasible method for segmenting temporal bone anatomical structures needs to be determined so routine pre-operative imaging can be used to efficiently generate high fidelity virtual models for VRSS. This requires considerations of the different segmentation approaches in relation to accuracy, work flow, and technical integration. In this study, we therefore wanted to explore three different approaches (manual, guided, and automated) to segmentation of temporal bone surgical anatomy to determine how the resulting segmented models compare, as well as the feasibility for clinical use for patient-specific VRSS.

## Material and methods

*Imaging data and processing*

De-identified clinical computed tomography (CT) imaging studies obtained for nine pediatric patients (age range 3 months to 12 years; 5 females and 4 males) with normal temporal bone anatomy were retrospectively requested. The imaging data represents naturalistic clinical conditions and were therefore obtained on different CT scanners (GE Medical LightSpeed VCT, Toshiba Aquilion One, and GE Medical Discovery CT750 HD) using different inner ear/temporal bone acquisition protocols.

Next, we used the Fiji[9] open source image processing software to convert the data from DICOM to the Nifti image format, to crop the imaging studies to include only the temporal bone of one side (resulting in two imaging series for each patient), and to re-scale these imaging series to the same resolution of .3x.3x.5 mm. Finally, the imaging series were re-converted to DICOMs for use with the OTOPLAN. The conversion between file formats does not introduce a loss of image quality since we did not use compression. Rescaling affects image quality in the up/down sampling process. However, this was of limited consequence for this study since the original (varying) resolutions of the datasets were close to the uniform resolution we chose.

*Manual segmentation*

The resulting 18 imaging series (9 left/9 right temporal bones) were manually segmented by the first author (S.A.) using the ITK-SNAP[10] open source software on a commercially available pen display (Wacom Cintiq Pro 13, Wacom Co. Ltd., Japan) (Figure 1A). The manual segmentation was done by outlining each structure in successive images without any other segmentation assistance such as thresholding. Time to segment each structure was measured. The incus and malleus, facial nerve, and cochlea were segmented. The resulting volume-based models were converted to STL format surface models in the ITK-SNAP software to allow comparison.

*Guided segmentation*

The OTOPLAN (CAScination AG, Bern, Switzerland) software on a Microsoft Surface Pro tablet device was used for guided segmentation (Figure 1B).[7] The software is currently not commercially available in the U.S. as it is not FDA approved for clinical use. The guided approach combines structured manual inputs (such as selection of specific structures and borders) with automated routines (such as segmentation based on thresholds).

The incus and malleus, stapes, facial nerve, and cochlea were segmented individually by two clinicians (S.A., M.B.) using the OTOPLAN, i.e. one clinician experienced with manual segmentation and one clinician representing the typical end-user with limited experience in segmentation. Time to segment each anatomical structure was measured. The software also allows for segmentation of chorda tympani, but since it was not visible at the obtained CT resolutions, we excluded it. The software uses different segmentation approaches depending on the structure: for the incus and malleus, a single combined segment is created based on the user's selection of a single point around the incudomalleolar joint and is based on thresholding; for the stapes the user selects three points (tip of the lenticular process, the anterior and posterior border of the oval window) and an idealized stapes model is then visualized using these coordinates; for the facial nerve, the user manually selects a number of points along the center of the facial nerve canal and a tube model is created, for which the diameter at different points can be manually adjusted in a panoramic view. Cochlear segmentation requires several steps: first, the cochlear view is defined by the user using the center of the modiolus, the basal turn and the round window; next, cochlear parameters (diameter, width and height) are calculated based on the user's selection of the following boundaries: round window, lateral wall, inferior and superior points on the lateral wall, and basal and apical center points; finally, the cochlea is segmented by selecting points along the basal turn, with fine tuning of the diameter similar to that used for the facial nerve. After these steps, a 3D model of the basal turn of the cochlea is rendered.

The OTOPLAN does not directly export the resulting surface models for the segmented structures but saves this information in an XML-format, which we used to reconstruct the models and calculate model volumes for comparison (details in Appendix 1).

*Automated segmentation*

The imaging series were segmented using the atlas-based approach to automated segmentation.[5] First, the user needed to rotate and orient the imaging series to match the OSU atlas orientation (within ~15 degrees). Then, the user only needs to manually select the incudomalleolar joint to translate the image for alignment. Next, the automated routine spatially registers the dataset to the OSU atlas space using the Elastix 4.7 open source toolbox[11] with registration performed within the otic capsule region. Then a second b-spline registration of the atlas into the patient imaging series space is performed. Finally, each segmented structure is transformed by this result and then refined: for the cochlea and ossicles, the routine involves identifying the bone surrounding the structures using otsu multilevel thresholding,[12] dilating each segment slightly and then masking the dilated segment against the otsu thresholded image; for the facial nerve this further requires adjusting the segment based on a factor of 0.75 of the standard deviation of the intensity value. The output of this automated segmentation were converted to surface models in the ITK-SNAP software for comparison.

*Outcomes and statistics*

The primary outcome was an evaluation of the usability of the tools for generating segmentations for the use in patient-specific VRSS, supported by feasibility considerations including time for segmentation. Quantitative comparisons were made of the segmentation volumes including comparison of guided models by the two clinicians and stapes and cochlear parameter point selections.

The sample-size was one of convenience as judged sufficient for the study purpose. Models were imported and compared using Python with the pymesh and numpy-stl libraries. The models from one of the 18 datasets would not output correctly due to re-scaling issues and was excluded. SPSS version 26 (IBM, Washington, USA) for MacOSX was used to analyze timing data (linear mixed models for repeated measurements) and sequence (i.e. learning curves)(linear regression). Volume differences were compared using paired samples t-test.

*Ethics*

The study was deemed exempt for ethical approval by the Nationwide Children's Hospital IRB (study ID 00000204).

**Results**

*Comparison of volume models*

Visual examples of the models resulting from segmentation using the three different approaches are provided in Figure 2.

The incus-malleus models resulting from the guided segmentation had a statistically significantly smaller volume than the corresponding manually segmented models (Table 1). For the facial nerve, the guided models were generally larger than the manually segmented models, however this was not statistically significant. Unsurprisingly—since the OTOPLAN software only segments the basal turn of the cochlea—the corresponding volume was on average less than half of the manually segmented models.

The models based on the automated segmentation were generally more similar in volume to the manually segmented models: the average difference in the volume was only statistically significantly different for the cochlea for which the automated routine resulted in slightly larger models.

*Comparison of between rater segmentation (guided segmentation)*

For the guided segmentation using the OTOPLAN, the combined incus and malleus model was the most consistent with only a small average difference in volume between the two users (Table 1). In contrast, the facial nerve and cochlea models had large average differences in volume. For the selection of the three stapes points, the average difference in distance between the corresponding points selected by the two users was 2.7 mm. For the six cochlear parameter points, the average difference in distance between the corresponding points selected was 3.6 mm.

*Time for segmentation*

Descriptive data for the time used for segmentation are presented in Table 2. Manual segmentation of the included structures took a median of 16 minutes, ranging from 10.0 to 49.0 minutes. There was no effect of sequence in manual segmentation time ($p=0.14$), suggesting that the segmenter was sufficiently experienced and on a learning curve plateau. For the guided approach, the median time used for segmentation was just under 5 minutes, ranging from 3.8 to 13.8 minutes (excluding the time to load data and adjust contrast in the software). There was a significant effect of user on segmentation time (linear mixed models, $p<0.0001$) meaning that the user could be consistently faster or slower in using the software, which most likely is explained by the users' different segmentation experience. The effect of sequence was not statistically significant ($p=0.09$). Automated segmentation took less than a minute for the automated processing to run but required 2–5 minutes of manual preparation time (rotating and aligning the datasets).

*Experience with the different platforms (pros and cons)*

Manual segmentation allows full control of the data and segmentation of potentially every structure of the temporal bone within the technician/clinician's ability to recognize these structures. A plethora of software for this purpose is available (including open source software) dependent on operating system platform and user preference. Some software packages allow for reconstruction in different planes as well as integration of semi-automated processing routines to aid segmentation. Manual segmentation is very time consuming, which is dependent on the number of structures segmented but also imaging resolution (high resolution equals more slices to go through), interface hardware (using a computer mouse versus using a pen display), and the user's technical aptitude.

The OTOPLAN software running on a tablet is aimed at the clinician user with the purpose of easy visualization and pre-surgical planning mainly for cochlear implantation such as guided decision on electrode type and length. It is fairly easy to import imaging studies in the DICOM format either by network or using a USB memory stick. The software is highly intuitive to use and offers step-by-step guidance through the many different relevant functions. This, however, requires user interaction in selecting points of the structures and further manual adjustments for refinement, which contributes to inter-rater differences. The software is not specifically developed for segmentation of temporal bone structures for use outside of the OTOPLAN software and therefore there is no easy way to export the resulting models for integration with VRSS. Precision of point selection may be increased by connecting an external mouse rather than using the touch screen of the device.

The automated, atlas-based segmentation requires the user to rotate and line up the dataset so it is the same orientation (within ~15 degrees) as the atlas first (if needed) which takes a couple of minutes. Once this is done, the user needs to identify and select the incudomalleolar joint (within a tolerance of ~1.5 cm) after which the routine runs and outputs the segmentations. This is fast and performs consistently with a reasonable precision compared with manual segmentation, potentially with all key structures automatically being segmented. The main disadvantage is less control over the segmentation compared with more user-involved approaches and therefore a desire to verify the output manually.

## Discussion

### *Synthesis of overall findings*

In this study, we have investigated three different approaches to segmentation of temporal bone anatomical structures that are key for patient-specific VRSS. Overall, manual segmentation is time consuming but also the most flexible approach for which structures to segment and choice of software. The OTOPLAN otological planning software is not designed for VRSS and is therefore currently limited by the structures that can be segmented. The different algorithms results in user-to-user variation as well as large volume differences compared with manual segmentation. However, the guided segmentation is fast and provides a structured approach appealing to the clinician end-user. The automated approach potentially allows a full range of structures to be segmented and produces segmentations more comparable in volume to those of manual segmentation. It requires minimal time of the user, with a high reproducibility and without the variability and inconsistency that is introduced by human users.

### *Comparison with the current literature*

Currently, only two papers describe experiences with patient-specific VRSS and both find potential benefits for temporal bone surgical training. Locketz et al. explored the use of patient-specific VRSS in relation to cadaveric dissection training, and used both manual and automatic thresholding to segment the tegmen, facial nerve, chorda tympani, sigmoid sinus, internal carotid artery, and ossicular chain in image processing software before import into the CardinalSim platform.[2] They report that "*total time spent on data upload, segmentation, and model creation was approximately 30 minutes per specimen.*" Arora et al. similarly evaluated the use of patient-specific VRSS of cadaveric temporal bones and found that the VoxelMan Tempo Simulator's automated processing routine required substantial interaction and that "*user activity associated with the segmentation process accounted for the majority of the upload time. The time for each upload decreased with experience (mean time: 21 min, range 10–40 min [...])*".[4] The segmentation times in these studies are thereby comparable to our manual segmentation times.

The imaging data were in both cases obtained under standardized conditions on cadaveric specimens only. Arora et. al used a clinical CT-scanner (Philips iCT 256) and obtained a 0.33x0.33x1 mm resolution. Locketz et al. used a clinical cone beam CT scanner (3D Accuitomo 170 ENT) for a 0.2 mm isotropic resolution, resulting in an average of 504 slices per specimen. Our experience with manually segmenting datasets at such high resolutions is an increase in time by a factor three. This highlights the need for implementing automated routines in the processing and segmentation because the guided and automated approach is less affected by increases in resolution. Finally, high-resolution datasets are needed to achieve high-fidelity VR simulation graphics and also isotropic datasets are optimal as this avoids "butter stick" voxels (i.e. rectangular cuboid).

*Study strengths and limitations*

A strength of this study is that we used naturalistic clinical imaging data representing different patient ages, data acquisition protocols and imaging resolutions. The small sample size was sufficient for our study purpose of evaluating feasibility of the three different segmentation approaches for patient-specific VRSS. A major limitation is that we only compared volumes of the segmentations without consideration of spatial placement of these volumes. The reason for this was the limitation in exporting the output from the OTOPLAN software. Further, because we needed to reconstruct the OTOPLAN models there could be some minor differences with the actual models within the software itself. Finally, the OTOPLAN includes a limited number of structures we could compare in this study. Manual and automated segmentation can segment other key landmarks such as the semi-circular canals, the bony surface over the sigmoid sinus etc. Further, we only had one person do manual segmentation because this is so time consuming, and it was not the intent of the study to compare manual segmentations by multiple raters. However, it is also important to consider the use case: for a quick overview of patient anatomy, reproducibility may not be a key concern, but for the use in patient-specific VRSS simulation it might matter more.

*Implications and future research*

Segmentation of surgical landmarks is fundamental for patient-specific VRSS for pre-surgical planning and rehearsal. The next step is to integrate the pre-processing routines (i.e. segmentation) for high-quality visualization in the VR simulation environment and evaluating the effects on training and surgical performance before implementation for routine clinical use. There is also a need to investigate how the segmentation algorithms perform in relation to abnormal anatomy such as vestibulocochlear malformations and facial nerve dehiscence.  This would be highly relevant since vestibulocochlear malformations constitutes a substantial number of for example pediatric cochlear implantation cases and also constitutes the cases where the added value of patient-specific VRSS pre-surgical rehearsal and planning is potentially highest. Currently, none of the automated tools are clinically available, which is due to FDA regulations. For the OSU automated routine, next step is to pilot the algorithms in a clinical setting for further evidence, which would be required before broad dissemination.

**Conclusion**

Feasible and accurate segmentation is the foundation for patient-specific VRSS for rehearsal and planning in temporal bone and we have investigated three different approaches to segmenting key surgical landmarks. Manual segmentation is time consuming and consequently impractical for routine use. The OTOPLAN is an exciting tool for surgeons and provides a structured approach to pre-surgical planning and guided decision making on cochlear implantation, but has limitations that currently makes it suboptimal for segmentation for VRSS. An automated algorithm performs well and requires only limited user interaction, which makes it the most feasible option for segmentation of temporal bone surgical landmarks and key structures for patient-specific VRSS. Regardless of approach, there is still considerable work in integrating the segmentation for high-quality visualization in the VR simulation environment and how to handle cases of abnormal anatomy.

**Acknowledgments:** The authors would like to thank Houchun Harry Hu, PhD, Dept. of Radiology, Nationwide Children's Hospital, Columbus, OH, for his assistance with providing the deidentified CT scans. The authors would also like to thank MED-EL for loan of the Otoplan device.

**References**

1. Aggarwal R, Brown KM, de Groen PC, et al. Simulation Research in Gastrointestinal and Urologic Care-Challenges and Opportunities: Summary of a National Institute of

Diabetes and Digestive and Kidney Diseases and National Institute of Biomedical Imaging and Bioengineering Workshop. *Ann Surg.* May 2017:1.

2. Locketz GD, Lui JT, Chan S, et al. Anatomy-Specific Virtual Reality Simulation in Temporal Bone Dissection: Perceived Utility and Impact on Surgeon Confidence. *Otolaryngol Head Neck Surg.* 2017;156(6):1142–1149.
3. Sethia R, Wiet GJ. Preoperative preparation for otologic surgery: temporal bone simulation. *Curr Opin Otolaryngol Head Neck Surg*. 2015;23(5):355–359.
4. Arora A, Swords C, Khemani S, et al. Virtual reality case-specific rehearsal in temporal bone surgery: a preliminary evaluation. *Int J Surg*. 2014;12(2):141–145.
5. Powell KA, Liang T, Hittle B, Stredney D, Kerwin T, Wiet GJ. Atlas-Based Segmentation of Temporal Bone Anatomy. *Int J CARS*. 2017;12(11):1937–1944.
6. Powell KA, Kashikar T, Hittle B, Stredney D, Kerwin T, Wiet GJ. Atlas-based segmentation of temporal bone surface structures. *Int J CARS*. 2019;14(8):1267–1273.
7. Otoplan Otological software, CAScination, Bern, Switzerland. https://www.otoplan.ch/ [Last accessed 10 Jan 2020].
8. Ping Lu, Barazzetti L, Chandran V, et al. Facial nerve image enhancement from CBCT using supervised learning technique. *Conf Proc IEEE Eng Med Biol Soc*. 2015;2015:2964–2967.
9. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676–682.
10. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116–1128.

11. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging. 2010;29(1):196–205.
12. Jianzhuang L, Wenqing L, Yupeng T. Automatic thresholding of gray-level pictures using two-dimension Otsu method. 1991 International Conference on Circuits and Systems, Shenzhen, China, 1991, pp. 325-327 vol.1

## FIGURE LEGENDS

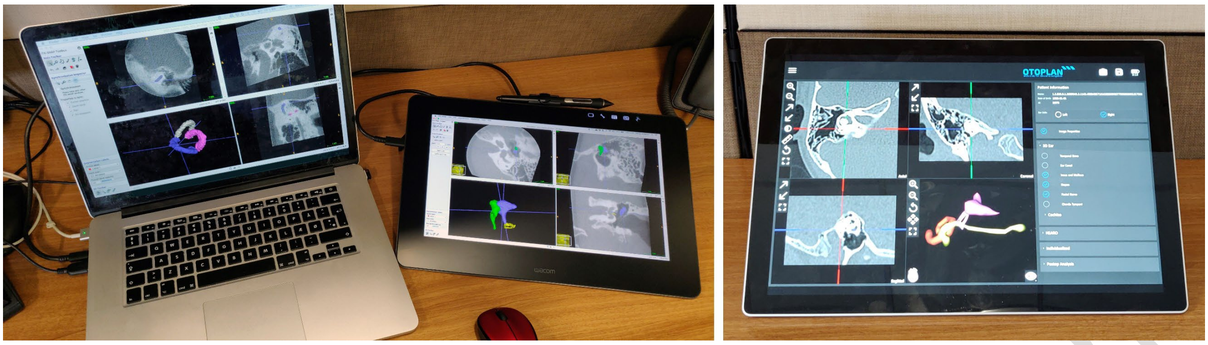


**Figure 1.** A) Laptop with Wacom pen display (left) for manual segmentation of the temporal bone anatomical structures using the ITK-SNAP software, B) the Otoplan for guided segmentation (right).

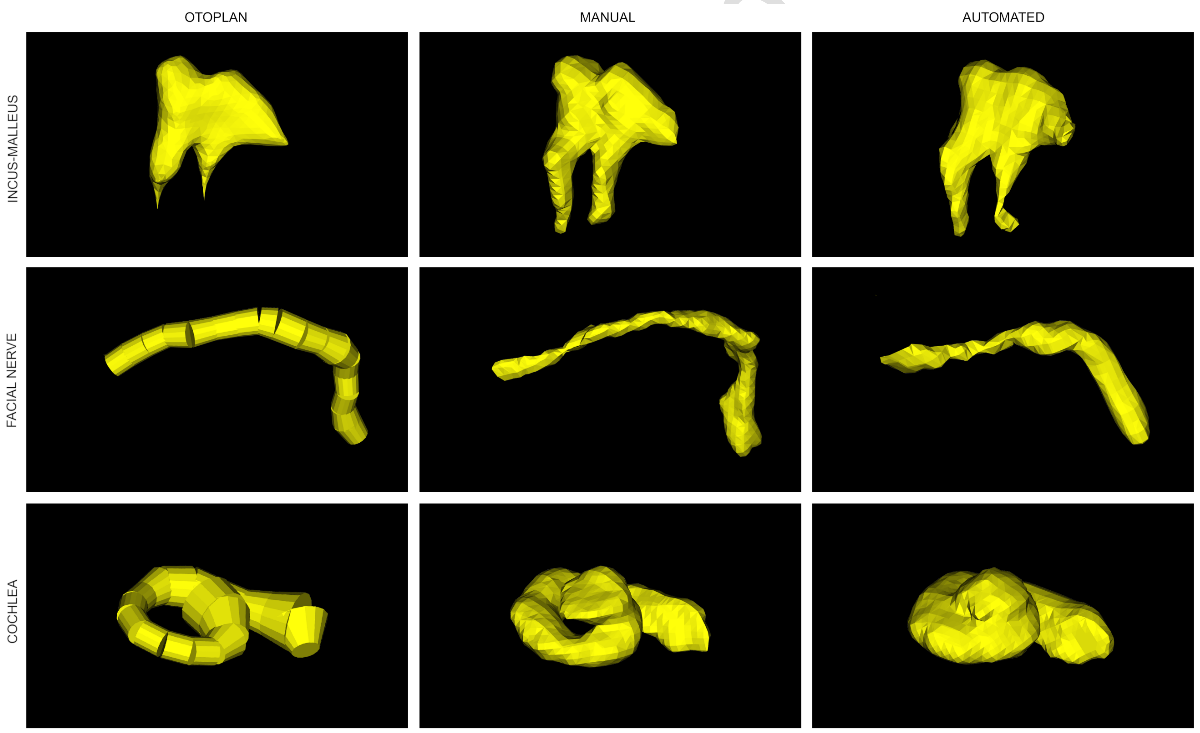


**Figure 2.** Visual example of segmented volume models.

**Table 1.** Comparison of segmented volumes.

| | Volume difference | | | |
|---|---|---|---|---|
| | Absolute ($mm^3$) | SD | Relative (%) | p |
| *OTOPLAN - between raters* | | | | |
| Combined incus-malleus model | 0.3 | 0.5 | 2.3 | 0.23 |
| Facial nerve model | 12.1 | 9.1 | 51.3 | **<0.001** |
| Cochlea model | 19.7 | 15.3 | 61.1 | 0.90 |
| *OTOPLAN vs. manual segmentation* | | | | |
| Combined incus-malleus model | 15.3 | 8.9 | 51.8 | **<0.001** |
| Facial nerve model | 12.8 | 11.1 | 145.2 | 0.23 |
| Cochlea model | 35.9 | 16.1 | 46.5 | **<0.001** |
| *Automated vs. manual segmentation* | | | | |
| Combined incus-malleus model | 3.7 | 3.0 | 15.9 | 0.32 |
| Facial nerve model | 7.8 | 8.0 | 39.3 | 0.23 |
| Cochlea model | 10.6 | 9.2 | 18.4 | **0.03** |

**Table 2.** Segmentation times.

| | Median (minutes) | Range (minutes) |
|---|---|---|
| **Otoplan** | | |
| Combined incus-malleus segment | 0.2 | 0.02–1.3 |
| Stapes points | 0.5 | 0.4–3.5 |
| Facial nerve | 1.5 | 1.0–3.8 |
| Cochlea | 2.7 | 2.1–6.0 |
| *All segments* | 4.9 | 3.8–13.8 |
| **Manual** | | |
| Incus | 1.0 | 1.0–2.0 |
| Malleus | 2.0 | 1.0–3.0 |
| Stapes | 2.0 | 1.0–2.0 |
| Facial nerve | 7.0 | 3.0–39.0 |
| Cochlea | 5.0 | 3.0–8.0 |
| *All segments* | 16.0 | 10.0–49.0 |
| **Automated** | | |
| Data preparation | 3.5 | 2.0–5.0 |
| Processing time (automated) | 0.9 | 0.8–1.0 |
| *All segments* | 4.4 | 2.8–6.0 |