THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE

PREPAREE A Laboratoire ImVIA

Ecole doctorale n°37

Sciences Pour l'Ingénieur et Microtechniques (SPIM)

Doctorat de Instrumentation, Informatique de l'image

Par

M. Raabid HUSSAIN

Augmented Reality for Middle and Inner Ear Surgical Procedures

Thèse présentée et soutenue à Dijon, le 26/11/2020

Composition du Jury :

Pr. CAVERSACCIO Marco	PU-PH, Universität Bern, Suisse	Président
Pr. PADOY Nicolas	PU, Université de Strasbourg, France	Rapporteur
Pr. NGUYEN Yann	PU-PH, Sorbonne Université, France	Rapporteur
Dr. GNANSIA Dan	Directeur R&D, Oticon Medical, France	Examinateur
Pr. BOZORG GRAYELI Alexis	PU-PH, Université de Bourgogne Franche-Comté, France	Directeur de thèse
Dr. LALANDE Alain	MCU-PH, Université de Bourgogne Franche-Comté, France	Co-directeur de thèse

école doctorale SPIM sciences pour l'ingénieur et microtechnique

Titre : Utilisation de la réalité augmentée lors des interventions chirurgicales de l'oreille moyenne et interne

Mots clés : chirurgie d'implant cochléaire, chirurgie mini-invasive, otologie, procédures transtympaniques, réalité augmentée, segmentation d'image de l'oreille.

Résumé: Les procédures otologiques impliquent la manipulation de petites structures délicates et complexes de l'anatomie de l'os temporal qui se trouvent à proximité de nerfs et de vaisseaux sanguins critiques. La réalité augmentée (RA) peut grandement être benéfique au domaine otologique en fournissant des informations supplémentaires anatomiques et de navigation fusionnée sur un seul écran. Cependant, bien que la navigation conventionnelle ait prouvé son utilité en otologie, le développement de la RA dans ce domaine reste limitée. Ce projet vise à développer des solutions RA pour les interventions chirurgicales de l'oreille moyenne et interne.
Nous proposons deux applications de la RA à cet égard. Dans la première application, des informations sur les structures de l'oreille moyenne sont obtenues à partir d'un examen tomodensitométrique préopératoire et sont superposées à la vidéo chirurgicale de la membrane tympanique. Cela permet au chirurgien d'avoir des informations en temps réel sur les structures anatomiques cibles et l'instrument chirurgical localisés derrière la membrane tympanique sans élévation du volet tympanoméatal. En prolongement de ce système, nous proposons également de visualiser le modiolus cochléaire sur la vidéo chirurgicale de l'oreille moyenne et interne permettant l'implantation transmodiolaire de l'implant cochléaire à travers le conduit auditif externe.
Les deux systèmes de RA proposés sont conçus de manière mini-invasive et sont uniquement basés sur des algorithmes de vision, éliminant la nécessité de systèmes traditionnels de suivi magnétique et/ou optique que permettant une installation dans l'environnement du bloc opératoire. Ce travail ouvre des perspectives importantes sur les procédures otologiques mini-invasives grâce à des solutions basées sur la RA.

Title : Augmented reality based middle and inner ear surgical procedures

Keywords : augmented reality, cochlear implant surgery, medical image segmentation , minimally invasive surgery, otology, transtympanic procedures.

Abstract: Otologic procedures involve manipulation of small, delicate and complex structures in the temporal bone anatomy which are in close proxmity of critical nerves and blood vessels. Augmented reality (AR) can highly benefit the otological domain by providing supplementary anatomical and navigational information unified on a single display. However, despite being composed of mainly rigid bony structures, the awareness and acceptance of possibilities of AR systems in otology is fairly low. This project aims at developing video-based AR solutions for middle and inner ear surgical procedures.
We propose two applications of AR in this regard. In the first application, information about middle ear cleft structures is obtained from a preoperative CT-scan exam and overlayed onto the surgical video of the tympanic membrane. This system provides the surgeon with real-time information about the anatomical target structures and the surgical instrument behind the tympanic membrane without tympanomeatal flap elevation. As an extension of this system, we also propose to visualize the cochlear modiolus in the real-time surgical video of the middle and inner ear cleft enabling transmodiolar implantation of the cochlear implant through the external auditory canal.
Both proposed AR systems are designed in a minimally invasive manner and are solely based on vision algorithms eliminating the need for traditional magnetic and optical tracking systems. The first trials showed an easy installation in the operating room environment. This work opens important perspectives into minimally invasive otologic procedures through AR-based solutions.

UBFC UNIVERSITE BOURGOGNE FRANCHE-COMTE
Université Bourgogne Franche-Comté
Batiment I3M, 64b rue Sully
21000 Dijon, France

To my 13 years old self who was against opting for medicine as he did not want to directly "play" with people's lives. Ironically, a decade later he ended up specializing in surgineering...

Acknowledgments

First and foremost, I would like to thank the Almighty, for giving me the strength, patience and knowledge that enabled me to efficiently and effectively tackle this project in time.

I would like to show my deepest gratitude towards my thesis directors, Alexis Bozorg Grayeli and Alain Lalande, for their motivation and relentless support throughout the course of this project. They continually conveyed a spirit of adventure in regard to research and implementation. Their enthusiastic nature towards the project kept my spirits high and they provided valuable information and guidance which proved vital in accomplishing various tasks of the project.

I would like to thank the project collaborators Caroline Guigou, David Fofi, Ioannis Stefanis and Roberto Marroquin for their expertise and help regarding data acquisition, visual perception and various medical facets of the project. This work would not have been possible without their assistance and guidance.

I would also like to thank the members of the thesis committee, Marco Caversaccio, Nicolas Padoy, Yann Nguyen and Dan Gnansia for taking time to read the thesis and provide encouragement and insightful comments. I am also indebted to Oticon Medical for providing financial and commercial support towards the project.

I thank my fellow labmates Kibrom, Arslan, Youssef, Clément, Antoine, Adrian, Nérée and Meriem for stimulating discussions and their companionship during the relentless working nights before the deadlines and for helping me in situations where my basic understanding of French proved futile. I would also like to thank the small Pakistani community in Dijon, particularly Babar, Danish and Abdul Qayyum, for making my time in Dijon enjoyable. Lastly, I would like to thank everyone for the fond memories and all the fun we had during the past three years.

Last but not the least, I would like to thank my family for their appreciation, support and well wishes. Their motivating words always prompted me to work harder and in turn doing better.

Contents

List of Figures

Glossary

μCT Micro Computed Tomography.

AR Augmented Reality.

CBCT Cone Beam Computed Tomography.

CNN Convolutional Neural Network.

CT Computed Tomography/Tomodensitometry.

DoF Degree of Freedom.

IF Impact Factor.

MIP Maximum Intensity Projection.

MIS Minimally Invasive Surgery.

MPR Multi-Planar Reformatting.

MRI Magnetic Resonance Imaging.

OR Operating Room.

Q Quartile Index.

SJR Scientific Journal Ranking.

SSD Shaded Surface Display.

VRT Volume Rendering Technique.

Chapter 1

Introduction

Augmented reality (AR) is a promising technology that enriches the sensory perception of the user by augmenting virtual contents directly on the environment. In contrast to conventional virtual reality, AR enhances reality rather than completely replacing it with a virtual environment. These auxiliary cues can be in different forms: visual, audio, haptics, taste and smell. In medical imaging, these virtual cues are most-often in the form of reconstructions from preoperative imaging data. AR is being used for patient and doctor education, disease simulation and surgical visualization with the aim of enhancing patient treatments and outcomes.

During the last two decades, AR has gained immense popularity but its use in the operating room is still under development and investigation. During surgery, AR can be delivered through goggles, screens, loud-speakers, gloves, joysticks or co-manipulated robots, etc. AR allows surgeons to incorporate additional pre or intraoperative data into the physical surgical field thus improving localization, approach, treatment, efficiency and safety [1].

Surgeons have endorsed image guided interventions as a means to avert perceptual distortions that downplay the impact of traditional imaging schemes such as endoscopy and microscopy. Image guided interventions have proven to outperform conventional procedures both in terms of better outcomes and reduced complications [2]. The main contribution of AR in surgery is the ability to see through structures and access hidden information without interfering with the surgical process. AR has been employed in surgical planning, intraoperative imaging, surgical navigation and target structure localization [2–4]. It can particularly be useful in highlighting critical structures, disease location, pathologies and risk regions in an intuitive manner. AR also has the potential to facilitate minimally invasive procedures by allowing surgeons to visualize structures without exposing them [5–7]. The main advantage that AR based procedures offer over traditional image guided procedures is the significant improvement in ergonomics of the system. With AR, everything can be available on a single view thus eliminating the need for the surgeon to go back and forth between different sensorial systems.

AR has been successfully applied in different surgical domains [1, 8, 9]. However, very few applications have been designed for otological procedures. Despite offering many advantages such as rigid bony structures and limited inter-structural movements, limitations in workspace and manoeuvrability coupled with high precision requirements have been the main hindrance in developing AR solutions for otological procedures. Thus, the aim of this work is to identify potential applications for the use of AR in this surgical domain.

In otological procedures, most of the concerned structures are prominently visible in computed tomography (CT) images and minimal deformation is observed between pre and intra-operative imaging. Thus, AR can be achieved by combining information from preoperative CT-scan images with the real-time video from the surgical microscope or endoscope. In this regard, the following objectives and applications were identified and addressed in our work:

- Identification and extraction of natural anatomical landmarks in middle ear that may be used for a point-based registration between CT-scan and surgical video.
- Providing accurate information about anatomical structures and surgical instruments behind the tympanic membrane during transtympanic middle ear procedures.
- Evaluating the possibility of implantation of an electrode array inside the cochlear modiolus through a transmodiolar approach.
- Automatic segmentation of inner ear structures for combining information from μCT-scan images of cadaver bones with clinical CT-scan images for detailed description of the anatomical structures.

The multi-disciplinary research undertaken in this thesis required interaction and coordination between researchers from different professional fields. Therefore, this work was accomplished by the combined efforts of computer vision researchers from ImViA laboratory (Dijon, France), medical expertise of surgeons from the ENT department at University Hospital of Dijon (Dijon, France) and our industrial partner Oticon Medical (Vallauris, France).

This dissertation is organized as follows: Firstly, the clinical background about the otological anatomy and different surgical procedures is presented in chapter 2. Ear is part of the cranial-base domain thus chapter 3 provides a literature review on the use of AR in cranial-base surgeries with emphasis on technical characteristics and applications. In chapter 4, we present a vision-based AR system for transtympanic middle ear procedures. This chapter entails different steps undertaken during development and validation of the system. This is followed by an extension of our AR system to realize transmodiolar cochlear implant surgery in chapter 5. In chapter 6, an auto-context inspired segmentation technique is proposed for automatic inner ear structure segmentation. Finally, in chapter 7 a summary of the outcomes and limitations of the developed systems is provided along with some propositions for future directions.

Chapter 2

Clinical Background

This chapter presents the clinical context related to our work. The first part of this section is dedicated to the anatomical information about the ear whereas the latter part focuses on different surgical approaches to otological structures that will be addressed in our work. These concepts are crucial to apprehend the principle of minimally invasive AR-based surgical solutions.

2.1 Ear Anatomy

Ear is situated bilaterally on the human skull in a symmetric arrangement as part of the temporal bone cavity. The main functions of the ear include hearing and maintaining a physical balance. Therefore, some common diseases related to the ear include deafness, motion sickness, vertigo, otitis, auricular hematoma, impacted cerumen, tinnitus, etc. The ear can be anatomically divided into three sections (outer ear, middle ear, and inner ear), each serving its own specific function [10] (Fig. 2.1):

2.1.1 Outer Ear

The outer ear consists of the auricle (or pinna) and the external acoustic meatus (or external auditory canal). Its purpose is to direct the sound waves onto the tympanic membrane (or ear drum). The main structures in the outer ear are:

- **Auricle:** It is the only part of the ear that is present outside the skull. It is a mostly cartilaginous structure covered by skin. The lateral aspect is concave and presents numerous grooves and ridges. Different parts of the auricle are depicted in Fig. 2.2. Its function

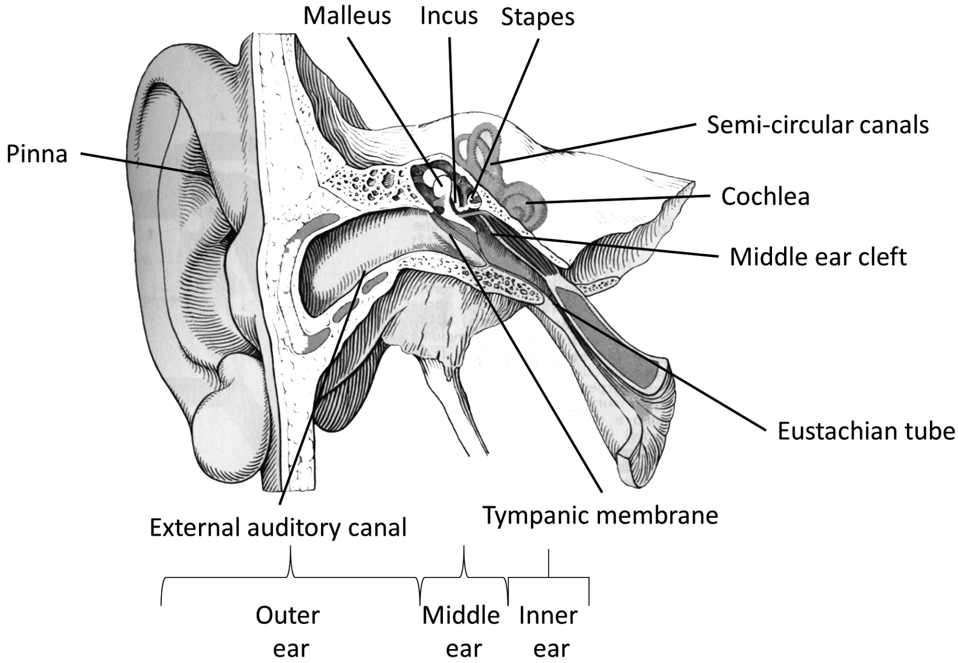


Figure 2.1: Anatomical description of the auditory structures.

is to capture sound waves from the environment and direct them towards the external acoustic meatus.

- **External acoustic meatus:** It is approximately an 8 mm diameter and 25 mm long oval shaped irregular hollow cylindrical structure, extending from the concha towards the tympanic membrane. The walls of the external acoustic meatus is composed of cartilage (1/3) and temporal bone (2/3). Its purpose is to direct sound waves collected at the auricle towards the tympanic membrane. It follows a sigmoid-shaped curvilinear path.

- **Tympanic membrane:** It is a connective tissue covered by skin (outside) and mucus (inside), connected to the surrounding temporal bone through a fibrocartilaginous ring which helps in maintaining the overall strength and integrity of the membrane. It is a very thin oval shaped translucent structure (with a diameter of 8-10 mm) connected at the distal end of the external acoustic meatus. Different parts of the tympanic membrane are depicted in Fig. 2.3.

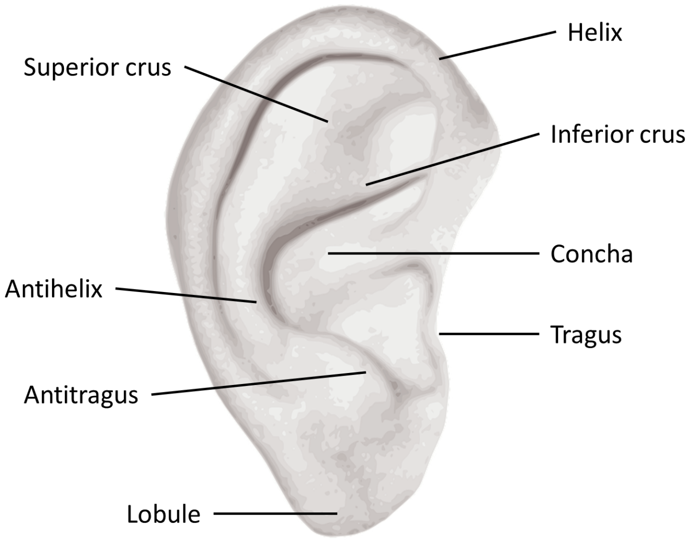


Figure 2.2: Anatomy of the auricle.

2.1.2 Middle Ear

The middle ear is located inside the temporal bone and is composed of the tympanic cavity containing the ossicles and the mastoid (Fig. 2.4). The ossicles are suspended in the cavity with the help of ligaments and tendons. The sound waves apply pressure on the tympanic membrane, enabling a vibratory motion in the middle ear chain/ossicles. The motion is then transferred to the fluids in the inner ear. The mechanical properties of the ossicles and the tympanic membrane/oval window surface ratio amplify the acoustic energy delivered to the inner ear. The main structures in the middle ear are:

- **Tympanic cavity:** It is a bony-membranous cavity in the middle ear, shaped like a biconcave lens. It contains the auditory ossicles (malleus, incus and stapes) which are connected in a chain-like manner and transmit sound vibrations from the tympanic membrane to the oval window through oscillatory motion. The six sides of the tympanic cavity are called: tegmental wall, jugular wall, carotid wall, membranous wall, labyrinthine wall, and mastoid wall.

- **Malleus:** The malleus is the largest and most lateral bone in the ear and is attached to the tympanic membrane via the handle of malleus. The head of the malleus lies in the epitympanic recess, where it articulates with the next auditory ossicle: the incus.

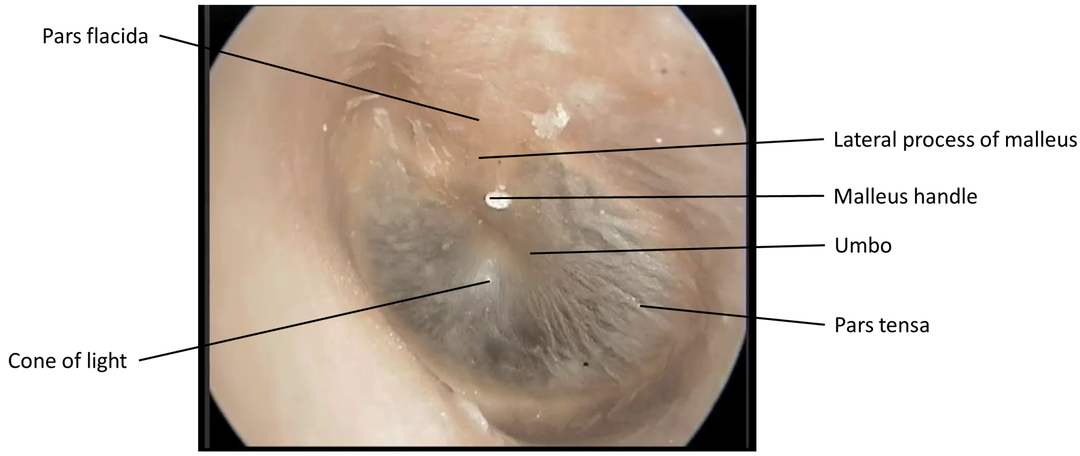


Figure 2.3: Anatomy of the tympanic membrane.

- **Incus:** The incus consists of a body and two limbs. The body articulates with the malleus, the short limb attaches to the posterior wall of the tympanic cavity, and the long limb joins the last of the ossicles: the stapes.
- **Stapes:** It is a stirrup-shaped structure (having a head, two limbs, and a base) which articulates with the incus. It joins the incus with the oval window and is the smallest bone in the human body.
- **Oval window:** It connects the middle ear to the vestibule in the inner ear and is covered by a membrane.
- **Round window:** It also connects with the inner ear and is covered by a membrane whose vibrations allow the fluid within the cochlea to move.
- **Promontory:** It is the projection of the first coils of the cochlea and the tympanic nerve passes through its groove.
- **Epitympanic recess:** It is located superior to the tympanic cavity. Malleus and incus partially extend into this recess.
- **Mastoid air cells:** They are a collection of air-pockets located in the mastoid antrum, posterior to the epitympanic recess. They act as a buffer system by releasing air into the tympanic cavity when the pressure becomes low.
- **Muscles:** The muscles, connected to the stapes and the malleus, contract in response to loud noise, reducing vibrations of the ossicles, and reducing the transmission of sound

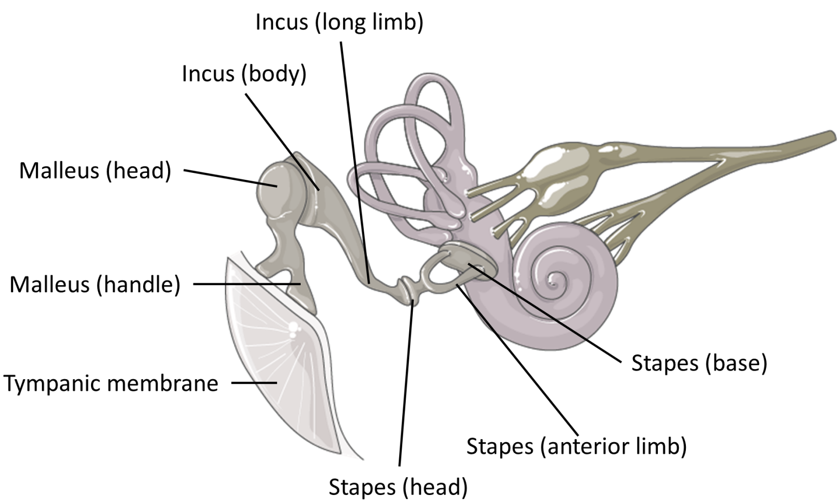


Figure 2.4: Anatomy of the middle ear (Adapted from `https://smart.servier.com`).

towards the inner ear.

- **Auditory tube:** It is a cartilaginous and bony tube connecting middle ear with the nasopharynx and is responsible for equalising the pressure between the middle ear and the external auditory meatus. It is also known as the eustachian tube. The tube is shorter and more straight in children, leading to a higher rate of occurrence of middle ear infections as compared to adults (since upper respiratory infection can spread through this tube to the middle ear).

2.1.3 Inner Ear

The inner ear is located in the petrous part of the temporal bone and comprises of the vestibulo-cochlear organs such as bony labyrinthe, semi-circular canals, utricle, saccule and cochlea (Fig. 2.5 and 2.6). The inner ear has two openings into the middle ear via oval window (between middle ear and vestibule) and round window (between middle ear and scala tympani). It transforms sound signals into electrical signals that are conveyed to the brain with the help of the auditory nerve. Cochlea helps in hearing while utricle, saccule and the semi-circular canals aid in balancing. Semi-circular canals also aid in eye-tracking. The main structures in the inner ear are:

- **Bony labyrinth:** It is the series of the following bony cavities: vestibule, cochlea and three semi-circular canals.

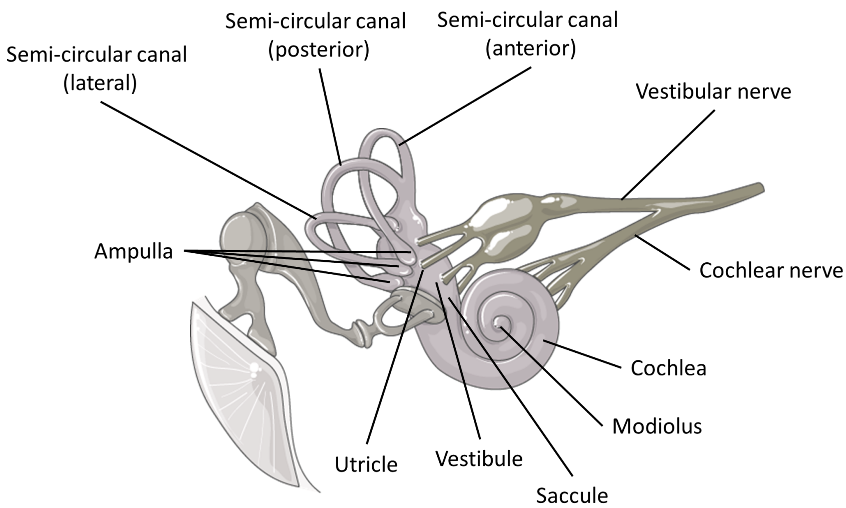


Figure 2.5: Anatomy of the inner ear (Adapted from `https://smart.servier.com`).

- **Vestibule:** It is the central part of the bony labyrinth separated from the middle ear by the oval window. It contains the saccule and utricle parts of the vestibular labyrinth. It also communicates with the posterior cranial fossa via the vestibular aqueduct.
- **Cochlea:** It is a spiral shaped structure in the inner ear responsible for hearing. It houses the triangular shaped cochlear duct (part of the membranous labyrinth) separating the scala tympani (located inferiorly and terminates at the round window) and the scala vestibuli (located superiorly and is continuous with the vestibule.) chambers. The basilar membrane of the cochlear duct houses the epithelial cells responsible for hearing. The cochlea is considered as the main part of the ear responsible for hearing. The basal portion of the cochlea is more sensitive to high pitch sounds whereas the apical portion is more sensitive to low pitch sounds. The auditory or cochlear nerve is attached at the cochlear modiolus.
- **Membranous labyrinth:** It is a continuous system of ducts present within the bony labyrinth. It comprises the saccule, the utricle, the cochlear and the semi-circular ducts.
- **Semi-circular canals and ducts:** The three orthogonal semi-circular canals contain the semi-circular ducts which are responsible for balance. Sensory receptors in the ampullae of the semi-circular canals detect movement and send respective signals to the brain.
- **Saccule and utricle:** They are membranous sacs located in the vestibule and are responsible for keeping balance by detecting movement of the head in vertical (saccule) and

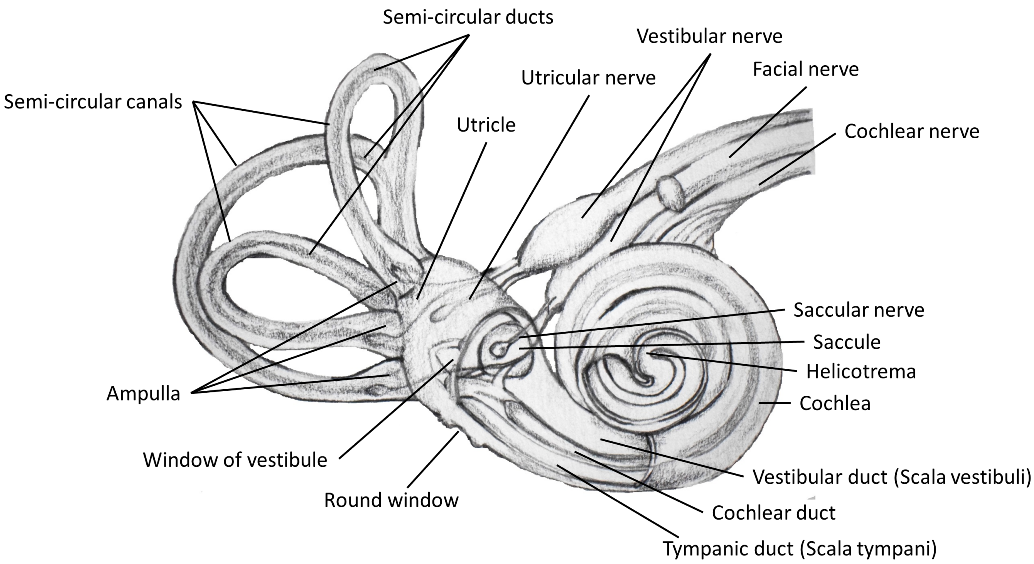


Figure 2.6: Intra-structural anatomy of the inner ear.

horizontal (utricle) directions.

- **Nerves:** Cochlear nerve is present at the base of the modiolus of the cochlea and transfers the electrical impulses transmitted from the the cochlea to the auditory receptors in the brain. Vestibular (responsible for balance) and facial nerves are also present in the inner ear.

2.2 Computed Tomography/Tomodensitometry

A CT-scan is of particular interest in otological domain as high contrast can be achieved between high density bony structures (such as the auditory canal and ossicles) and the low density areas/tissues (such as tympanic membrane and air). Pre-operative CT-scan images have been frequently used in image guided otological procedures as they can provide precise anatomical information during surgery.

CT is a non-invasive medical imaging procedure that uses special X-ray equipment to capture detailed images of the areas inside the body. Similar to regular X-ray imaging, the presence and the visibility of bodily structures in the output image depends on their individual ability to block X-ray beams. A CT-scan agglomerates a series of X-ray signals captured at different angles through a mobile X-ray tube and a stationary detector ring (Fig. 2.7). The detectors measure the intensity profile of the residual X-ray beam (after passing through the body of the patient and/or the target structure) at different angles of projection. Back-projection algorithm is then performed on the intensity profiles to construct a 3D matrix of the volume under study (typically stored in DICOM format). On this 3D matrix, the following reconstruction algorithms are commonly applied for appropriate visualizations:

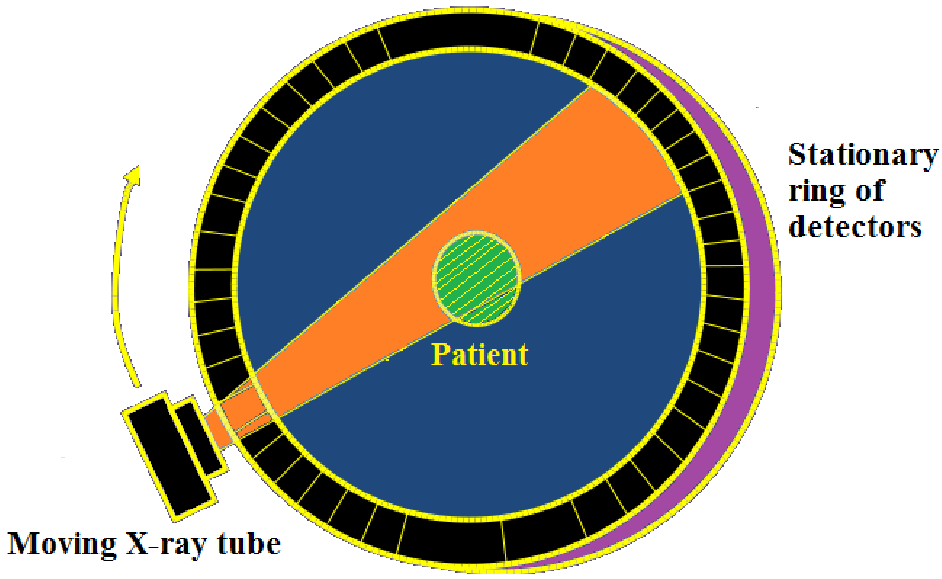


Figure 2.7: CT acquisition. The moving X-ray tube projects X-ray beams towards the patient (or more specifically the temporal bone in our case). The residual beams reach the detector creating a pseudo X-ray image. The acquisitions from multiple tube locations can provide a 3D image.

- **Multi-Planar Reformatting (MPR):** MPR is commonly used on thin-slice data from volumetric CT imaging. It involves the process of converting data from a 3D volume into a given plane such as axial, coronal and sagittal. An MPR view of a typical otological CT-scan is shown in Fig. 2.8. Different manipulations can be performed on the reconstructed data for specific purposes. Curvillinear MPR is an example of this which involves tracing a curved structure in the matrix with the aim of generating a planar image that transects the structure along one of its axis. A part of this work employed this functionality to help in optimizing the cochlear implant choice by computing patient-specific cochlear lumen characteristics [11] (Appendix A).

- **Maximum Intensity Projection (MIP):** MIP is useful for visualizing all the dense structures in a given volume. It refers to the process of projecting voxels with highest attenuation value onto the visualization planes throughout the CT volume. This implies that the MIP images from opposite viewpoints would be symmetrical. MIP reconstructions are fast however, they do not provide a good sense of depth. Alternatively, local MIP may be employed which provides a sense of occlusion as well.

- **Volume Rendering Technique (VRT):** VRT is a technique to display a projection of a discretely-sampled 3D volume data according to Hounsfield units of the different structures onto a camera plane. VRT reconstructions may be realized by either extracting isosurfaces from the volume and rendering them as polygonals or by rendering the entire volume as a single block. VRT can yield better object contours and semi-transparent displays but is computationally expensive.

- **Shaded Surface Display (SSD):** SSD is a surface rendered display obtained after segmenting the concerned structures in the CT volume from the background according to a unique threshold (in Hounsfield units). SSD provides a realistic 3D representation of the data. Virtual endoscopy is a useful application of SSD where a 3D reconstruction of the CT volumetric data is considered and a virtual endoscope camera can be traversed inside and around the reconstructed volume. An example of a SSD reconstruction of a temporal bone cadaver for virtual endoscopy is shown in Fig. 2.9.

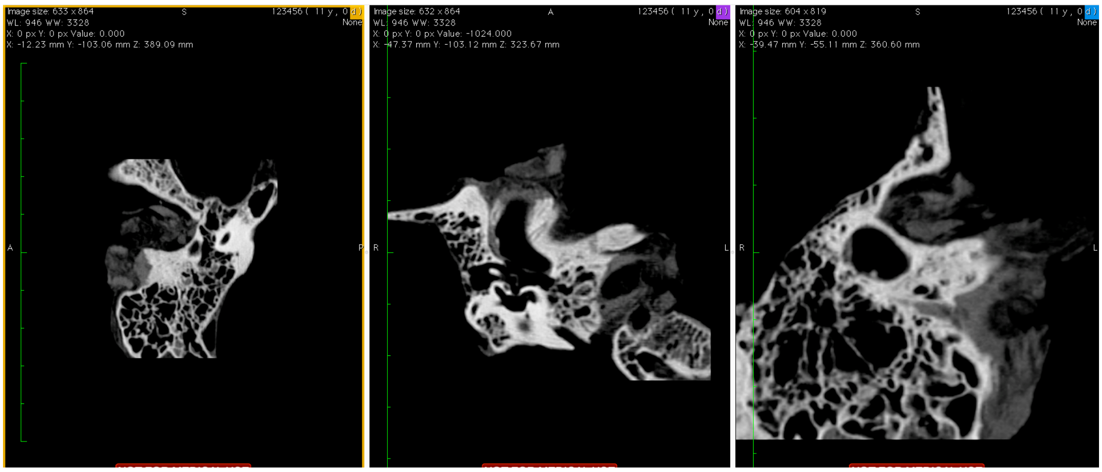
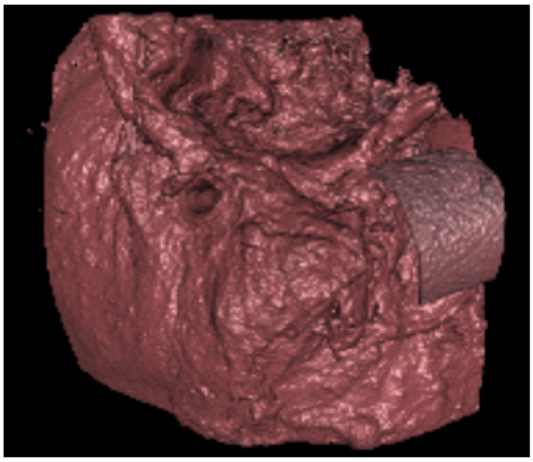


Figure 2.8: MPR reconstruction of a human temporal bone cadaver (axial, coronal, and sagittal views).

Clinical high resolution CTs (typically 0.2 mm) provide good resolution for inter-structural details of the ear anatomy. Cone beam CTs (CBCTs) are also increasingly being used in the operating room (OR). In contrast to clinical CT scanners which employ fan X-ray beam tube, CBCT scanners use a cone X-ray beam tube covering a relatively larger volume during

Right human temporal bone specimen.



3D modelling of the specimen.

Figure 2.9: 3D reconstruction of a temporal bone cadaver used for virtual endoscopy.

one single rotation around the patient. CBCT offers many advantages over clinical CTs such as faster acquisition time and lower radiation exposure, however, they produce low-resolution images that are not as anatomically correct as clinical CTs.

Clinical CTs and CBCTs are acquired at relatively lower resolutions for accurate intra-structural modelling of sub-millimetric structures in the ear. Micro-CTs (μCTs) can be used alternatively which provides a higher resolution substitute (typically 5-50 μm). However, in-vivo imaging is often not possible due to the small acquisition workspace of μCTs (100 mm diameter sphere). Thus, this technology is only applied on cropped cadaveric specimens. The detailed structural models derived from μCTs can be registered with clinical CTs to obtain detailed information about the ear [12]. Recently, a μCT dataset was published for inner ear structures [13]. Manual annotation of μCT images takes a significant amount of time since there are a lot of slices. A part of this thesis focused on developing convolutional neural netwtork (CNN) based automatic segmentation algorithms for inner ear segmentation. Manual annotation of data is mandatory to train a CNN.

In otology, although CT is the preferred imaging technology, magnetic resonance imaging (MRI) may sometimes also be used as a complementary imaging modality. It can help in identifying fluids within the membranous labyrinth, chorda tympani and facial nerves [14–16].

2.3 Operating Room Environment

An OR is a sterile, well-ventilated and organized environment. ORs are designed for the surgical team to perform surgical procedures which require focus, patience, time and safety. Attention should be paid in designing the OR setup such that the different components enable a well-coordinated operation. The surgical team needs to adhere to different protocols before entering the OR in order to minimize the spread of micro-organisms and maintain a sterile environment. A traditional otological OR is depicted in Fig. 2.10. Different components of a typical OR environment are:

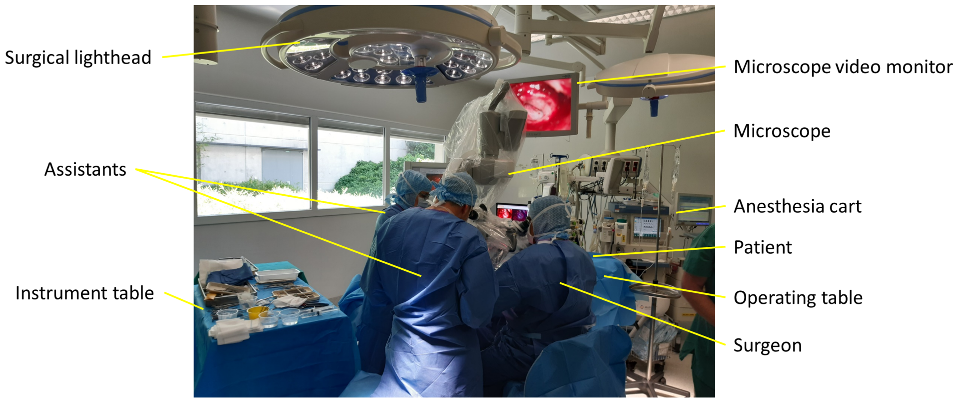


Figure 2.10: Operating room environment.

- **Operating Table:** An OR table is essential in any OR and different components of the OR are positioned around the table. The purpose of the table is to keep the patient in the correct position and, if required, help in moving the patient. During an otological procedure, the patient normally lies on his back with his head turned such that the ear being operated is on the top. The OR table can be stationary or mobile and has three components: a table column, the table top and the transporter.

- **Anesthesia Machine and Cart:** Anesthesia refers to the administration of drugs to inhibit the patient's mind and prevent the patient from feeling pain during the procedure. The anesthesia machine contains ventilator and oxygen support system and is a means for administering anesthetic gases for sedation or total anesthesia. The anesthesia cart holds additional medications for anesthesia and the respiratory support system hardware. These carts should be well-organized to allow anethesiologists to have easy access to all anesthesia tools for administration and monitoring.

- **Instrument Table:** This is a sterile table where the sterile surgical instruments are placed. This table should be positioned appropriately to allow the assistant nurse to pass the instruments quickly to the surgeon.

- **Sterile Personnels:** These include the surgeon, surgical assistants and the scrub nurse. The sterile OR area is normally created just before starting the procedure and should be continuously monitored. The sterile personnels should remain within the sterile area, carefully manoeuvring in the area.

- **Non-Sterile Personnels:** These include the anesthesiologist, circulating nurses, technicians, potentially students and other observers. It is recommended for the non-sterile personnels to keep a distance of at least 1 foot from the sterile area and be careful while traversing around the area.

- **Microscope/Endoscope:** Otological procedures involve manipulation of very small structures raising a need for surgical microscopes or oto-endoscopes in order to effectively observe the surgical targets. The choice of which imaging modality to use depends on the site of the body to be observed and the type of surgery being performed. Some procedures require visualization of the surgical scene by other surgical team members, thus these imaging systems often come with high definition display monitors as well. The microscope should be sterile and covered by a plastic film.

- **Surgical Lighthead:** High-quality lighting is essential for performing intricate surgical procedures. LED and halogen lights with low heat dissipation are generally used in the OR. Surgical lightheads are positioned on top of the surgical scene and are designed to provide bright white light to illuminate the surgical scene and eliminate any shadows.

- **Nurse Station:** These are used by OR nurses to document the relevant medical information which is important in keeping the patient's medical record up-to-date. These stations normally house electronic equipments and a desk for reporting purposes.

Integration of any surgical or navigational system should be seamless into this environment and not obstruct the surgical workflow. Some AR systems employ external tracking systems that need to placed near the patient and may require rearranging the OR setup. This work focuses on vision-based AR systems that only require a computer as an additional equipment which can be placed next to the microscope or instrument table requiring no rearrangement of the setup.

2.4 Minimally Invasive Surgery

Open surgery is considered as the most conventional option to examine and repair lesions in the human body. Despite providing the largest access to the target structures, this strategy causes more tissue damage around the approach, and may entail increased bleeding, infection, pain complications and longer postoperative stays at the hospital [17]. Additionally, the incisions have to be made carefully in order to avoid damaging vital structures present in the region.

The concept of keyhole or minimally invasive surgeries (MIS) was introduced in the 1980s as an effective and safe alternative to traditional open surgeries. MIS pertains to the process in which surgical instruments and an endoscope are passed through a small puncture in the skin (and sometimes in other structures such as the skull) to access target structures. Although it counters most of the drawbacks pertaining to traditional approaches, it also usually introduces additional instruments into the framework thus increasing complexity. The limited view provided by the endoscope, the lack of haptics and intricate movements introduce new complications into the surgical setup: the lack of 3D view may alter the estimation of the target position in relation with vital structures, whereas the lack of haptic feedback may interfere with the assessment of tissue resistance and modify the way the surgeons manipulate them [1]. Moreover, key-hole access may also include injuries along the trajectory since the lateral view in the approach tunnel is nearly absent [18]. These limitations are especially significant in otological surgery as it deals with critical vascular and neural edifices within a confined workspace. These factors may influence surgical time and surgeons comfort and stamina ultimately leading to a negative effect on the surgical outcome and efficiency [17, 19]. Consequently, tools that improve visualization, location and orientation are highly beneficial particularly when natural anatomy is eclipsed.

MIS has been applied in different domains like colonoscopy, bronchoscopy, rhinoscopy, etc. However, few works have been presented in the literature related to image guided minimally invasive otologic surgery [1]. Surgeons dealing with otological diseases use surgical microscope for magnification and focused lighting. The advantage of this tool is the absence of encumbrance in the surgical field, the binocular 3D vision, and the possibility of using both hands for surgery. Although the endoscopes provide a 2D view of a limited field and encumber the route, they can improve other aspects of vision by providing an angled view and/or a wide-angle exposure [20]. Image guided navigation can be coupled to both microscope and endoscope. The focal point of the microscope or the tip of the endoscope and a pointer instrument can be tracked on preoperative images in real-time thanks to infrared or magnetic tracking of these tools [21, 22]. These developments have enabled MIS procedures to be more precise and safe.

Different robotic systems are also being designed for otological surgeries that will enable MIS approaches [23–25]. The potential of robotic assistance in this field for enhanced performance is immense as micro-surgical gestures are needed with sub-millimetric precision. However, dedicated manipulators are required owing to the geometrical complexity of the otological anatomy [23]. AR also has the potential to improve MIS procedures by providing additional navigational guidance and visualizations of obstructed anatomy as well as extended visualizations of surrounding structures [1, 26, 27].

2.5 Transtympanic Procedures

Transtympanic surgery is a form of minimally invasive technique used in otology [28]. Different transtympanic procedures have been designed for diagnosis and therapeutic purposes. These procedures involve penetration of micro-instruments through a very small and deep cavity inside the ear by creating an opening in the tympanic membrane with the aim of accessing the middle ear cleft structures such as labyrinthine windows and ossicles [29–32]. These techniques have been used in different applications such as ossicular chain repair, drug administration, labyrinthe fistula diagnosis, extensive cholesteatoma removal, vertigo prevention, electrocochleography, etc. [33, 34].

Figure 2.11 depicts the view of the tympanic membrane from an oto-endoscope during a tanstympanic procedure. This approach offers many advantages compared to traditional surgery: faster procedure, preservation of tympanic membrane, reduced bleeding, simpler and less painful post-operative process, etc. However, such procedures introduce complications into the surgical process by minimising the available operative space and field of view, and limiting the instrument movement. Also, the instruments (surgical tools and endoscope) have to be very carefully manoeuvred as such not to rupture the delicate tympanic membrane.

AR is seen as a possible solution to address the aforementioned issues as it can provide an enhanced view of the middle ear cleft with enhanced ergonomy and high precision. A typical surgical AR system comprises an endoscope or microscope that captures the real image of the surgical target, a processing module that tracks the relation between the target and the camera positions and processes virtual objects, and a displaying unit that displays the enhanced environment by merging the information from the video camera and the virtual objects [3, 5, 35, 36]. Virtual objects in transtympanic surgery may refer to preoperative reconstructions of the ossicles and labyrinthine windows. A part of this thesis focuses on developing and evaluating an AR system for transtympanic procedures.

2.6 Cochlear Implant Surgery

A cochlear implant is a surgically implanted electronic device that replaces the function of the ear. For patients with severe or profound hearing loss, cochlear implants have been considered

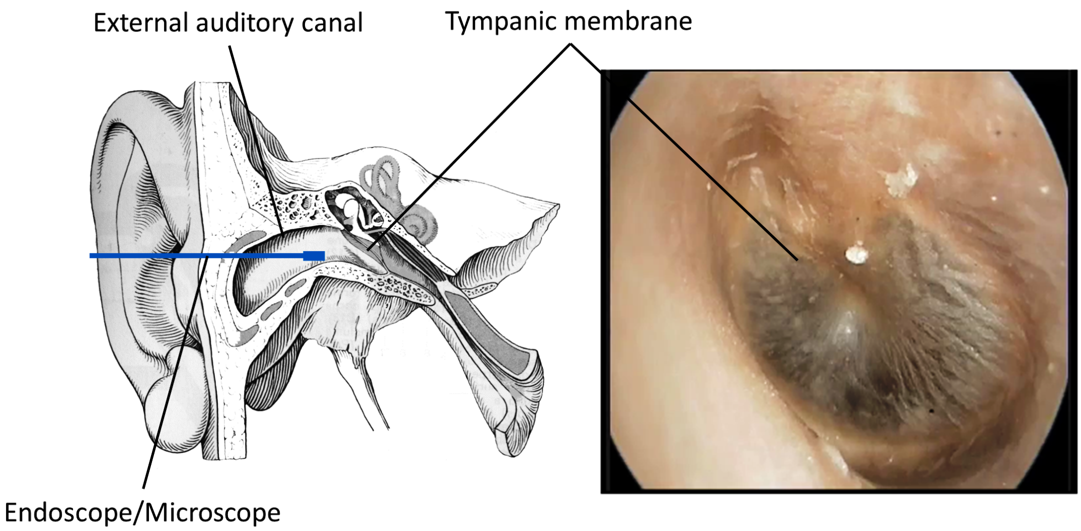


Figure 2.11: Tympanic membrane visualization using an endoscope or a microscope.

standard clinical treatment for several decades. The cochlear implant supplants the cochlea and transforms sound signals to electrical stimulation and delivers it to the auditory nerve. The implanted patient has to go undergo auditory training and speech therapy to adapt to the new auditory stimuli.

The implant is composed of two main parts (Fig. 2.12). The outer part, worn behind the ear, receives, processes, and encodes the acoustic signal from the environment. The main processing steps include automatic gain control and noise reduction. It then transmits the information and the energy to the inner part by magnetic induction through closed skin. The inner part delivers the encoded data to the auditory nerve endings by an electrode carrier placed inside the cochlea. The electrode carrier is typically an array of 20 electrodes which produces significant electrical stimulation around the damaged hair cells in the scala tympani leading to direct stimulation of the residual auditory neurons. The electrode needs to be inserted carefully as it may incur lesions of the modiolus, osseous spiral lamina or basilar membrane which may result in ganglion cell degeneration [37].

Conventional cochlear implants are inserted into the tympanic ramp (scala tympani) of the cochlea through the round window or a cochleostomy. This entry point is reached after drilling a bony cavity behind the auditory canal (mastoidectomy). The insertion is performed manually with limited visual and sensitive feedback. This approach, designed several decades ago, was best suited to limited technical possibilities and the absence of navigation in 60s and 70s [38]. It is still used in routine practice. However, with this technique, the electrode array is several

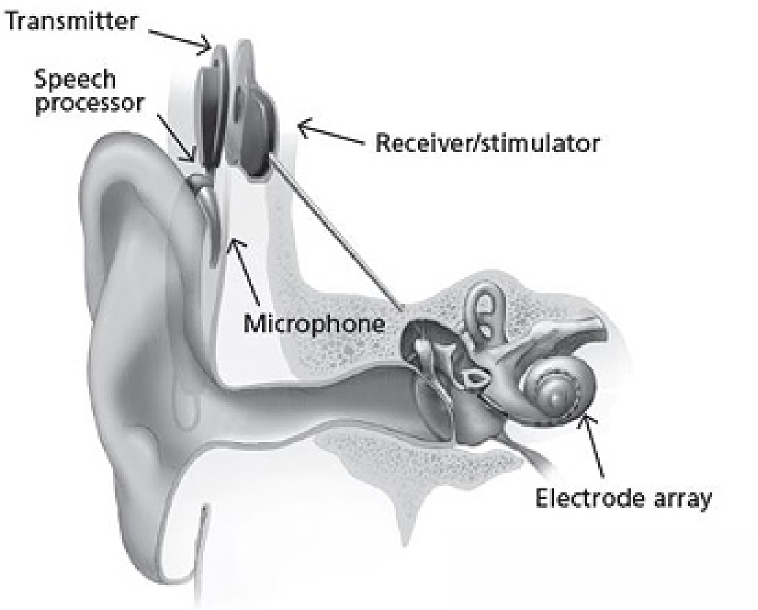


Figure 2.12: Different components of a cochlear implant. (Retrieved from `https://www.nidcd.nih.gov/`).

millimetres away from the nerve endings, making electrical stimulation less specific and less efficient [39] (Fig. 2.13).

Different variants of cochlear implants have been proposed in order to increase the performance. The number of electrodes, rigidity, curvature, electrode length, width and spacing may be altered according to the patient-specific cochlea model to better align the electrode array in the scala tympani. Different implantation techniques have also been proposed such as the total drill-out technique, the intact canal wall drill-out procedure, the short inferior tunnel insertion, the double electrode array, etc. [40–43]. However, the nerve-electrode contact is still poor yielding sub-optimal results [44].

These drawbacks led to the idea of the auditory nerve implantation (Fig. 2.13). Placing the array in contact with the nerve would theoretically increase specificity and reduce energy consumption [45, 46]. However, a precise localization of the nerve required complete drilling of the cochlea and this technique was abandoned [47]. More recently, the feasibility of cochlear nerve implantation without cochlear destruction via a transmodiolar approach was investigated with the help of neuronavigation [48]. This approach provides a direct access to the auditory nerve in the modiolus at the center of the cochlea without middle ear or cochlear destruction. In this technique, the cochlear apex (entry point) is reached through the external auditory canal since it is projected at the anterior part of the middle ear cavity. One of the surgical challenges

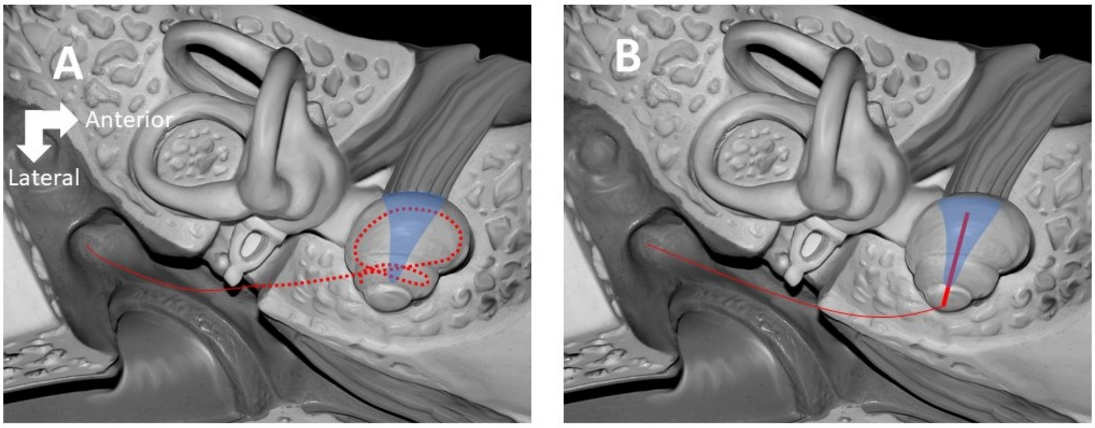


Figure 2.13: Comparison between a conventional (A) and transmodiolar (B) cochlear implant.

is that there is no visible anatomical landmark to indicate the cochlear apex and the axis of the modiolus. A part of this thesis focuses on proposing an AR-based solution for realizing a safe and reproducible transmodiolar implantation by augmenting precise location of the modiolar axis on the surgical view.

Chapter 3

Augmented Reality in Cranial-base Surgery

Ear anatomy is a part of the cranial base domain. Cranial base surgery is one of the surgical domains that has benefited a lot from the progress in imaging technology [49]. This region separates the brain and the posterior fossa (cerebellum and brainstem) from the mid facial region and the neck. It gives passage to a multitude of cranial nerves and vessels with vital functions. Operations in the cranial base region often require re-establishing aesthetics and functional anatomy by implants or by grafting, contouring or displacing skeletal elements [50].

AR is a promising technology that has been applied in various surgical domains. AR provides surgeons with additional cues by incorporating virtual information from pre and intraoperative imaging techniques to the real-time surgical environment. These additional cues can greatly enhance the surgical workflow in the form of improving localization and approach, treatment, efficiency and safety [1]. The potential for AR applications is immense in cranial base domain as the anatomy mainly consists of rigid bony structures yielding small deformations and allowing easy pre to intraoperative registration. However, only few works have been reported in this regard due to complicated workspace and high precision requirements. In this chapter, we present an overview of different works that have been reported pertaining to AR-based procedures in this domain. In this systematic review, we will first describe the contribution of AR to current neuro-navigation systems and its application to the minimally invasive cranial base surgery. Then, we will discuss various technical aspects underlying AR. Finally, we will discuss and confront different approaches to AR and future AR developments before it can be widely accepted in cranial base surgery.

Our analysis of the relevant literature has been published in the following article [1]:

"R. Hussain, A. Lalande, C. Guigou, A. Bozorg-Grayeli. Contribution of Augmented Reality to Minimally Invasive Computer-Assisted Cranial Base Surgery. *IEEE J Biomed Health Inform*, 24(7):2093-2106, 2020."

Discussion

In this article, we provided an overview of various AR-based surgical systems that have been proposed for different cranial base procedures. Researchers have employed AR in diagnosis, surgical planning, intraoperative visualization and guidance, navigation and post-operative analysis. The combination of AR with medical imaging can achieve precise anatomical localization and unification of pre and intraoperative data in an ergonomic and efficient manner. AR has been shown to not only facilitate surgical accuracy and help in decision making but also reduce operation time.

The use of AR has been mainly based on two premises: improving the outcomes of current surgical procedures and opening horizons for procedures that were not previously determined as feasible. In our work, we employed AR to help surgeons improve transtympanic procedures and study the possibility of transmodiolar cochlear implantation through additional visual cues. Based on this literature review and after consultation with different experts involved in the project, the following criteria were defined and considered during development of the AR systems:

- The system should exhibit millimetric precision which is essential for otological procedures.
- The system should be real-time with low latency.
- The system should provide visual cues in an ergonomic manner without interrupting the flow of the surgery.
- The system should not employ any specialised and dedicated hardware so that it would be easy to integrate into an OR.
- The calibration process of the setup should be minimal and easy to implement.
- The system should not use any bulky and expensive tracking systems allowing easy access for all ORs.
- The system should be able to manage most surgical manoeuvres without constraining the surgical workspace.
- The system should be robust to different acquisition systems available in different ORs all over the world.
- The system should encompass a precautionary feedback system for warning the surgeon if a system bug has been encountered or if undesired manipulations are being performed.

- The system should be able to cope with unexpected situations during the procedure such as change of approach, vibrations, etc.

Conventional AR systems employed optical and electromagnetic devices to establish correspondence between virtual and real information. Although, these systems provide high accuracy, they are bulky and expensive and require special markers and frames for tracking and an extensive calibration process before use. Alternatively, the premise of the work conducted in this dissertation is to develop AR solutions that employ vision-only based algorithms. This would help to improve the acceptability and ergonomics of AR systems since no specialised equipment is required, allowing simple installation into the OR setup.

Chapter 4

Augmented Reality for Transtympanic Middle Ear Surgery

This chapter presents an AR-based system for transtympanic middle ear surgical procedures based only on vision algorithms. A transtympanic surgical procedure aims at accessing middle ear cleft structures through a minimal opening in the tympanic membrane. The procedure is performed in a small and deep cavity involving fragile structures of the human body and requires high precision micro-surgical gestures. The outcomes of such surgeries can be potentially improved using computer vision and robotic-assistance systems.

Our goal is to implement a real-time AR-based system for robotic-assisted transtympanic surgery by combining information from a preoperative CT-scan image and real-time surgical video with the aim of visualizing the auditory ossciles and labyrinthine windows which are concealed by the opaque tympanic membrane (Fig. 4.1). Different possible applications of the proposed system include in-situ drug delivery and ossicular chain reconstruction. The system has been tested using endoscopes, laboratory microscopes and surgical microscopes on temporal bone resin phantoms, cadaver bones and during different middle ear procedures in the operating room. The subsequent sub-chapters provide details about the different processes undertaken during the development of this system.

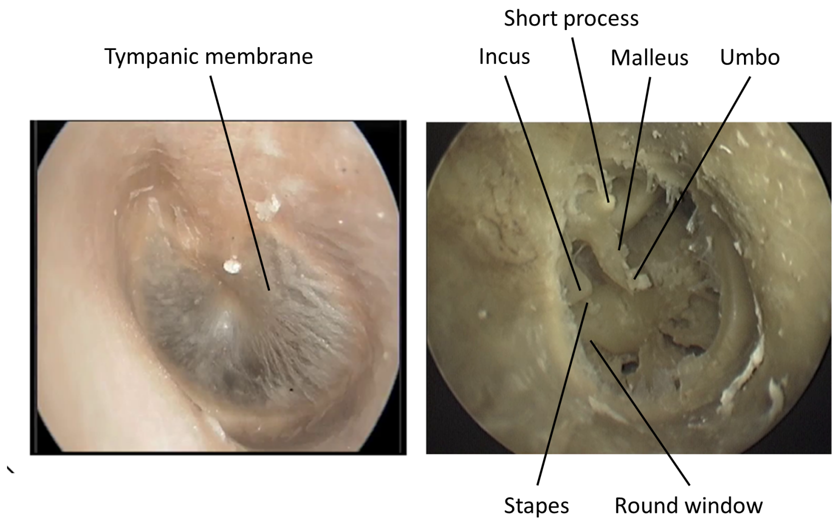


Figure 4.1: Endoscopic visualization through the auditory canal with and without the tympanic membrane.

4.1 Pilot Study: Comparison of Tracking Strategies

In this sub-chapter, we introduce the concept of AR-based transtympanic procedures and test our hypothesis in an elementary setting. This involved a manual registration between CT-scan image and pre-recorded endoscopic image followed by an analysis of different image-based tracking schemes for movement tracking. This sub-chapter has been published in the following article [5]:

"R. Marroquin, A. Lalande, R. Hussain, C. Guigou, A. Bozorg-Grayeli. Augmented Reality of the Middle Ear Combining Otoendoscopy and Temporal Bone Computed Tomography. *Otol Neurotol*, 39(8):931-939, 2018."

Discussion

In this work, we presented a proof of concept of an AR-based transtympanic surgical system. The study conducted on pre-recorded endoscopic videos of human cadaver bones yielded sub-millimetric accuracy providing motivation for further development of the system. The study yielded similar results between intact and elevated tympanic membrane. The experiments conducted without the tympanic membrane revealing an exact synchronization between the video and the CT endoscopic image of the visible ossicles further ascertained the practicality and usefulness of the system. Among the different tracking strategies employed, SURF based scheme was seen as the optimal candidate in terms of accuracy, robustness and latency. In the next step, we will extend this system by incorporating fiducial markers for automatizing the registration and an instrument identification and tracking scheme.

4.2 Vision-based AR System

In this sub-chapter, we present our proposed framework for AR-based transtympanic procedures. The framework was devised on the basis of the previous study. We analysed the possibility of a fiducial marker-based semi-automatic registration strategy to ease and speed up the registration process. A 5 degree of freedom (DoF) surgical instrument pose identification and tracking scheme was also considered in this study. This sub-chapter extends our previous study by validating the system's tracking scheme in near-realistic and challenging scenarios, and by imitating an actual procedure (drug administration). Contrary to the previous study, this study was performed in real-time. These developments brought the system several steps closer to its application in the operating room. This sub-chapter has been published in the following article [36]:

"R. Hussain, A. Lalande, R. Marroquin, C. Guigou, A. Bozorg-Grayeli. Video-based augmented reality combining CT-scan and instrument position data to microscope view in middle ear surgery. *Sci Rep*, 10:6767, 2020."

Discussion

In this work, we proposed an AR system for minimally invasive middle ear procedures via a transtympanic approach. The system employed fiducial markers attached at the periphery of the intact tympanic membrane for a point-based semi-automatic registration. An image feature-based tracking scheme was used for movement tracking and a three collinear point framework was used for surgical instrument identification and tracking behind the tympanic membrane. The study conducted on real-time microscopic videos of temporal bone resin phantom models also yielded sub-millimetric precision which is important in otological applications. Different target structures were successfully approached behind the tympanic membrane during an in-situ drug administration simulation. The tracking was unaffected by moderate instrument and patient movements and fluid leakages. Another study conducted for evaluation of the proposed system is presented in Appendix B with additional experimental results. In the next part, we will evaluate some features of the system in a clinical environment.

4.3 Clinical Evaluation

The next step of our work is the evaluation of the registration and movement tracking processes of the AR system in a clinical environment. The system was seamlessly deployed during different surgical procedures in the OR such as tympanoplasty and otosclerosis by simply connecting the output of the microscope video into the computer. For a better embedding, some additional features were also introduced to improve the ergonomics and safety of the system. These included a pre-registration step to help in selecting the registration points, automatic tracking recovery step to counter the effects of blurry and occluded video frames and an automatic re-registration step for efficient registration in case if the system fails due to extreme movements. This study provided important perspectives about transformation of a system from "bench-to-bedside". This work is under submission for IEEE J Biomed Health Inform.

Discussion

In this work, we evaluated the AR system in the OR on real-time surgical microscope videos of patients undergoing different middle and inner ear procedures. The system was employed with realistic time and ergonomic constraints in the OR and yielded sub-millimetric registration and tracking errors. The system was tolerant towards moderate hand and instrument movements, and blood and cerebrospinal fluid leakages. The main obstacles in smooth operation of the system were observed to be inconsistent reflections in the surgical scene and a rapid occlusion of a significant part of video frame by the surgeon's hands. The relative increase in surgical errors from laboratory to clinical experiments was synonymous with the performance of other AR systems [1]. This study provided a proof of concept of the use of the AR system in the OR and more cases should be included to provide a comprehensive analysis. The study also revealed insights into possible application of the AR system in other middle and inner ear procedures as well.

4.4 Benefits of Augmented Reality for Transtympanic Middle Ear Surgery

In this chapter, we presented a vision-based AR surgical system for transtympanic procedures. The system combines information from a preoperative CT-scan image with the real-time surgical video of the external auditory canal with the aim of visualizing middle ear cleft structures concealed behind the tympanic membrane. The system starts with a point-based initial microscope-CT registration followed by image feature-based patient-microscope movement tracking. The instrument behind the tympanic membrane was also identified and tracked using a three collinear marker framework.

Both anatomical and fiducial marker-based registrations were possible and both options are provided in the system. Similar registration errors were observed between the two types of registrations however, the time for corresponding points' selection was significantly lesser when markers were used. On the contrary, the use of markers is not simple since the markers need to be attached on the periphery of the tympanic membrane before the preoperative CT-scan acquisition which takes place some days before the surgical procedure in routine practice. It would be difficult for the markers to be immobile during this period. A patient-specific circular frame housing these markers could be a possible solution for this problem.

The system was tested using an oto-endoscope, a laboratory microscope and three surgical microscopes and yielded similar results on all platforms making it robust to image acquisition systems. The experiments conducted on temporal bone resin phantoms, human cadaver bones and patients also yielded near-similar results. Image feature-based tracking was observed to be sufficient for moderate movement tracking. Less image features could be extracted from the phantom models since their surface was smooth and uniformly coloured as compared to human samples. However, the extracted features were sufficient for robust movement tracking. A re-registration was required when extreme movements such as jerks were applied or the surgeon's hands were introduced rapidly into the video frame.

The relevant target structures were successfully approached behind the tympanic membrane using a tracked surgical instrument. The markers attached near the tip of the instrument have to be carefully positioned such that they are visible in the video frame when the instrument tip is behind the membrane. This information can be obtained using the distance between the membrane and the concerned target structure in preoprative CT-scan image. For good performance, it is important for the virtual endoscopic image to be taken at nearly the same viewpoint as the microscope orientation during the procedure.

The main advantage of the system is that no additional equipment (such as external trackers and calibration setup) is required enabling a smooth installation of the system in the OR. The system only requires an input from the microscope camera which is readily available in most surgical microscopes. This is crucial in acceptance of new technological systems into the OR. The system also did not interfere with the ergonomics of the surgery and the procedures could be carried out without any hindrance. However, when shifting from laboratory environment to an OR, a small decrease in performance was observed which is analogous to other similar studies [1].

In the next chapter, we will study the possibility of extending the vision-based AR system to transcanal inner ear procedures.

Chapter 5

Augmented Reality for Transmodiolar Implant Surgery

In the previous chapter, we presented an AR system for middle ear procedures. In this chapter, we will extend the AR system to inner ear procedures, particularly for cochlear implantation surgery.

Conventional cochlear implants are inserted into the scala tympani part of the cochlea through the round window or a cochleostomy. The point of entry is reached after drilling a bony cavity behind the auditory canal (mastoidectomy). This approach is highly invasive and the inserted electrode array is several millimetres away from the nerve endings leading to sub-optimal stimulation. An alternative approach is the transmodiolar auditory nerve insertion approach in which the electrode is implanted directly onto the auditory nerve through the external auditory canal. However, obtaining precise information about the auditory nerve is very difficult using traditional approaches as the cochlea is concealed behind the promontory temporal bone.

The aim of this work is to realize a safe and reproducible transmodiolar insertion using an AR framework by combining information from preoperative CT-scan images with the video feed from the surgical microscope. AR can be used to visualize the modiolar axis and provide navigational cues during insertion in an ergonomic manner. Contrary to the previous chapter where anatomical structures and surgical instruments were tracked and augmented on to surgical scene, only the cochlear modiolus and respective navigational information will be augmented in this study. In this regard, we proposed two different AR approaches: the first one was based on a stereoscopic-like camera system whereas the second one was similar to our transtympanic AR system using mono-vision. Both the systems were tested on temporal bone resin phantoms. This chapter presents our two approaches to transmodiolar implantation.

5.1 Stereo-like Framework

In this sub-chapter, we present an AR system for visualization of the cochlear modiolus from the external auditory canal using two microscope cameras placed on each eyepiece of a stereo-microscope. Firstly, we identify anatomical landmarks that can be used for registration between a preoperative CT-scan and surgical videos. A deep reinforcement learning inspired network was used to automatically extract information about different landmarks. Secondly, these landmarks were then used for registration between the CT-scan and stereo-microscopic camera images to visualize the modiolar axis on the surgical viewpoint in an AR setup. An image feature-based framework similar to the previous AR system was also employed to maintain correspondence between the two modalities. This sub-chapter has been published in the following article [51]:

R. Hussain, A. Lalande, K.B. Girum, C. Guigou, A. Bozorg-Grayeli. Augmented reality for inner ear procedures: visualization of the cochlear central axis in microscopic videos. *Int J Comput Assist Radiol Surg*, 15:1703-1711, 2020.

Discussion

In this work, we presented a two camera vision based AR system for modiolus visualization on surgical video. The approach employed an uncalibrated stereo-camera pair attached on each eyepiece of a surgical microscope. Seven anatomical landmarks (round window niche, incus tip, umbo, short process of malleus, tip of pyramid, cochlear apex, and cochlear base point) were automatically identified from a preoperative CT-scan image using deep Q network. Six of these landmarks were used to register the information from the CT-scan with the surgical videos using a point-based registration approach. The location of the modiolar axis was then obtained and overlaid onto the surgical scene using these registration projection matrices. An image feature-based tracking was also incorporated to manage patient-microscope movements. The system yielded millimetric difference compared to manual identification of anatomical points. The study provided an important first evaluation of the possibility of AR based transmodiolar cochlear implant surgery.

The stereo-like vision system, presented in this sub-chapter, is very sensitive to accurate registration point selection and it is difficult to obtain a good angle of approach such that all registration landmarks are simultaneously visible in both camera frames. To counter these drawbacks, we will present a mono-vision AR system for transmodiolar implantation in the next sub-chapter.

5.2 Mono-vision Framework

In this sub-chapter, we present an AR system for transmodiolar nerve implantation using a single camera video. This approach is an extension of the AR system for transtympanic procedures. Firstly, the modiolus is identified on the preoperative CT-scan image. A reconstructed CT-scan image of the middle and inner ear cleft structures was obtained and registered onto the surgical video using a point-based scheme. Patient-microscope movements were tracked using an image feature-based tracking scheme. Orthogonal visual cues were also projected onto the surgical scene for navigational guidance during insertion. The modiolus was drilled using the visual cues and the implant was inserted into the drilled cavity. Post-implantation CT-scans were used to analyze the performance of the system. This work is under submission for IEEE J Biomed Health Inform.

Discussion

In this work, we presented a mono-vision based AR system for auditory nerve implantation via a transmodiolar approach. A sham implant was inserted into the cochlear modiolus on temporal bone resin phantoms based on visualization and navigational cues about the modiolar axis on the real-time microscope video. Information from a preoperative CT-scan was relayed onto the surgical scene using a point-based registration and image feature tracking scheme. The orthogonal navigational cues were pertinent for transmodiolar implantation with millimetric accuracy. Post-implantation CT-scan imaging revealed that the implants had been inserted at the correct locations with small angular errors which were within the anatomical tolerance. The study yielded encouraging results for implantation using an AR approach. This is the first study on auditory nerve implantation on human samples without cochlear destruction.

5.3 Benefits of Augmented Reality for Transmodiolar Auditory Nerve Implantation

In this chapter, we presented two vision-based AR surgical systems for auditory nerve implantation using a transmodiolar approach. The systems combined information from a preoperative CT-scan image with the real-time surgical video of the external auditory canal with the aim of visualizing the cochlear modiolus concealed behind the promontory bone. The systems start with a point-based initial microscope-CT registration followed by image feature-based patient-microscope movement tracking during insertion.

The automatic landmark detection process based on deep Q learning strategy yielded an error around 2 mm which is large for millimetric otological manipulations. However, this error is inherent to CT acquisition parameters and non-conical nature of otological structures. Moreover, it is equivalent to only 3–4 voxels in clinical CT-scan images. The main motivation for using a reinforcement learning strategy was that it appeared to adapt to different experimental data such phantoms, cadavers and patients. An alternative semi-automatic CNN approach is presented in Appendix C which yielded better accuracy. However, it requires a manual region of interest extraction stage to obtain a cropped volume of middle and inner ear structures only. Higher resolution CT-scan acquisitions and networks trained on larger datasets might improve the landmark extraction accuracy.

The first system employed an uncalibrated stereo camera pair attached on each eyepiece of a stereo microscope. Although this system does not involve any calibration step, most surgical microscopes do not come with stereo cameras so some installation steps needed to be undertaken such as custom-designed frames to house the cameras on the eyepieces. On the contrary, the second system used a single camera only which is readily available on most surgical microscopes requiring no calibration or installation steps. Systems that do not require any calibration steps offer a huge advantage not only in terms of ergonomy but also because calibration errors are not applicable. An example of a setup requiring calibration steps is presented in Appendix D.

Since, the two systems are calibration-free, another advantage that the mono-vision approach offered over the stereo-like approach is in the selection of the registration points. In stereo-like setup, the 3D anatomical points have to be very carefully and precisely selected since it is more sensitive to selection errors. A small deviation might incur a large error in depicting the location of the modiolus. On the other hand, the mono-vision setup registers two images inhibiting a larger tolerance for registration point selection errors. However, the CT-scan reconstruction image needs to be captured at a similar viewpoint with the microscope video. Irrespective of the system, an automatic registration scheme would be highly beneficial. However, care would need to be taken in designing an automatic scheme since the structures in the ear are not conical and offer very less distinctive features making such a process highly prone to error.

In the stereo-like setup, it might be very difficult to obtain a good angle of approach such that all registration landmark points are visible in both camera frames at the same time. The external auditory canal will have to be enlarged in such cases which is a routine procedure practised during different surgical procedures. Nevertheless, during the implantation performed in the second study, it was observed that the anterior part of the auditory canal would also need to be enlarged for properly drilling the modiolus because of the inherent angle of insertion and bony characteristics of ear structures. Innovations in surgical tool design, such as compact angled drills, would be helpful in avoiding auditory canal enlargement.

This chapter provided important insights towards advancement in realizing a transmodiolar auditory nerve implantation. With current vision-based technology, millimetric precision is possible which might be improved by employing more advanced surgical systems and image acquisition tools. Since, the cochlea is concealed behind the promontory, there is no intra and post implantation visual feedback through the surgical scene. Considering CBCT-scans during implantation could also be a good possible future direction to provide a visual feedback and confirm if the implant is being inserted correctly.

In some cases, information about the modiolus may not be sufficient and visualization of the entire inner ear anatomy might be needed for improved localization. Therefore, in the next chapter, we will propose a technique for automatic segmentation of inner ear structures from CT-scan images that may be used in an AR setup.

Chapter 6

Inner Ear Segmentation

During conventional cochlear implantation, the information about the anatomical variations of the cochlea (e.g. malformations, lumen obstruction or narrowing) or its size influences the choice of the array [52]. Clinical CT-scans are acquired at relatively low resolutions which is sufficient for inter-structural details but does not provide intra-structural details. μCT images offer a significant improvement in terms of resolution over clinical CT-scan images providing sub-structural information about the cochlea, semi-circular canals and vestibule to the surgeons. However, obtaining this information through manual segmentation requires a significant amount of time and effort due to the presence of a large number of image slices.

This chapter presents a CNN based technique for automatic inner ear segmentation from μCT images. We propose an auto-context inspired cascaded CNN framework for fast and automatic inner ear segmentation. This chapter is under review in the Sci Rep journal.

Discussion

In this work, we proposed a fully automatic CNN framework for segmenting inner ear structures from μCT images. The proposed approach employed an auto-context based cascaded 2D U-net architecture with 3D connected component refinement to segment the cochlear scalae, semicircular canals and the vestibule. The system yielded precise 3D segmentations with an average Dice coefficient of 0.90 and Hausdorff distance of 0.74 mm, despite the limited size of the dataset. Different variants of the proposed AutoCasNet framework outperformed their corresponding state of the art algorithms (whose best performance in respect to slice orientation was listed in the results section). The framework's segmentation time was fairly low which is motivating to apply it on larger dataset and use it in clinical practice. It is important to mention that this study provided a proof of concept for such a system and it is important for the framework to be trained and tested on a larger dataset before being employed in routine practice.

In future, the framework can be used to build a detailed robust inner ear model using statistical shape modelling from high number of μCT segmentations. Statistical shape modelling algorithms are often used to represent anatomical variations of concerned structures in a compact parametric model by first registering segmentations obtained from different patients and generating a mesh structure through deformation field generation. This detailed statistical model can then be registered with the segmentation from patient's clinical CT-scan using automatically extracted anatomical points or intensity-based algorithms [12, 51]. The resulting high-quality co-registered data set of the human bony labyrinth can be used to study microscopic inner ear morphology in detail, for developing efficient design of neuroprostheses and for surgical planning during minimally invasive treatment [12, 13, 53]. In addition to providing crucial details about the ear structures, the model can also be used for virtual augmentation in AR surgical systems.

Chapter 7

Conclusion & Perspectives

This project aimed at developing video-based AR solutions for middle and inner ear surgical procedures. The systems were designed such that they do not require any calibration or installation steps and can be integrated into any OR.

The first application that was addressed in this dissertation was transtympanic procedures that aim at accessing middle ear cleft structures through a minimal opening in the tympanic membrane [36]. The procedure is performed in a small and deep cavity involving fragile structures of the human body and requires high precision micro-surgical gestures. We developed an AR system combining preoperative CT-scan based virtual endoscopy image and the real-time microscopic video with the aim of visualizing middle ear structures behind the tympanic membrane. A tracked surgical instrument may be passed through a small puncture in the tympanic membrane for drug administration, labyrinthine fistula diagnosis and ossicular chain repair. The system was evaluated on temporal bone resin phantoms, human cadaver bones and during middle ear surgeries in operating room environment. The system yielded sub-millimetric accuracy which is compatible with otological surgery.

As an extension of this system, we also proposed an AR system for an inner ear procedure [51]. Conventional cochlear implants are inserted into the tympanic ramp (scala tympani) of the cochlea through the round window or a cochleostomy. With this approach, the electrode array is several millimetres away from the nerve endings making the electrical stimulation less specific and less efficient. Transmodiolar approach provides a good alternative for cochlear implantation by providing a direct access to the auditory nerve in the modiolus at the center of the cochlea. We developed an AR system that combined information from a preoperative CT-scan image with the real-time video from the surgical microscope with the aim of providing navigational cues during implantation. The system tested on temporal bone phantoms showed promising results with millimetric accuracy and improved ergonomy compared to conventional neuronavigation systems.

In some cases, information about the modiolus may not be sufficient. Therefore, we proposed a technique for automatic segmentation of inner ear structures from CT-scan images. Clinical CT-scans are acquired at relatively low resolutions, and improved performance can be achieved by registering patient specific CT-scan to a high-resolution inner ear model built from accurate 3D segmentations based on μCT of human temporal bone specimens. In this regard, we proposed a framework based on CNN for human inner ear segmentation from μCT images which can be used to build such a model from an extensive database. The proposed approach employed an auto-context based cascaded 2D U-net architecture with 3D connected component refinement to segment the cochlear scalae, semicircular canals and the vestibule. This system was formulated on a public data set composed of 17 μCT-scans. The system yielded precise and fast automatic inner-ear segmentations with a processing time of 11 seconds per 3D μCT image, an average Dice coefficient of 0.90, and Hausdorff distance of 0.74 mm.

The developed concepts in this work open important perspectives into minimally invasive otological procedures through AR-based solutions. The following future works are envisioned for the continuation of the work conducted under this dissertation:

- One of the main limitations of the proposed AR systems that was observed during experiments in the OR, was the registration process. Currently, manual and semi-automatic point-based approaches have been integrated into the system. More work should be undertaken to further automatize this step. A patient-specific custom ring designed to house the fiducial markers would be very useful so that the markers can be removed after the CT-scan acquisition and put back in the same place before the surgical procedure. It might also be useful to conduct topological feature analysis or employ structure from motion techniques, persistent homology, or deep learning strategies to realize fully automatic microscope-CT registrations. For deep learning strategies, an extensive dataset would need to be built. Automatic registration process would be beneficial in terms of both ergonomics and registration time.

- For the development of AR systems, image feature tracking schemes were proposed in a cumulative framework. Although these techniques were observed to exhibit good performance during the experiments, they suffer from minimal system drift errors. These drifts are tolerable for the currently envisioned applications such as drug administration and electrode insertion since they can be performed relatively quickly. For these systems to be applied on longer procedures, bundle adjustment techniques might need to be integrated into the tracking schemes to avoid system drifts.

- Currently, only visual feedback is available to the surgeon for inferring whether the system has failed. While this is sufficient for significant system drifts, millimetric errors can not be reliably judged solely on visual feedback. A precautionary feedback system should be designed to alarm the user if the synchronization/tracking error exceeds a specific threshold.

- The AR system has been validated on resin phantoms, cadaver bones and small number of patients undergoing ear surgeries. An extensive study should be conducted on a large number of patients and procedures. This would increase the confidence and reveal limitations of the system for use in the OR environment. It would also be interesting to quantify the errors as a function of different visual challenges such as fluid leakages, patient-microscope movements, system jerks, hand obstruction, reflectance etc. This would also define the next developments that need to be undertaken for the system to be usable in any given scenario.

- In this work, we proposed a framework for inner ear segmentation to build anatomical models for AR applications. The framework should be extended to segment other structures in the ear as well which would be helpful during a higher range of surgical applications.

- Integration of a robotic manipulator would improve the surgical manoeuvres during transtympanic procedures, since the pivot of motion can be adjusted to be exactly at the point of entry in tympanic membrane. This would avoid tympanic rupture during insertion of the surgical instrument and during movements when it is behind the tympanic membrane. RobOtol (Collin SAS, Paris, France) might be a possible candidate for the AR system [23].

- A virtual reality based patient-specific simulator could be designed to simulate the surgical process for training and pre-planning purposes.

- Finally, it would be interesting to compare the proposed vision-based AR guidance systems with conventional neuronavigation systems and other AR solutions employing traditional external tracking devices to study the drawbacks and benefits of each type of system in terms of ergonomy, accuracy, mental workload, repeatability etc. Different types of approaches are being studied for auditory nerve implantation at our laboratory and a comparative study would be highly insightful.

Appendix A

Useful Length and Diameter of Cochlear Lumen for Cochlear Implantation

During conventional cochlear implantation, the insertion of the electrode array into the scala tympani can be traumatic for the spiral ligament, the basilar membrane, among others. Variability in cochlear size (length and diameter) is an important factor that needs to be considered before the insertion. These specifications can be obtained by unfolding the cochlear lumen using curvillinear MPR reconstructions. In this chapter, we studied the inter-individual variability in cochlear shape to determine the useful length of the scala tympani and the diameter of the cochlear lumen using routine preoperative CT-scan imaging before cochlear implantation.

This chapter has been accepted in the form of the following article:

"C. Guigou, A. Schein, P. Trouilloud, A. Lalande, R. Hussain, A. Bozorg-Grayeli. Curvilinear Multiplanar Reconstruction to Predict Useful Length and Diameter of Cochlear Lumen for Cochlear Implantation. *Otol Neurotol*, 41(10):e1207-e1213, 2020."

Appendix B

Additional Results for Transtympanic Middle Ear Surgery

This chapter presents some additional experiments that were performed to further validate the AR-based system for transtympanic surgery to visualize the auditory ossicles behind the intact tympanic membrane. The system under evaluation consisted of a fiducial marker-based microscope-CT registration followed by simultaneous image feature-based patient-microscope movement tracking and surgical instrument pose estimation. The real-time experiments conducted in this study included an analysis on initial registration, and microscope movement and instrument tracking using temporal bone resin phantoms and a laboratory microscope.

These additional results were presented at the MICCAI conference [3]:

"R. Hussain, A. Lalande, R. Marroquin, K.B. Girum, C. Guigou, A. Bozorg-Grayeli. Real-Time Augmented Reality for Ear Surgery. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 324-31, Granada, Spain, 2018."

Appendix C

Semi-automatic Landmark Detection for Otological Procedures

In most AR systems, particularly in the cranial base domain, point based pre to intraoperative image registration is still one of the most popular approaches [1]. However, the ear does not contain sufficient number of distinct anatomical structures usable for automatic registration of a preoperative CT-scan with a surgical video. This enforces manual annotation of anatomical landmarks to be prone to error and not trivial. In this chapter, we propose a semi-automatic CNN-based 3D landmark detection framework that can be applied on preoperative head CT-scans to extract information about round window niche, pyramid, incus tip, umbo, short process of malleus and cochlear modiolus. These landmarks are visible in endoscopic or microscopic surgical videos obtained via the external auditory canal and can be used for registration between a preoperative CT-scan image and a corresponding surgical video.

This work was presented as a long abstract in Surgetica conference [54]:

"R. Hussain, C. Guigou, K.B. Girum, A. Lalande, A. Bozorg-Grayeli. 3D landmark detection for augmented reality based otologic procedures. *Surgetica Conference*, Rennes, France, 2019."

Appendix D

Stereo-calibration between Non-overlapping Heterogeneous Camera Systems

Calibration is the first crucial process in most image guided surgical systems. Traditional surgical systems employ bulky and expensive optical or electromagnetic tracking systems to establish real-time correspondence between patient and surgical instruments. Alternatively, some systems may only use vision-based tracking systems which typically employ externally placed cameras to determine and track poses of different entities in the surgical setup. In this chapter, we propose a technique for stereo-calibration between a surgical microscope camera and an external tracking camera rigidly attached to the microscope. This method may be applied on stereo systems with both partial and non-overlapping field of views.

This work was presented as a long abstract in CARS conference:

"R. Hussain, A. Lalande, I. Stefanis, K.B. Girum, C. Guigou, D. Fofi, A. Bozorg-Grayeli. Stereo calibration of non-overlapping field of view heterogeneous cameras for calibrating surgical microscope with external tracking camera. *CARS Conference*, S187-S188, Munich, Germany, 2020."

Appendix E

List of Publications

E.1 Journals

- **R. Hussain**, A. Lalande, K.B. Girum, C. Guigou, A. Bozorg-Grayeli. Augmented reality for inner ear procedures: visualization of the cochlear central axis in microscopic videos. *Int J Comput Assist Radiol Surg*, 15:1703-1711, 2020. DOI: `https://doi.org/10.1007/s11548-020-02240-w`. [Q1(SJR: 0.68); IF: 2.47]

- **R. Hussain**, A. Lalande, C. Guigou, A. Bozorg-Grayeli. Contribution of Augmented Reality to Minimally Invasive Computer-Assisted Cranial Base Surgery. *IEEE J Biomed Health Inform*, 24(7):2093-2106, 2020. DOI: `https://doi.org/10.1109/JBHI.2019.2954003`. [Q1(SJR: 1.31); IF: 5.22]

- **R. Hussain**, A. Lalande, R. Marroquin, C. Guigou, A. Bozorg-Grayeli. Video-based augmented reality combining ct-scan and instrument position data to microscope view in middle ear surgery. *Sci Rep*, 10:6767, 2020. DOI: `https://doi.org/10.1038/s41598-020-63839-2`. [Q1(SJR: 1.34); IF: 3.99]

- **R. Hussain**, A. Lalande, K.B. Girum, C. Guigou, A. Bozorg-Grayeli. Automatic Segmentation of Inner Ear on CT-Scan Using Auto-Context Convolutional Neural Network. *Sci Rep*, 2020 (under review).

- **R. Hussain**, C. Guigou, A. Lalande, A. Bozorg-Grayeli. Vision based Augmented Reality System for Middle Ear Surgery: Evaluation in Operating Room Environment. *IEEE J Biomed Health Inform*, 2020 (submission under process).

- C. Guigou, A. Schein, P. Trouilloud, A. Lalande, **R. Hussain**, A. Bozorg-Grayeli. Curvilinear Multiplanar Reconstruction to Predict Useful Length and Diameter of Cochlear

Lumen for Cochlear Implantation. *Otol Neurotol*, 41(10):e1207-e1213, 2020. DOI: `https://doi.org/10.1097/MAO.0000000000002829` [Q1(SJR: 1.17); IF: 1.71]

- K.B. Girum, A. Lalande, **R. Hussain**, G. Crehange. A deep learning method for real-time intraoperative US image segmentation in prostate brachytherapy. *Int J Comput Assist Radiol Surg*, 15:14671476, 2020. DOI: `https://doi.org/10.1007/s11548-020-02231-x`. [Q1(SJR: 0.68); IF: 2.47]

- K.B. Girum, G. Crehange, **R. Hussain**, A. Lalande. Fast interactive medical image segmentation with weakly-supervised deep learning method. *Int J Comput Assist Radiol Surg*, 15:14371444, 2020. DOI: `https://doi.org/10.1007/s11548-020-02223-x`. [Q1(SJR: 0.68); IF: 2.47]

- R. Marroquin, A. Lalande, **R. Hussain**, C. Guigou, A. Bozorg-Grayeli. Augmented Reality of the Middle Ear Combining Otoendoscopy and Temporal Bone Computed Tomography. *Otol Neurotol*, 39(8):931-939, 2018. DOI: `https://doi.org/10.1097/MAO.0000000000001922`. [Q1(SJR: 1.17); IF: 1.71]

- C. Guigou, **R. Hussain**, A. Lalande, A. Bozorg-Grayeli. Augmented Reality based Transmodiolar Cochlear Implantation. *IEEE J Biomed Health Inform*, 2020 (submission under process).

E.2 Conference Proceedings

- **R. Hussain**, A. Lalande, R. Marroquin, K.B. Girum, C. Guigou, A. Bozorg-Grayeli. Real-Time Augmented Reality for Ear Surgery. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 324-31, Granada, Spain, 2018. DOI: `https://doi.org/10.1007/978-3-030-00937-3_38`.

- K.B. Girum, G. Crehange, **R. Hussain**, P.M. Walker, A. Lalande. Deep generative model-driven multimodal prostate segmentation in radiotherapy. *Artificial Intelligence in Radiation Therapy Workshop* held in conjunction with *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 119-127, Shenzen, China, 2019. DOI: `https://doi.org/10.1007/978-3-030-32486-5_15`.

E.3 Conference Abstracts / Short Papers

- **R. Hussain**, A. Lalande, I. Stefanis, K.B. Girum, C. Guigou, D. Fofi, A. Bozorg-Grayeli. Stereo calibration of non-overlapping field of view heterogeneous cameras for calibrating surgical microscope with external tracking camera. *CARS Conference*, S187-S188, Munich, Germany, 2020.

- **R. Hussain**, A. Lalande, R. Marroquin, C. Guigou, A. Bozorg-Grayeli. Navigation par la réalité augmentée combinant le scanner et la position des instruments avec la vue microscopique pour la chirurgie de l'oreille moyenne. *SFORL Congress*, 00645, Paris, France, 2019.

- **R. Hussain**, A. Lalande, K.B. Girum, C. Guigou, A. Bozorg-Grayeli. Segmentation automatique de l'oreille interne á partir des données tomodensitométriques par l'apprentissage profond. *SFORL Congress*, 00669, Paris, France, 2019.

- **R. Hussain**, C. Guigou, K.B. Girum, A. Lalande, A. Bozorg-Grayeli. 3D landmark detection for augmented reality based otologic procedures. *Surgetica Conference*, Rennes, France, 2019.

- C. Guigou, **R. Hussain**, A. Lalande, A. Bozorg-Grayeli. Réalité augmentée au bloc opératoire pour la chirurgie de l'oreille moyenne. *SFORL Congress*, Paris, France, 2020.

E.4 Challenges

- Winner of category 2 (Train on VR, Test on Clinical-like). R. Hussain, **R. Hussain**, Y. Skandarani, K.B. Girum. VGG-3DConvNet for surgical action recognition. Surgical Visual Domain Adaptation (MICCAI-EndoVis-SurgVisDom challenge), 2020.

- 2^{nd} Runner-up of classification contest. K.B. Girum, Y. Skandarani, **R. Hussain**, A. Bozorg Grayeli, G. Crehange, A. Lalande. Automatic myocardial infarction evaluation from DE-MRI using deep convolutional networks. automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI (MICCAI-STACOM-EMIDEC challenge), 2020.

- Top 10 challengers. Y. Skandarani, **R. Hussain**, K.B. Girum, A. Lalande. On the effectiveness of U-Net for coronary artery segmentation. Automated Segmentation of Coronary Arteries (MICCAI-ASOCA challenge), 2020.

Bibliography

[1] R Hussain, A Lalande, C Guigou, and A Bozorg Grayeli. Contribution of augmented reality to minimally invasive computer-assisted cranial base surgery. *IEEE Journal of Biomedical and Health Informatics*, 2019.

[2] MJ Citardi, A Agbetoba, JL Bigcas, and A Luong. Augmented reality for endoscopic sinus surgery with surgical navigation: a cadaver study. In *International Forum of Allergy & Rhinology*, volume 6, pages 523–528. Wiley Online Library, 2016.

[3] R Hussain, A Lalande, R Marroquin, KB Girum, C Guigou, and A Bozorg Grayeli. Real-time augmented reality for ear surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 324–331. Springer, 2018.

[4] YP Murugesan, A Alsadoon, P Manoranjan, and PWC Prasad. A novel rotational matrix and translation vector algorithm: geometric accuracy for augmented reality in oral and maxillofacial surgeries. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 14(3):e1889, 2018.

[5] R Marroquin, A Lalande, R Hussain, C Guigou, and A Bozorg Grayeli. Augmented reality of the middle ear combining otoendoscopy and temporal bone computed tomography. *Otology & Neurotology*, 39(8):931–939, 2018.

[6] J Wang, H Suenaga, K Hoshi, L Yang, E Kobayashi, I Sakuma, and H Liao. Augmented reality navigation with automatic marker-free image registration using 3-d image overlay for dental surgery. *IEEE Transactions on Biomedical Engineering*, 61(4):1295–1304, 2014.

[7] S Bernhardt, SA Nicolau, L Soler, and C Doignon. The status of augmented reality in laparoscopic surgery as of 2016. *Medical Image Analysis*, 37:66–90, 2017.

[8] P Vávra, J Roman, P Zonča, P Ihnát, M Němec, J Kumar, N Habib, and A El-Gendi. Recent development of augmented reality in surgery: a review. *Journal of Healthcare Engineering*, 2017, 2017.

[9] JW Yoon, RE Chen, EJ Kim, OO Akinduro, P Kerezoudis, PK Han, P Si, WD Freeman, RJ Diaz, RJ Komotar, et al. Augmented reality for the surgeon: systematic review. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 14(4):e1914, 2018.

[10] S Standring, H Ellis, J Healy, D Johnson, A Williams, P Collins, and C Wigley. Gray's anatomy: the anatomical basis of clinical practice. *American Journal of Neuroradiology*, 26(10):2703, 2005.

[11] C Guigou, A Schein, P Trouilloud, A Lalande, R Hussain, and A Bozorg Grayeli. Curvilinear multiplanar reconstruction to predict useful length and diameter of cochlear lumen for cochlear implantation. *Otology & Neurotology*, In Press, 2020.

[12] W Wimmer, L Anschuetz, S Weder, F Wagner, H Delingette, and M Caversaccio. Human bony labyrinth dataset: Co-registered ct and micro-ct images, surface models and anatomical landmarks. *Data in Brief*, 27:104782, 2019.

[13] N Gerber, M Reyes, L Barazzetti, HM Kjer, S Vera, M Stauber, P Mistrik, M Ceresa, N Mangado, W Wimmer, et al. A multiscale imaging and modelling dataset of the human inner ear. *Scientific Data*, 4:170132, 2017.

[14] SH Bartling, K Peldschus, T Rodt, F Kral, H Matthies, R Kikinis, and H Becker. Registration and fusion of ct and mri of the temporal bone. *Journal of Computer Assisted Tomography*, 29(3):305–310, 2005.

[15] JI Lane, RJ Witte, OW Henson, CLW Driscoll, J Camp, and RA Robb. Imaging microscopy of the middle and inner ear: Part ii: Mr microscopy. *Clinical Anatomy: The Official Journal of the American Association of Clinical Anatomists and the British Association of Clinical Anatomists*, 18(6):409–415, 2005.

[16] L Shi, D Wang, WCW Chu, GR Burwell, TT Wong, PA Heng, and JCY Cheng. Automatic mri segmentation and morphoanatomy analysis of the vestibular system in adolescent idiopathic scoliosis. *Neuroimage*, 54:S180–S188, 2011.

[17] JH Bong, HJ Song, Y Oh, N Park, H Kim, and S Park. Endoscopic navigation system with extended field of view using augmented reality technology. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 14(2):e1886, 2018.

[18] M Aref, A Martyniuk, S Nath, A Koziarz, J Badhiwala, A Algird, F Farrokhyar, SA Almenawer, and K Reddy. Endoscopic third ventriculostomy: outcome analysis of an anterior entry point. *World Neurosurgery*, 104:554–559, 2017.

[19] Y Chu, J Yang, S Ma, D Ai, W Li, H Song, L Li, D Chen, L Chen, and Y Wang. Registration and fusion quantification of augmented reality based nasal endoscopic surgery. *Medical Image Analysis*, 42:241–256, 2017.

[20] CT Tang, K Kurozumi, P Pillai, V Filipce, EA Chiocca, and M Ammirati. Quantitative analysis of surgical exposure and maneuverability associated with the endoscope and the microscope in the retrosigmoid and various posterior petrosectomy approaches to the petroclival region using computer tomograpy-based frameless stereotaxy. a cadaveric study. *Clinical Neurology and Neurosurgery*, 115(7):1058–1062, 2013.

[21] T Brinker, G Arango, J Kaminsky, A Samii, U Thorns, P Vorkapic, and M Samii. An experimental approach to image guided skull base surgery employing a microscope-based neuronavigation system. *Acta Neurochirurgica*, 140(9):883–889, 1998.

[22] X Wang, L Li, Y Wang, J Hu, J Zhou, Z Jing, and A Wu. Clinical application of multimodal neuronavigation system in neuroendoscope-assisted skull base chordoma resection. *Journal of Craniofacial Surgery*, 28(6):e554–e557, 2017.

[23] M Miroir, Y Nguyen, J Szewczyk, O Sterkers, and A Bozorg Grayeli. Design, kinematic optimization, and evaluation of a teleoperated system for middle ear microsurgery. *The Scientific World Journal*, 2012, 2012.

[24] M Miroir, J Szewczyk, Y Nguyen, S Mazalaigue, and O Sterkers. Design of a robotic system for minimally invasive surgery of the middle ear. In *2008 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 747–752. IEEE, 2008.

[25] J Pile and N Simaan. Modeling, design, and evaluation of a parallel robot for cochlear implant surgery. *IEEE/ASME Transactions on Mechatronics*, 19(6):1746–1755, 2014.

[26] M Caversaccio, F Langlotz, LP Nolte, and R Häusler. Impact of a self-developed planning and self-constructed navigation system on skull base surgery: 10 years experience. *Acta Oto-Laryngologica*, 127(4):403–407, 2007.

[27] H Essig, M Rana, A Meyer, AM Eckardt, H Kokemueller, C von See, D Lindhorst, F Tavassol, M Ruecker, and NC Gellrich. Virtual 3d tumor marking-exact intraoperative coordinate mapping improve post-operative radiotherapy. *Radiation Oncology*, 6(1):159, 2011.

[28] DS Poe. Transtympanic endoscopy of the middle ear. *Operative Techniques in Otolaryngology-Head and Neck Surgery*, 3(4):239–244, 1992.

[29] S Kakehata. Transtympanic endoscopy for diagnosis of middle ear pathology. *Otolaryngologic Clinics of North America*, 46(2):227–232, 2013.

[30] SK Plontke, PK Plinkert, B Plinkert, A Koitschev, HP Zenner, and H Löwenheim. Transtympanic endoscopy for drug delivery to the inner ear using a new microendoscope. *Advances in Oto-Rhino-Laryngology*, 59:149–155, 2002.

[31] ZA Mood and SJ Daniel. Use of a microendoscope for transtympanic drug delivery to the round window membrane in chinchillas. *Otology & Neurotology*, 33(8):1292–1296, 2012.

[32] Z Selmani, I Pyykkö, H Ishizaki, and TI Marttila. Role of transtympanic endoscopy of the middle ear in the diagnosis of perilymphatic fistula in patients with sensorineural hearing loss or vertigo. *ORL*, 64(5):301–306, 2002.

[33] P Montandon, P Guillemin, and R Häusler. Prevention of vertigo in menieres syndrome by means of transtympanic ventilation tubes. *ORL*, 50(6):377–381, 1988.

[34] X Ge and JJ Shea Jr. Transtympanic electrocochleography: a 10-year experience. *Otology & Neurotology*, 23(5):799–805, 2002.

[35] P Daponte, L De Vito, F Picariello, and M Riccio. State of the art and future developments of the augmented reality for measurement applications. *Measurement*, 57:53–70, 2014.

[36] R Hussain, A Lalande, R Marroquin, C Guigou, and A Bozorg Grayeli. Video-based augmented reality combining ct-scan and instrument position data to microscope view in middle ear surgery. *Scientific Reports*, 10(1):1–11, 2020.

[37] S Weber, K Gavaghan, W Wimmer, T Williamson, N Gerber, J Anso, B Bell, A Feldmann, C Rathgeb, M Matulic, et al. Instrument flight to the inner ear. *Science Robotics*, 2(4), 2017.

[38] JH Doyle, JB Doyle, and FM Turnbull. Electrical stimulation of eighth cranial nerve. *Archives of Otolaryngology*, 80(4):388–391, 1964.

[39] AG Micco and CP Richter. Tissue resistivities determine the current flow in the cochlea. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 14(5):352–355, 2006.

[40] BJ Gantz, BF Mccabe, and RS Tyler. Use of multichannel cochlear implants in obstructed and obliterated cochleas. *OtolaryngologyHead and Neck Surgery*, 98(1):72–81, 1988.

[41] T Balkany, M Luntz, FF Telischi, and AV Hodges. Intact canal wall drill-out procedure for implantation of the totally ossified cochlea. *The American Journal of Otology*, 18(6 Suppl):S58–9, 1997.

[42] NL Cohen and SB Waltzman. Partial insertion of the nucleus multichannel cochlear implant: technique and results. *Otology & Neurotology*, 14(4):357–361, 1993.

[43] T Lenarz, A Lesinski-Schiedat, BP Weber, PR Issing, C Frohne, A Büchner, RD Battmer, JV Parker, and EV Wallenberg. The nucleus double array cochlear implant: a new concept for the obliterated cochlea. *Otology & Neurotology*, 22(1):24–32, 2001.

[44] K Assenova, S Stoyanov, and T Karchev. Anatomic considerations for electrode implantation into the modiolus. *The Mediterranean Journal of Otology*, 3:133–139, 2007.

[45] JC Middlebrooks and RL Snyder. A wider range of sounds for the deaf. In *EurekAlert*. Future Drugs Ltd, 2007.

[46] C Guigou, G Leterme, B Pasquis, L Martin, G Tourrel, L Bretillon, and A Bozorg Grayeli. Implantation of the auditory nerve via the middle ear cavity in rats with partial hearing preservation. *Audiology and Neurotology*, 21(2):98–104, 2016.

[47] FB Simmons, JM Epley, RC Lummis, N Guttman, LS Frishkopf, LD Harmon, and E Zwicker. Auditory nerve: electrical stimulation in man. *Science*, 148(3666):104–106, 1965.

[48] WFS Afifi, C Guigou, S Mazalaigue, JP Camuset, F Ricolfi, and A Bozorg Grayeli. Navigation-guided transmodiolar approach for auditory nerve implantation via the middle ear in humans. *Audiology and Neurotology*, 20(2):128–135, 2015.

[49] HR Kelly and HD Curtin. Imaging of skull base lesions. In *Handbook of Clinical Neurology*, volume 135, pages 637–657. Elsevier, 2016.

[50] G Widmann. Image-guided surgery and medical robotics in the cranial area. *Biomedical Imaging and Intervention Journal*, 3(1), 2007.

[51] R Hussain, A Lalande, KB Girum, C Guigou, and A Bozorg Grayeli. Augmented reality for inner ear procedures: visualization of the cochlear central axis in microscopic videos. *International Journal of Computer Assisted Radiology and Surgery*, 15:1703–1711, 2020.

[52] TH Vu, C Perazzini, M Puechmaille, A Bachy, A Mulliez, L Boyer, T Mom, and J Gabrillargues. Ct-scan contouring technique allows for direct and reliable measurements of the cochlear duct length: implication in cochlear implantation with straight electrode-arrays. *European Archives of Oto-Rhino-Laryngology*, 276(8):2135–2140, 2019.

[53] D Sieber, P Erfurt, S John, GR Dos Santos, D Schurzig, M Sørensen, and T Lenarz. The openear library of 3d models of the human temporal bone based on computed tomography and micro-slicing. *Scientific Data*, 6:180297, 2019.

[54] R Hussain, A Lalande, KB Girum, C Guigou, and A Bozorg Grayeli. 3d landmark detection for augmented reality based otologic procedures. In *Surgetica*, 2019.