Human Middle Ear Anatomy Based on Micro-Computed Tomography and Reconstruction: An Immersive Virtual Reality Development

Kai Cheng 1,*, Ian Curthoys 2, Hamish MacDougall 3, Jonathan Robert Clark 1,4,5, and Payal Mukherjee 1

1 Royal Prince Alfred Institute of Academic Surgery, Sydney Local Health District, Sydney, NSW 2050, Australia
2 Vestibular Research Laboratory, School of Psychology, The University of Sydney, Sydney, NSW 2006, Australia
3 School of Psychology, The University of Sydney, Sydney, NSW 2006, Australia
4 Department of Head and Neck Surgery, Chris O'Brien Lifehouse Cancer Centre, Sydney, NSW 2050, Australia
5 Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW 2006, Australia
* Correspondence: kai.cheng@health.nsw.gov.au

Abstract: Background: For almost a decade, virtual reality (VR) has been employed in otology simulation. The realism and accuracy of traditional three-dimensional (3D) mesh models of the middle ear from clinical CT have suffered because of their low resolution. Although micro-computed tomography (micro-CT) imaging overcomes resolution issues, its usage in virtual reality platforms has been limited due to the high computational requirements. The aim of this study was to optimize a high-resolution 3D human middle ear mesh model suitable for viewing and manipulation in an immersive VR environment using an HTC VIVE VR headset (HTC and Valve Corporation, USA) to enable a seamless middle ear anatomical visualisation viewing experience in VR while preserving anatomical accuracy. Methods: A high-resolution 3D mesh model of the human middle ear was reconstructed using micro-CT data with 28 µm voxel resolution. The models were optimised by tailoring the surface model polygon counts, file size, loading time, and frame rate. Results: The optimized middle ear model and its surrounding structures (polygon counts reduced from 21 million polygons to 2.5 million) could be uploaded and visualised in immersive VR at 82 frames per second with no VR-related motion sickness reported. Conclusion: High-resolution micro-CT data can be visualized in an immersive VR environment after optimisation. To our knowledge, this is the first report on overcoming the translational hurdle in middle ear applications of VR.

Keywords: human middle ear; anatomy; 3D; virtual reality

1. Introduction

The middle ear is a highly complex structure that plays a critical role in the human auditory system. As such, gaining a thorough understanding of its intricate and variable anatomy is essential for the diagnosis and treatment of a range of ear-related conditions. Advances in surgical techniques, such as endoscopic ear surgery, have made it more important than ever to have a detailed understanding of middle ear anatomy. However, creating high-resolution 3D visualizations and simulations of the middle ear in virtual reality (VR) has proven challenging due to limitations in imaging resolution, data size, and computing hardware. Standard computed tomography (CT) techniques, which are commonly used clinically for medical imaging, do not provide adequate resolution for creating detailed and realistic three-dimensional (3D) reconstructions of the middle ear’s complex structures [1]. This is particularly true for areas such as the lenticular process of the incus and the stapes superstructure, which have reduced bone density compared to other parts of the middle ear. Soft tissue structures, such as membranes and muscles, are also challenging to reconstruct using standard petrous temporal bone clinical CT scans [2].
The use of histology in the preparation of temporal bones is resource-intensive and limits the ability to increase a range of VR assets. As such, the use of histology-based preparation for temporal bones is reducing in frequency [3]. Instead, utilisation of microtomography (micro-CT) to overcome these challenges has increased in recent years to generate higher-resolution anatomically accurate data. However, the large amount of data generated by micro-CT is difficult to manipulate and computationally intensive, making it challenging to utilize in VR applications. This is a barrier to further applications of VR in simulation for otology. As a result, cadaveric temporal bones remain the gold standard for technical training in otology [4], especially when applied to tasks involving middle ear surgery.

Despite this, VR technology has shown increasing trends in application in otology, demonstrating that there is an ongoing demand to innovate in this area [3]. If newer forms of data sets of the middle ear, such as micro-CT, could be adopted into VR, it would open the possibility of attaining a variety of VR assets of the middle ear to simulate anatomical variation and disease without compromising anatomical accuracy. Additional advantages would include a reduction in processing time and the cost of development, making it a more scalable innovation [5].

This study sought to address these challenges by optimizing a mesh model of the middle ear generated from micro-CT data using the adaptive polygon optimization (APO) technique. The study examined the polygon count, mesh model loading time, and VR frame rate of the optimized mesh model to assess its feasibility for creating an anatomically accurate middle ear model in VR. To the researchers’ knowledge, this is the first study to explore the use of human middle ear micro-CT in VR otology visualization. These findings have important implications for the future of medical education and surgical training, suggesting that VR technology could be an effective tool for teaching technical skills in middle ear surgery.

2. Methods

This study was carried out following institutional ethics approval from the Royal Prince Alfred Hospital Ethics Office through the Research Ethics and Governance Information System (REGIS), protocol numbers 2019/ETH13789 and X19-0480.

An overview of the major development steps is provided in Figure 1.

![Figure 1](image-url). Development of an immersive virtual reality human middle ear anatomy based on micro-CT data.

2.1. Specimen Processing and Scanning

A sample of six freshly frozen temporal bones was dissected to extract the otic capsules with an intact tympanic membrane, which contains the middle ear. The bones were placed in Karnovsky’s fixative (3% paraformaldehyde, 0.5% glutaraldehyde in phosphate buffer) to prevent shrinkage. The specimens were bathed in diluted osmium tetroxide to allow soft-tissue staining. The specimens were placed in an Xradia MicroXCT 400 Micro Tomography scanner (Carl Zeiss AG, Oberkochen, Germany). The bones were scanned at a resolution of [resolution].
28 microns, and cross-section image sequences (1024 px × 1024 px) were saved in Tag Image File Format (TIFF) at 24-bit color depth. The specimen that showed the best soft-tissue structures was chosen for reconstruction.

2.2. Segmentation and Surface Model Reconstruction

Segmentation and surface model reconstructions were performed using a previously validated protocol [1], and image sequences were loaded into AVIZO 2021 3D Visualization & Analysis Software (Thermo Fisher Scientific Inc, Waltham, MA, USA). Middle ear regions of interest (ROI) were specified and volumetric models were rendered. A semiautomatic image segmentation procedure was implemented to reconstruct a 3D surface model [Figure 2].

![Figure 2. The middle ear and its surrounding reconstruction using AVIZO 2021 software.](image)

2.3. Adaptive Polygon Optimisation

Small polygons were combined into larger polygons using adaptive polygon optimisation (APO) software (Mootools software, Saint Vincent de Cosse, France) while maintaining surface details [Figure 3]. Using this method, the surface model size could be reduced dramatically. Eight surface models of the human ear and its surroundings were created with polygon counts ranging from 8.5 million to 600,000. The size of the mesh model was also reduced from 128 Mb to 5.5 Mb [Figure 3].
After the APO process, the surface model of the middle ear and its surroundings were reoriented using transform tools in Autodesk 3dsMax 2021 (Autodesk, San Rafael, CA, USA). Middle ear structures were extracted using the detaching tool. Some additional landmark structures, such as nerves and blood vessels, were not included in the model. These structures were created in 3dsMax 2021 software using primitive shapes such as tubes to keep a low polygon profile and provide a contextual reference. Color was applied manually to ossicles, nerves, tympanic membranes, ossicular ligaments, and surrounding structures [Figure 4]. Completed models were checked by an otologist (PM) and exported in FBX (Filmbox) format with embedded media options activated to include color information. The optimal combination of variables in the workflow, including data segmentation, surface model optimization, VR loading time, and frame rate, was then evaluated.

**Figure 3.** Incus surface detail comparison before and after adaptive polygon optimisation.

After the APO process, the surface model of the middle ear and its surroundings were reoriented using transform tools in Autodesk 3dsMax 2021 (Autodesk, San Rafael, CA, USA). Middle ear structures were extracted using the detaching tool. Some additional landmark structures, such as nerves and blood vessels, were not included in the model. These structures were created in 3dsMax 2021 software using primitive shapes such as tubes to keep a low polygon profile and provide a contextual reference. Color was applied manually to ossicles, nerves, tympanic membranes, ossicular ligaments, and surrounding structures [Figure 4]. Completed models were checked by an otologist (PM) and exported in FBX (Filmbox) format with embedded media options activated to include color information. The optimal combination of variables in the workflow, including data segmentation, surface model optimization, VR loading time, and frame rate, was then evaluated.
3. Results

The original micro-CT surface model contained over 21 million polygons, and translation to VR was not possible. Thus, reducing the polygon count through surface model optimisation as shown in Table 1 enabled acceptable loading rates and frame rates for VR applications. By using the APO, polygon counts could be reduced by 90% with the preservation of surface details. However, once the optimisation ratio exceeded 90%, the surface details started to fade. Ideal optimisation of the middle ear surface model for VR was achieved between 2 million and 2.5 million polygons, which enabled a loading time of fewer than 3 min at a frame rate of 82 to 90 frames per second with no VR-related motion sickness experienced.

Table 1. Polygon numbers, model size, loading time, VR frame rate, polygon optimization ratio and surface detail assessment.

<table>
<thead>
<tr>
<th>Polygons</th>
<th>Model Size (MB)</th>
<th>VR Loading Time</th>
<th>Frame Rate (FPS)</th>
<th>Optimisation Ratio</th>
<th>Surface Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.5 million</td>
<td>128</td>
<td>12 min and 20 s</td>
<td>35</td>
<td>59.5%</td>
<td></td>
</tr>
<tr>
<td>6.5 million</td>
<td>96</td>
<td>9 min and 45 s</td>
<td>51</td>
<td>69%</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Polygons</th>
<th>Model Size (MB)</th>
<th>VR Loading Time</th>
<th>Frame Rate (FPS)</th>
<th>Optimisation Ratio</th>
<th>Surface Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3 million</td>
<td>64</td>
<td>5 min and 10 s</td>
<td>63</td>
<td>79.5%</td>
<td>✔</td>
</tr>
<tr>
<td>2.5 million</td>
<td>32</td>
<td>2 min and 55 s</td>
<td>82</td>
<td>88%</td>
<td>✔</td>
</tr>
<tr>
<td>2 million</td>
<td>25</td>
<td>2 min and 30 s</td>
<td>90</td>
<td>90%</td>
<td>✔</td>
</tr>
<tr>
<td>1.5 million</td>
<td>19</td>
<td>1 min and 55 s</td>
<td>90</td>
<td>93%</td>
<td>✔</td>
</tr>
<tr>
<td>1 million</td>
<td>12</td>
<td>1 min and 20 s</td>
<td>90</td>
<td>95.2%</td>
<td>✔</td>
</tr>
<tr>
<td>600 k</td>
<td>5.5</td>
<td>30 s</td>
<td>90</td>
<td>97.1%</td>
<td>✔</td>
</tr>
</tbody>
</table>

4. Discussion

Simulation has been widely applied in otolaryngology [4,6–8]. In the last 2 decades, the application of 3D technology in simulation for education has increased [3]. This has been particularly prevalent in rhinology and otology, where reliance on technology has been necessary to attain minimally invasive routes for surgery. Due to the increased cost and reduced availability of traditional teaching resources such as human cadavers, there is a need for other technologies that are more readily available, cost-effective, and reproducible to facilitate the uniform assessment of competency [6,7]. Therefore, the use of 3D technologies such as VR has grown in prominence, particularly in temporal bone surgical training [3]. Advantages of VR include the development of standardized assessments, objective evaluation of experience level [8], and enhancement in surgical skills for both novice and experienced surgeons [9], as well as case-specific surgical rehearsal leading to improving confidence [10]. Limitations include attaining high-resolution data to create anatomical accuracy, attaining a range of anatomical models to simulate interindividual variation, simulating disease, and doing so in a manner that is not resource-intensive. Therefore, the application of VR has largely been limited to the technical skills that can be taught, such as mastoid surgery and cochlear implantation, but applications for use in middle-ear-specific tasks have been limited [11].

4.1. Strengths of the Reported Method

Sorensen et al. successfully demonstrated the use of VR in ear simulation over two decades ago [12]. The visible ear simulator [12] used images taken from one specimen to generate a volume-rendered model of the temporal bone to enable mastoidectomy surgery. Images were gathered using microslicing with slice width ranging from 50–100 µm. Segmentation of the specimen alone was reported to take 100–150 h, and a total of 450 h of labor was required to generate the entire visible ear model from one temporal bone. Almost 15 years later, using a protocol combining CBCT and microslicing data, 6 additional temporal bones were added to the open ear library [5]. Despite this advancement, image loss appearing as dehiscence and the need for manual segmentation was reported in important structures such as semicircular canals, facial nerves, and tegmen, demonstrating that limitations of image processing and labor intensity continue to be limitations in attaining high-resolution and anatomically accurate data sets. In addition, despite advances in CBCT resolution, the resolution of clinical imaging, especially for middle ear visualisation, still imposes limitations if imaging resolution is greater than 100 µm [1].

The micro-CT dataset used in this experiment used a resolution of 28 µm, and total image processing time to publish the dataset in VR was 30 h. Advantages of micro-CT include having access to a higher-resolution dataset, avoiding temporal bone processing artefacts (such as those created due to microslicing) and reduced segmentation time. This also allows the ability to increase the number of temporal bones, although micro-CT datasets are still limited to cadaveric bones.

The use of micro-CT to study middle ear anatomy dates back to 2003. Applications include finite element studies [13,14], prosthesis design [15], 3D printing [16], and anatomical evaluation of ossicles [17]. Neural anatomical structures of the middle ear such as the facial nerve [18] and chorda tympani [19] have also been described. Due to laborious data
processing [20] and expensive visualization software [21], studying the middle ear and its surrounding structures has been challenging for large groups of students and trainees. Combining the superior imaging quality of micro-CT with the simulation advantages of VR is a possible method for making such a high-resolution data set more readily accessible to a wider group of people, especially as VR headsets become cheaper and more user-friendly. As demonstrated in this study, if this dataset is processed by selecting specific strategies in rendering, image optimisation, and VR visualisation techniques tailoring to the simulation intent, this is indeed feasible.

4.2. Rendering, Image Optimisation and Visualisation Techniques

This study used surface rendering techniques with APO. The strengths and weaknesses of rendering techniques using volumetric data (voxels) compared to the surface models (polygons) have been reported in the literature (15). In representing anatomical details, Udupa and Hung (1991) [21] examined surface and volume rendering methods and concluded that the surface process has a slight advantage. In comparison, the volume rendering suggested by van Ooijen et al. [22] produced better image quality without loss of information compared to surface modelling, which only used a limited portion (i.e., the surface detail) of the available data. A recent study showed that while volume models gave superior volumetric data, detail of surface anatomy was sacrificed [2].

For virtual temporal bone drilling exercises, 3D volumetric voxel data is utilised, as it displays the entire raw 3D dataset without human interpretation (such as imaging segmentation by defining a threshold and 3D reconstruction). At each step of the virtual exercise, as the drill removes a volume of bone, a 3D array of voxels is removed and the whole model is updated to display the result of each surgical move [23,24]. However, rendering entire volumetric data stacks is a time-consuming, memory-intensive, and computationally expensive task, especially when dealing with large micro-CT data sets. As an alternative approach, a hybrid data structure is commonly used in some virtual procedure simulators [25], in which surface model vertices directly correspond to the volumetric representation. The graphically rendered surface model is dynamic and is updated while the voxels are being ‘drilled away’.

When studying surface anatomy, the inherent nature of the surface modelling technique allows good depth perspective and a higher definition of surface detail, which has been shown to be a significant advantage compared to volume modelling. Surface models have been reported [26–28] to allow precise clinical measurements of anatomical structures and are suited to the fabrication of tangible educational models using 3D printers. Other advantages of surface modelling include features such as geometry optimisation and the creation of colour and texture maps. In addition, with short computational times, relatively small data size and less memory and storage requirements, complex surface anatomy can be visualised when the surface modelling approach is deployed compared to volume modelling [29], which makes it more feasible for tasks requiring detailed visualisation.

Ultimately, the choice of volume, surface, or hybrid rendering can be customised to address the task at hand. In this study, where the purpose was to visualise anatomical structures of the middle ear and interact with whole shapes (i.e., ossicles) that did not involve drilling (typically required in virtual mastoidectomy), volume data was not necessary. Surface rendering was used to reduce file sizes while preserving the fine details of structures such as the posterior wall of the middle ear, the lenticular process and stapes, and the tendons and nerves traversing the middle ear [30,31]. Combining surface-rendering techniques with APO allowed image optimisation to attain a sufficient frame rate for an immersive VR experience.

A low VR frame rate (FPS) can cause motion sickness, resulting in nausea, headaches, sweating, and dizziness [32]. As a result, balancing other criteria such as polygon counts, surface model loading time, and surface model detail is crucial to obtaining an optimal frame rate. VR performance indicators reported in this study, such as polygon numbers,
model size, loading time, VR frame rate, polygon optimisation ratio, and surface detail assessment, offer useful references to help researchers choose an optimal VR parameter setup.

4.3. Limitations of the Reported Method

Reconstructing finer low-attenuation structures such as the middle ear membranes and tympanic diaphragm continues to be challenging. Increasing the greyscale threshold to include all soft tissue structures created artifacts by capturing osmium sedimentation. However, reducing it did not accurately allow the whole structure to be reconstructed. In this study, this required a manual design, which added to the processing time. These limitations can be overcome by rescanning the bone before and after staining and overlapping of imaging data, which in the future has the potential to further reduce processing time. In addition, as the clinical imaging resolution of CBCT improves, using patient-specific datasets that are also anatomically accurate may also become feasible.

5. Conclusions

In recent years, 3D animation simulators have become increasingly popular for medical education and training, particularly in the field of temporal bone training. However, the range and complexity of surgical tasks that can be simulated are limited by the lack of variation in available anatomy and the resolution of imaging techniques. It is also not possible to create anatomically accurate VR assets for middle ear disease. The middle ear, with its intricate and variable anatomy, presents a significant challenge due to variations in bone density and the presence of soft tissue structures such as membranes, ligaments, and nerves. Therefore, the introduction of micro-CT as a source of data for VR is a small but critical technological challenge, which this paper reports being able to overcome. As demonstrated in this study, selective rendering strategies can be utilized to attain task-specific assets, such as surface rendering for anatomical studies. Further, by optimizing a mesh model generated from micro-CT data using the adaptive polygon optimization technique, the assets can be better optimised for an immersive VR experience without compromising anatomical accuracy. This technological milestone marks a significant step forward in the use of VR in otology applications, and has the potential to revolutionize medical training and education by reducing time taken to create the assets, and thereby costs. This method might be used on a wider variety of specimens with further development, enabling a wider range of technical tasks to be performed with more accuracy and realism.

Author Contributions: Conceptualization, K.C.; Methodology, K.C. and H.M.; Formal analysis, K.C.; Investigation, H.M. and J.R.C.; Writing—original draft, K.C.; Writing—review & editing, I.C., H.M., J.R.C. and P.M.; Visualization, K.C.; Supervision, P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Study protocol was approved by the Institutional Review Board of Royal Prince Alfred Hospital (protocol numbers 2019/ETH13789 and X19-0480, 13 July 2019).

Informed Consent Statement: Informed consent was waived due to the anonymous nature of the data used in this study.

Data Availability Statement: In the spirit of collaboration and advancing scientific knowledge, we are open to sharing the anonymized CT scan data utilized in this study upon reasonable request.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.