Modelling of the Human Inner Ear Anatomy and Variability for Cochlear Implant Applications

Kjer, Hans Martin

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Modelling of the Human Inner Ear Anatomy and Variability for Cochlear Implant Applications

Hans Martin Kjer

Kongens Lyngby 2015

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

This thesis takes the biomedical engineering approach to working with and understanding the anatomy and physiology of the inner ear. The purpose is to apply the acquired knowledge in the development of implantable hearing aids.

The so-called Cochlear Implant (CI) is a fascinating technology that without underselling it provides hearing for the deaf. The technology faces a number of challenges, and a part of the solution to those is closely connected with an improved understanding of the inner ear anatomy, both with regards to the individual patient but also to the variation in the population.

The inner ear is a relatively small structure and even with modern medical scanners only the coarsest details are revealed about the specific patient anatomy. To study the anatomy it is required to work on specimens from deceased subjects scanned with for instance μCT. The anatomy is complex and presents several challenges concerning data processing and analysis.

Our approach is to describe the inner ear as a statistical shape model. The thesis covers our work with regards to data segmentation, shape characterization, development of image registration model suited for the inner ear and construction of statistical deformation models.

The thesis results in a series of applications relating to CIs. The shape model can be used by CI-manufacturers for virtual product development and testing. At the same time, it can be applied to estimate the detailed inner ear shape from a clinical patient CT scan. This opens up for tools to optimize the programming of the CI, such that the hearing restoration is improved.

Summary (Danish)

Denne afhandling tager mediko-ingeniørens tilgang til at arbejde med og forstå anatomien og fysiologien af det indre øre. Formålet er at den tilegnede viden kan benyttes i udviklingen af bedre implanterbare høreapparater.

De såkaldte Cochlear Implants (CI) er en fascinerende teknologi der uden at underdrive giver døve deres hørelse tilbage. Teknologien har dog en række udfordringer, og en del af løsningen af disse hænger nøje sammen med en bedre forståelse det indre øres anatomi, både med hensyn til den enkelte patient men også variationen i befolkningen.

Det indre øre er en forholdsvis lille struktur, og selv med nyere medicinske scannere kan man kun se de groveste detaljer for den enkelte patient. Hvis man vil studere anatomien er man derfor nød til at arbejde med eksemplarer udtrukket fra afdøde personer og skannet med f.eks. μCT. Anatomien er kompleks og stiller adskillige udfordringer i forhold til data processering og analyse.

Vores tilgang i denne afhandling er at beskrive det indre øre ved en statistisk form-model. Afhandlingen dækker vores arbejde med data segmentering, form karakterisering, udvikling af billede-registrerings-modeller egnet til det indre øre samt konstruktion af statistiske deformations-modeller.

Afhandlingen munder ud i række anvendelsesmuligheder relateret til CIs. Formmodellen kan benyttes af CI-producenterne til virtuel produkt udvikling. Samtidig kan den benyttes i forbindelse med at estimere den detaljerede indre øre form fra en patient CT skanning. Dette åbner op for værktøjer til at optimere CI-programmeringen, således at hørelsen kan genskabes bedre.

Preface

The work presented here was carried out primarily at the Section for Image Analysis and Computer Graphics at the Department for Applied Mathematics and Computer Science at the Technical University of Denmark (DTU COMPUTE). The work is done in accordance with the programme of department's PhD School (ITMAN) for acquiring the PhD degree.

Associate Professor Rasmus Reinhold Paulsen has been the main supervisor during the project, assisted by postdoctoral fellow Jens Fagertun.

A part of the research took place at the SIMBioSys group of the Department of Information and Communication Technologies at the Universitat Pompeu Fabra (UPF) in Barcelona under supervision by Professor Miguel Ángel González Ballester.

The research was funded from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement n°304857. The project was titled HEAR-EU and beyond DTU the consortium consisted of The University Pompeu Fabra and Alma IT systems in Barcelona, Spain, The University of Bern and Scanco Medical AG in Switzerland and finally MED-EL in Austria. Each of these collaborators have contributed to the material and research being presented.

The thesis revolves around data processing of inner ear specimens imaged with μCT, from which anatomical statistical shape models are built using image registration techniques. The building process, the models and their usage in conjunction with applications relating to Cochlear Implants are addressed.

Lyngby, 31-August-2015

Hans Martin Kjer

Contributions

Contributions included in thesis

A H.M. Kjer, J. Fagertun, S. Vera, D. Gil, M.A. González Ballester and R.R. Paulsen. Free-form Image Registration of Human Cochlear μCT Data Using Skeleton Similarity as Anatomical Prior.
Pattern Recognition Letters, Special Issue on Skeletonization and Applications, DOI: 10.1016/j.patrec.2015.07.017, 2015.

B H.M. Kjer, S. Vera, J. Fagertun, D. Gil, M.A. González Ballester and R.R. Paulsen. Image Registration of Cochlear micro-CT Data Using Heat Distribution Similarity.
Lecture Notes in Computer Science, Scandinavian Conference on Image Analysis (SCIA), DOI: 10.1007/978-3-319-19665-7_20, 2015.

C H.M. Kjer, J. Fagertun, S. Vera, D. Gil, M. A. González Ballester and R. R. Paulsen. Shape Modelling of the Inner Ear from Micro-CT Data.
Shape Symposium, 2014.

D H.M. Kjer, S. Vera, J. Fagertun, F. Perez, J. Herrero, M.A. González Ballester and R.R. Paulsen. Predicting Detailed Inner Ear Anatomy From Pre-Operational CT for Cochlear Implant Surgery.
Proceedings of Computer Assisted Radiology and Surgery (CARS), 2015.

H J.R. Romero, H.M. Kjer, M. Ceresa and M.A. González Ballester. Multi-Region Statistical Shape Model for Cochlear Implantation.
(Submitted to) *SPIE Medical Imaging*, 2016.

E M. Ceresa, N. Mangado, H.D. Velardo, N.C. Herrezuelo, P. Mistrik, H.M. Kjer, S. Vera, R.R. Paulsen and M.A. González Ballester. Patient-Specific Simulation of Implant Placement and Function for Cochlear Implantation Surgery Planning.
Lecture Notes in Computer Science, Medical Image Computing and Computer Assisted Interventions (MICCAI), 2014.

F S. Vera, F. Perez, C. Balust, R. Trueba, J. Rubió, R. Calvo, X. Mazaira, A. Danasingh, L. Barazzetti, M. Reyes, M. Ceresa, J. Fagertun, H.M. Kjer, R.R. Paulsen and M.A. González Ballester. Patient Specific Simulation for Planning of Cochlear Implantation Surgery.
Lecture Notes in Computer Science, Clinical Image-based Procedures (CLIP), MICCAI workshop, 2014.

G N. Mangado, M. Ceresa, H.D. Velardo, H.M. Kjer, S. Vera, R.R. Paulsen, J. Fagertun, P. Mistrik, G. Piella and M.A. González Ballester. Monopolar Stimulation of the Implanted Cochlea: a Synthetic Population-Based Study.
Lecture Notes in Computer Science, Clinical Image-based Procedures (CLIP), MICCAI workshop, 2015.

I (In Preparation) H.M. Kjer, J. Fagertun, M. A. González Ballester and R. R. Paulsen and others. Patient Specific Estimation of Detailed Cochlea Shape From CT Images, 2015.

J (In Preparation) B. Braithwaite, H.M. Kjer, J. Fagertun and R. R. Paulsen. Cochlear Implant Electrode Localization in Post-Operative CT Using a Spherical Measure, 2015.

Other contributions by author

- H. M. Kjer, M. Ceresa, N. Carranza, S. Vera, M. A. González Ballester and R. R. Paulsen. Cochlear Finite Element Modelling, Mesh Quality under SSM-Driven Deformations.
  Mesh Processing in Medical Image Analysis (MESHMED), MICCAI workshop, 2013.

- H. M. Kjer, S. Vera, F. Perez, M. A. González-Ballester, R. R. Paulsen, Human Cochlea, Semi-Automatic Anatomical Measurements on μCT 3D Surface Models.
  Proceedings of the 13th International Conference on Cochlear Implants, 2014.

- M. Ceresa, H.M. Kjer, S. Vera, N. Carranza, F. Perez, L. Barazzetti, P. Mistrik, A. Dhanasingh, M. Caversaccio, M. Stauber, M. Reyes, R. Paulsen, M. A. González-Ballester. Finite element model for patient-specific functional simulations of cochlear implants, *Mesh Processing in Medical Image Analysis (MESHMED), MICCAI workshop*, 2013.

- N. Mangado, M. Ceresa, N. Duchateau, H. Dejea, H. M. Kjer, R. R. Paulsen, S. Vera, P. Mistrik, J. Herrero, M. A. González Ballester. Automatic Generation of a Computational Model for Monopolar Stimulation of Cochlear Implants. *Proceedings of Computer Assisted Radiology and Surgery (CARS)*, 2015.

- N. Mangado, N. Duchateau, M. Ceresa, H. M. Kjer, S. Vera, P. Mistrik, J. Herrero, M. A. González Ballester. Patient-Specific Virtual Insertion of Electrode Array for Electrical Simulations of Cochlear Implants. *Proceedings of Computer Assisted Radiology and Surgery (CARS)*, 2015.

- E. Ruiz, H. M. Kjer, S. Vera, M. Ceresa, R. Paulsen. M. A. González Ballester. Random Walks with Shape Prior for Cochlea Segmentation. *Proceedings of Computer Assisted Radiology and Surgery (CARS)*, 2015.

- S. Vera, R. Caro, F. Perez, M. Bordone, J. Herrero, H. M. Kjer, J. Fagertun, R. Paulsen, A. Dhanasingh, L. Barazzetti, M. Reyes, M. Ceresa, M. A. González Ballester. Cochlear Implant Planning, Selection and Simulation with Patient Specific Data. *Proceedings of Computer Assisted Radiology and Surgery (CARS)*, 2015.

Acknowledgements

First and foremost my sincere gratitude goes to my supervisor Associate Professor Rasmus Paulsen. His supervision and mentoring through the past many years have given me grounds for developing myself both academically and professionally but certainly also on a personal level. I will remember my years at DTU with fondness due to the projects that I have been so privileged to work with under his supervision.

Secondly, I thank Jens Fagertun, who as a postdoctoral researcher on the project for the past two years have given me much guidance, feedback and plenty of discussions about your work.

A special thank goes to Professor Miguel Ángel González Ballester, who as the Principal Investigator of the HEAR-EU research project and as my supervisor during my external stay in Barcelona has made a significant contribution to my work.

All of the members and collaborators from the HEAR-EU project also deserves a nod of acknowledgment. The collaboration has over the years provided me endless motivation, and from our interesting plenary meetings I take away with me valuable experiences concerning international cooperation. The long list of people includes Benjamin Braithwaite from DTU. Martin Stauber and Bruno Koller from Scanco Medical AG. Wilhelm Wimmer, Nicolas Gerber, Brett Bell and Marco Caversaccio from the ARTORG Center at the University of Bern. Sergio Vera, Frederic Perez, Maurizio Bordone from Alma IT systems. Livia Barazzetti and Mauricio Reyes from the University of Bern. Anandhan Dhanasingh and Pavel Mistrik from MED-EL, Austria. Mario Ceresa, Nerea Mangado,

Alexis Bagué Roldán, Esmeralda Ruiz Pujadas and Jordi Romero from Universidad Pompeu Fabra (UPF).

Many other colleagues from UPF deserves to mentioned as well. Sergio Sánchez Martínez, Carlos Yagüe Méndez, Quim Fernandez, Antonio Porras, Paula Rudenick, Martha Nuñez García, Sara Noureldin, Oualid Benkarim, Bruno Paun, Veronica Zimmer and more. The hospitality and openness shown towards a stranger from the north, made it easy to feel welcome and ensured that I had wonderful and memorable external stay.

No acknowledgment would be complete without thanking the countless past and present colleagues at DTU Compute. I have had the honour of sharing office with a lot of fine people during the years, who each in their own way have contributed to making it a pleasurable working environment. I am extremely grateful for the good times we have had. Not only on a daily basis, but also at social events such as boardgames and movie-nights. Not forgetting the many off-work physical activities such as obstacle races, bouldering, running etc. that I have enjoyed tremendously doing.

Contents

Chapter 1

Introduction

The senses have a major impact on the life of all organisms. They are responsible for the perception of the world around us, they help shape our experiences and impacts our ability to interact with the surroundings. Generally life is richer with sharp and well-functioning senses, and their function is often taken for granted.

Humans are traditionally considered to have five senses, with sight (*ophthalmoception*), hearing (*audioception*) and touch (*tactioception*) as the most important or impact-full and secondarily smell (*olfacception*) and taste (*gustaoception*). By some viewpoint balance and acceleration (*equilibrioception*), awareness of limbs positions (*proprioception*), temperature (*thermoception*) and pain (*nociception*) could also be defined as senses. Whatever the definition, a loss of a sensory function is debilitating to a lesser or greater extent. Performance of the senses tend to gradually decline with age, meaning that it affects us all in some way sooner or later. Age related loss of functionality happens on a slow enough time-scale that our brains can adjust to the change of perception, and we are often not able to recognize the loss until accounted or corrected for.

Unsurprisingly, recovery of lost sensory function is today a big industry, at least when it comes to correction of impaired vision and hearing. The markets for eyewear and hearing aids are enormous. Optical lenses in the form of spectacles and devices for sound amplification are common and have been around for ages.

Taking a look at the historical development of these devices is a study on its own, basically telling the story of human technological improvements [Mil11] and of changes in society and culture. Considering the projected demographic development with regards to increasing lifespans and a higher focus on geriatric care, then these markets and industries are unlikely to shrink in size.

The devices and technologies holds interesting futuristic perspectives as well. Not only will they be able to compensate for impairments, but also enrich our sensory input with additional information. Augmented Reality (AR) is coming closer. Devices will pick up informative or additional sensory signals and project them into our naturally perceived sensory-landscape as an overlay or alteration, which is not too intrusive nor disruptive. AR is a research field on its own topic [Azu97, VKP10] and can be illustrated by the following recent cases:

- **Augmented sight:** Google Glasses is arguably the most prominent recent case of augmented vision. The concept is basically a spectacle functioning as a head mounted display where "the see-through lens could display everything from text messages to maps to reminders. They may be capable of showing video chats, providing turn-by-turn directions, taking photos and recording notes – all through simple voice commands" [Gol12].

- **Augmented hearing:** New Scientist published a feature concerning a man with a modern hearing aid, which he had hacked. The augmented signal was in this case information about near-by wi-fi networks that smart phones collect. His hearing was augmented with the wi-fi network information by letting "distant signals click and pop like hits on a Geiger counter, while the strongest bleat their network ID in a looped melody" [Swa14].

These kind of AR technologies are perhaps more fascinating than useful and practical at present. The hype of the Google glasses have died, and the device saw a lot of issues during its short lifetime [New15]. It might take years for these technologies to mature, but the future perspective is nevertheless highly interesting.

These cases were simply meant as an introductory appetizer for why biomedical technologies and applications are interesting with regards to the human senses, even for someone who might not be in the immediate risk group of facing severe sensory impairment within a short timespan.

Sensory Bionics

However common and helpful eyewear and hearing aids are today, the amount of sensory loss they can compensate for is fundamentally limited. They work by shaping or altering the incoming sensory signal enough to allow the perceiving organ to receive and interpret it properly. However, the sensory system must be functional for this to work. At some point, the hearing or the sight can become so impaired that no amount of sound amplification nor shaping of the light is able to recover the sensory function adequately.

When a sensory function in general has degraded enough or is lost, more advanced biomedical technologies can come into play. This could be devices or implants that either changes the shape, function or the activation of the sensory system it-self or altogether replacing it. Such technologies might be referred to as *sensory bionics* or *neural-* or *sensory prosthetics*. Below is a list of such devices, which are either already on the market or being researched.

- **Sight:** Modifications of the eyes can take place in two ways depending on the debilitating cause. Degradations in the focusing system (i.e. the cornea and lens) are very common. Treatment of cataract by replacing the natural clouded lens with an Intraocular Lens (IOL) has been an established procedure for years [AML+84]. However, as a bionic device the retinal implant is much more fascinating. They provide vision by using external cameras to record the incoming light, project the signal into the eye, where an array of implanted electrodes stimulate the remaining and functioning retina neurons. "These devices have evolved to a level where in current clinical trials they are beginning to restore basic visual function to blind individuals" [OdC12].
- **Hearing:** A selection of implantable hearing aids is already on the markets. Most notable and widely used is the Cochlear Implant (CI), where an array of electrodes is surgically inserted into the cochlear. The CI recovers the sense of hearing by recording sounds externally and then electrically stimulating the cochlear nerves in accordance with the recorded sound signal. Only the cochlear nerve and brain is required to be functioning properly for this to work [WD08]. It is not an exaggerated statement to say that this device is allowing deaf people to hear.
- **Balance:** Vestibular implants are an emerging technology leaning on the experiences from CIs. The target of electric stimulation is the vestibular nerves instead of the cochlear nerve, in order to treat balance disorders. Conditions leading to a loss of balance can be severely debilitating, so even though a lot of research is still required, the technology has promising prospects [ML12].

A thing that all of these fascinating bionic devices have in common, is the fact that they can only be realized through collaborative and multi-disciplinary research. They exist on the interface between the biology and the technology, the clinician and the engineer. When one grows stronger so does the other. It is within this exciting area of biomedical engineering research that the present thesis will dig into.

1.1 Focus of the Thesis & HEAR-EU

The project focuses on research and applications relating to the Cochlear Implants (CI). The work presented here was part of a European research project - called HEAR-EU. With collaborators from institutions and companies in Switzerland, Austria, Spain and Denmark, this fact will undoubtedly make its' mark on the present thesis. Certain aspects of the work presented has to considered in the light of that.

Within the hearing aid industry, the CI-area still only occupies a minor part. With 300,000+ recipients worldwide as of 2012 [11-13] it is not considered a large area. However, within the field of bionics this is certainly not just research - it is a well established technology with competing products and companies, which in it-self is remarkable. The CI-area is rapidly growing (in 2008 the number of CI-user exceeded just 120,000 [WD08]), and perhaps one of the most noteworthy aspects about the device is the large percentage of children receiving the implant (in 2012 the adult/child implantee ratio was approximately 58,000/38,000 [11-13]).

The CI was presented very briefly above and a more comprehensive description will be given later. Without spoiling anything, it is safe to state that things are more complicated than the basic description suggests. The technology is challenged on multiple fronts and there is still a significant gap between natural- and CI-based hearing [WD08].

Within the HEAR-EU project and the present thesis the focus is on the CI challenges relating to biomedical engineering. The question that we seek to answer coan be formulated as,

- **What are the necessary steps of improvements that can bridge the current gap between natural- and CI-induced hearing?**

The following notion is a good starting point: Just as the lenses of ordinary glasses and settings of a standard hearing aid need to be tailored to the specific

user, then so should a Cochlear Implant. The CI is permanently implanted and something conceivably much more advanced and intricate. It would be natural to think that such a device needs to be even more rigorously user-specifically designed and modified. Remarkably, this is not what is happening today. There are reasons for that, and the current limitations of CIs will be more thoroughly described later on. However the above statement still points to the core idea of the thesis.

- **Improvements to the performance of CI can be achieved by taking into account patient-specific factors.**

Patient-specificity is becoming a important variable in modern medicine and biomedical research, as exemplified with the recent Precision Medicine Initiative [CV15]. Within the field of CI, the use of patient specific models could possibly lead to improvements in the following areas:

Clinical benefits: By having patient specific computational models before surgery it may be possible to advise on the implant type or design, which is suited best for the recipient.

Post-operatively, a combination of a patient specific anatomical model with an estimation or measurement of the actual electrode array placement, can provide the audiologist with a detailed understanding of the system and facilitate means for optimizing CI-programming [NLGD13].

- Realization of these benefits requires that an accurate and detailed model of the CI recipients' anatomy and physiology can be provided.

Manufacturer benefits: For the CI companies, having access to models of the human variability in cochlear anatomy and physiology opens up possibilities for virtual product design, testing and optimization. The electrode implant types and designs available today is based on a 'one-size-fits-all' scheme (which a bit harshly put could be called a 'one-size-fits-few' or even 'none' strategy).

In an ideal future, companies will be able to produce an implant designed specifically for a recipient. However, a more immediate plausible step forward could be the development of implants designs, which fit favorably for a certain group or type of persons.

- Realization of these aspects requires a good understanding of the anatomical variability in the population.

Patient benefits: All of the above should separately lead to improvements in the hearing of the CI-user. However, there is also another interesting point to make. To the normal human our hearing is basically a black box. It works the way it works and there is nothing to do about it really. However, this is not the situation for CI-users. They have a permanent implant with a computer controlling their perceived sound-scape. They can in theory tune and adjust their hearing based on specific situations and environments. Today this is limited to simple things like volume control. In order to expand upon the possibilities of adapting hearing one needs to know how a certain scenario affects everyone, but at the same time how that translates into specific changes for a particular user.

- To close the gap between CI- and natural hearing, one must understand both the statistical variability in the population and the patient specific anatomy and physiology.

The ideas and statements presented above lead to one of the concrete tasks of the HEAR-EU project and the fundamental topic of this thesis:

- **How to make a statistical model of detailed inner ear anatomy and physiology?**

The proposed solution and procedure is presented in Figure 1.1. It is a pipeline going from the physical preparation of sample specimens to be scanned with μCT, through various image processing steps, to end up with a statistical shape model (SSM). The inner ear anatomy has to be studied with μCT from ex-vivo specimens in order to generate realistic shape models with enough anatomical detail. The shape model can describe the population variability in a sensible manner and further be used in estimation of patient specific anatomy from CT images.

Note, that while the pipeline (Figure 1.1) is an integral part of the HEAR-EU project, the scope of HEAR-EU is larger and wider. The illustrated tasks does in no way represent an equal amount of work nor difficulty. The illustration is simply expanded to show the work-flow and the areas in which this thesis is focused and where contributions have been made.

In general, the research conducted and presented in the thesis falls under the field of computational anatomy. Aspects of image analysis (segmentation, registration), image processing and handling, optimization (model fitting, registration), multivariate statistics and statistical shape models and mesh processing will appear as parts of the thesis.

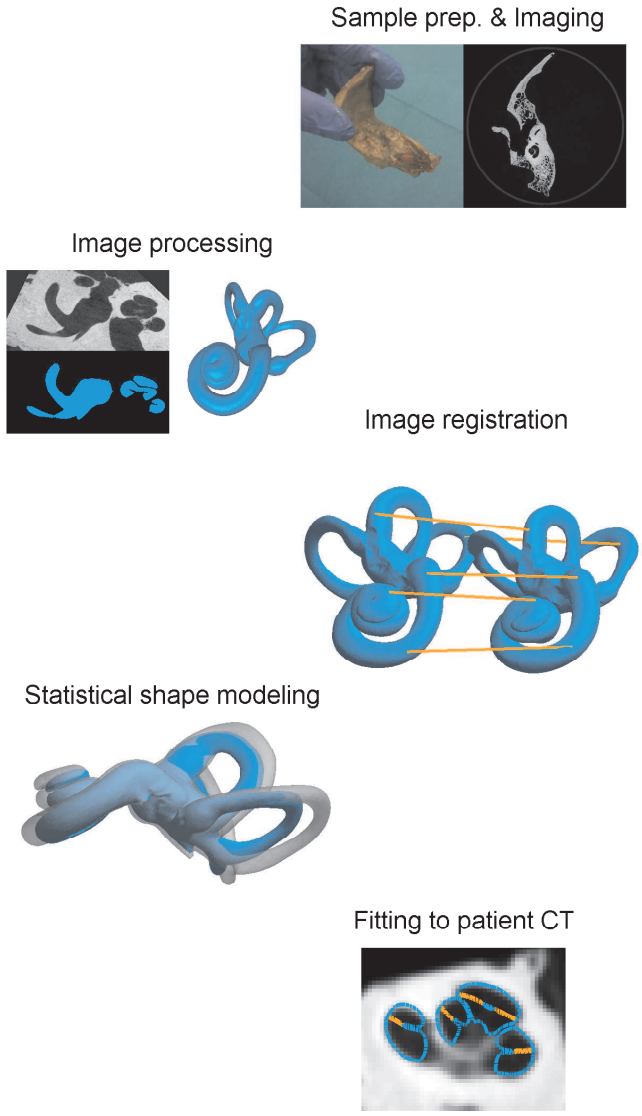


Figure 1.1: Thesis overview. A part of the HEAR-EU pipeline, but the illustration is modified to show the flow of the thesis and the areas in which contributions have been made.

1.2 Thesis Objective

Condensing the statements from above leads to the main objectives of the thesis,

- Construct statistical shape models of the human inner ear and cochlea from μCT data
  - Develop registration models suited for the anatomy and modality.
  - Build statistical shape models from the available data
  - Set-up a fitting procedure, in order to make patient specific prediction of cochlear shape from CT images.
- Explore how the models can be used in applications relating to Cochlear Implants.

1.3 Thesis Overview and Reading Guidelines

The thesis is intended as a chronological report, where a given chapter usually relies on the preceding ones, largely following the flow presented in Figure 1.1. Certain chapters are structured in a manner that allows a reader familiar with the particular subject to skip larger parts of it.

Most of the applied techniques have been used before in other fields, with other kinds of data or with different applications in mind. However, the data should not be treated as a black box, which was simply fed into a standard methodological framework. It is ultimately the available data and the intended applications that sets the limitations and controls the methodological choices of our work. This thesis is therefore also intended to convey a view on the anatomy and morphology of the inner ear, so that the reader hopefully can see the rationale behind the project and the process.

The overall structure of the thesis:

- **Chapters 2 & 3,** provides 'soft' background knowledge concerning inner ear anatomy, hearing physiology and Cochlear Implants.
- **Chapter 4,** looks at the CT and μCT image acquisition and the resulting datasets used throughout the thesis. Various types of data processing are described and applied, including inner ear shape characterization in a classic morphological sense.
- **Chapter 5,** dwells into image registration and the challenges of establishing point correspondences between the μCT data.

- **Chapter 6,** is concerned with constructing statistical shape models and procedures for fitting and using them.
- **Chapter 7,** goes into the various applications that the preceding chapters were building towards. Most importantly is the prediction of patient specific anatomy from CT data.
- **Chapter 8,** summaries and concludes the thesis.
- **Appendixes A - J,** holds a copy of the included scientific contributions.
- **Appendix K** - Table showing varying electrode designs.
- **Appendix L** - Table giving an overview of author contributions.

Chapter 2

The Inner Ear - Hearing and Balance

The goal of this chapter is to provide enough background information about the human hearing and balance to make the thesis self-contained for a reader unfamiliar with the subject and terminology. For the reader knowledgeable in the human inner ear and hearing, larger sections of this chapter can easily be skipped without significant loss of coherency with the remainder of the thesis.

In any case, the major points are summarized in Section 2.3. These points emphasize the presented details that are important for understanding the material in the later chapters.

For more comprehensive and in-depth sources on the topic, the reader is referred to textbooks like *Anatomy & Physiology* [SSP08], *Principles of Neural Science* [KSJ+12] and *The Temporal Bone: An Imaging Atlas* [LW10].

2.1 The Anatomy and Physiology of Hearing

An understanding of the hearing takes place on multiple scales. The most global perspective is illustrated in Figure 2.1, and formally divides the ear into three parts:

- **The outer ear:** Consisting of the *auricle* and the *external auditory meatus* this part is primarily seen as a passageway from the outside to the middle ear. Note that the auricle plays an important part in filtering incoming sounds, allowing for instance the direction to a sound source to be perceived.
- **The middle ear:** The system is comprised of the *tympanic membrane* and the *ossicles* bones (*malleus*, *incus* and *stapes*) and resides in the middle ear cavity. The function of the system is to convert an incoming sound wave into mechanical vibrations.
- **The inner ear:** Also called the *bony labyrinth* because it is a closed and fluid-filled tunnel system embedded within the temporal bone. It consists of three overall parts - the *cochlea*, the *vestibule* and the *semicircular canals* (SCC). Traversing the entire inner ear is a separate closed and fluid-filled chamber known as the *membranous labyrinth* (see Figure 2.4). The inner ear has two membrane-covered openings into the middle ear cavity - the *round* (RW) and the *oval window* (OW). The footplate of the stapes is connected to the oval window. This interface serves the purpose of converting the mechanical vibrations of the ossicles into fluid waves propagating in the cochlea. The motion of the fluid triggers electric stimulations of the innervating cochlear nerve. The neural signal travels to the auditory cortex of the brain resulting in a perceived sound.

2.1.1 Cochlear Anatomy

The cochlea needs to be described on finer scales for appreciating how exactly the hearing works. The illustration on Figure 2.2 serves this purpose by approximately showing the cochlea on a global-, intermediary- and finer scale.

Global scale: On this scale the human cochlea resembles a snail-like spiral structure with approximately two and half turns, that would roughly fit inside a $11 \times 9 \times 5$ mm box [EHWRA09]. Starting from the bottom the turns are referred to as the *basal-*, *middle* and *apical* turn. The most distant point on the spiral is called the *apex*. The turn direction of a left-sided cochlea is counter-clockwise when starting from the apex, and clockwise for a cochlea in the right side.

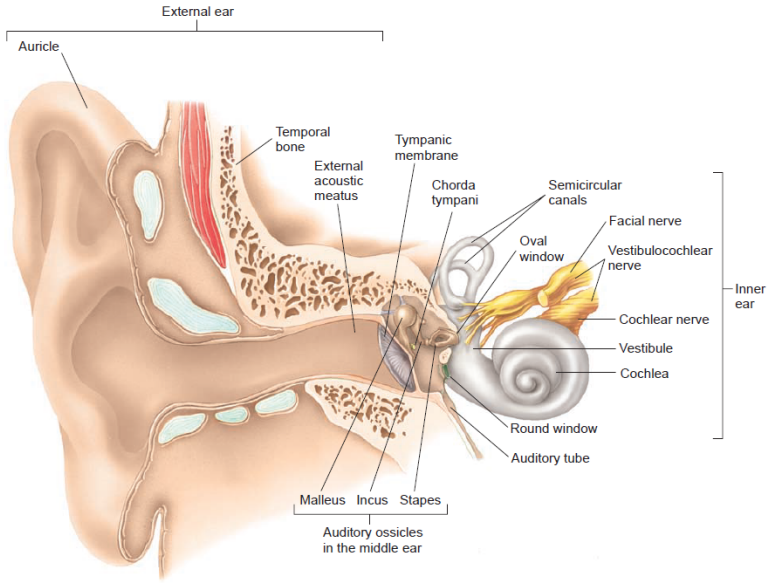


Figure 2.1: Overview of the human ear anatomy. Image from [SSP08].

The cochlea spins around the central bony part called the *modiolus*. The bone separating the different cochlear turns is referred to as the *interscalae septae*. The cochlear nerve goes from the brain through the *internal auditory canal* and into the modiolus at the cochlear basal turn. From here the cochlear nerve spreads out its fibers (*spiral ganglion*) to all of the cochlea.

The *cochlear aqueduct* supplying the bony labyrinth with fluid (perilymph) enters into the cochlear on the bottom of the basal turn close to the RW. The supply of fluid (endolymph) to the membranous labyrinth is through the *vestibular aqueduct* located in the vestibule (see Figure 2.4).

Intermediary scale: Observing a cross-section of a cochlear reveals a division into three chambers called *scala vestibuli*, *-tympani* and *-media* (cochlear duct). Scala vestibuli and scala tympani joins together at the *helicotrema* in the apical turn and is therefore technically a single chamber filled with perilymph. The scala media is the cochlear part of the membranous labyrinth, and thus filled with endolymph.

In a given cross-section the scalae are separated by a bony ridge called the *lamina spiralis*. It extends from the modiolus into the cochlea. The *basilar-* and *Reissner membrane* extends from the spiral lamina and goes across to join the

soft-tissue lining of the cochlear outer wall, which is called the *spiral ligament*. The region bound within the outer wall, the basilar- and the Reissner membrane defines the scala media.

The distance from the spiral lamina ridge to the outer wall (i.e. the length of the basilar membrane) varies linearly throughout the cochlea starting from ≈0.04 mm in the basal turn and ending at ≈0.5 mm at the helicotrema. A similar linear variation is present for the diameter of the collagen fibers making up the basilar membrane. The fibers are wide in the basal turn, making the membrane short and stiff, but becomes thinner towards the apex.

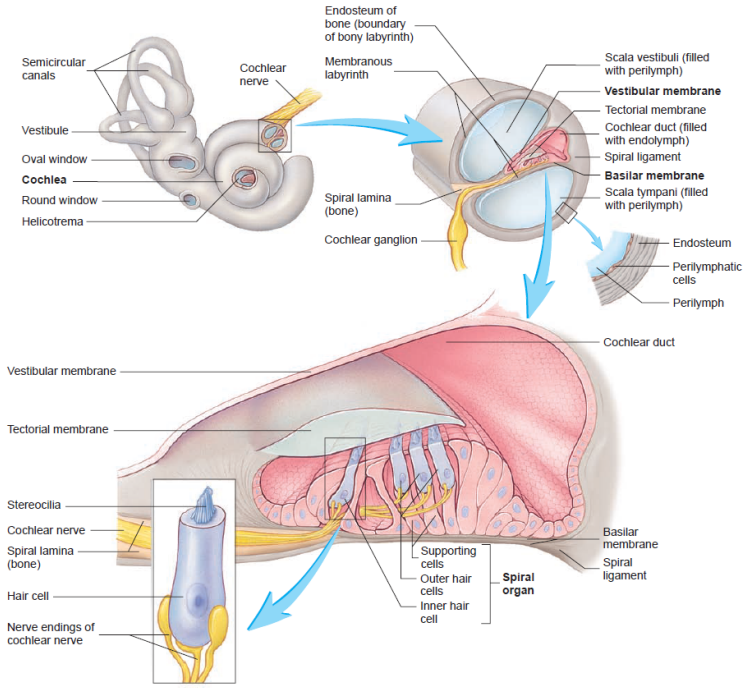


Figure 2.2: Cochlear Anatomy Illustration of the cochlea cross-section on gradually finer scales. Image from [SSP08].

Finer scale: Finally, it is worth taking a closer look into the scala media at the *organ of Corti* and the *tectorial membrane*, as this is the functional unit responsible for the last conversion step in the 'sound-to-neural stimuli' process.

The organ of Corti (spiral organ) resides on top of the basilar membrane towards the spiral lamina. It runs along the entire cochlea spiral and measures roughly

200×100 μm in a cross-section [RALE+12]. It is a structure consisting of supporting cells and specialized sensory cells, called hair-cells due to their hair-like processes extending from their apical end. The basal end of each hair cell is innervated by synaptic terminals of the spiral ganglion nerve fibers. The nerves run through the lamina spiralis to the modiolus where they join to form the cochlear nerve.

Attached to the spiral lamina and lying on top of the organ of Corti is the tectorial membrane. It a gelatinous acellular matrix pushing on the hair bundles of the hair cells. This interaction plays an important role in the generation of neural stimuli.

2.1.1.1 Anatomical variability

One of the thesis objectives (Section 1.2) is to build statistical shape models. Considering the anatomical variability therefore becomes important as well. Reading text-books on anatomy and physiology provides mostly an artistic view of the form and shape, and usually offer only a single stylish instance of the anatomy. While it is an accurate rendering of the real anatomical form, it provides no knowledge of the variability.

Understanding anatomical variances comes from observing multiple instances of real scans or specimens, which is typically beyond the scope of general text books to cover. It is possible to find observational studies on inner ear anatomical variability [EHWRA09, KSW+98, RALE+12, ANL+14, LSK+13, SLK+13, EJD+06]. Section 4.6.2 concerns characterization of inner ear anatomy and variability using classic morphology descriptions.

The following observations are relevant to take note of with regards to the work of this thesis,

- **Gender:** The morphology of the inner ear is independent of the gender. Some studies find the size of the male cochlear to be slightly larger, but considering the general inter-person variability, a distinction between male and female morphology is unnecessary [EJD+06, SLK+13].
- **Age:** The inner ear is fully developed in size around the 19th gestation week [JS04]. Postnatal growth occurs only in the surrounding structures and the temporal bone.
- **Left and right:** Anatomical studies find no significant differences between the morphology of the left and right sided cochlear [ERA13, SLK+13]. For a given person there is some suggestion of strong bilateral symmetry [RALE+12].

2.1.2 Cochlear Physiology

This section provides a description of the cochlear functionality with regards to the normal hearing. This should provide the physiological context to the above explained anatomy, and more importantly introduce an absolute key concept for understanding the Cochlear Implantation strategy.

Air-to-fluid wave conversion: Starting on the global scale, a sound wave passes through the outer ear and is converted into mechanical vibrations of ossicles. The system of small bones serves the purpose of making an efficient transformation of the air-wave into a fluid-wave in the cochlear. A direct conversion between the two mediums would be problematic due to the density differences. The tympanic membrane has a large surface area required for absorbing the sound pressure wave. In comparison the footplate of the stapes-bone has a small surface, which efficiently can induce fluid-motion in the scala vestibuli by pushing on the oval window membrane. The bone and membrane is held together by a ring of fibrous tissue called *the annular ligament*.

Mechano-eletrical transduction: The induced wave-motion of the perilymph in the scala vestibuli is propagated to the remaining scalae through the thin and flexible basilar- and Reissner membranes. The wave travels in the direction towards the apex, returning back through the scala tympani arriving finally at the location of the round window (RW). If the RW had been solidly closed the wave would be reflected back causing undesirable interference. Instead the RW is closed with a membrane that can dampen this effect. However, a more detailed description of the cochlea fluid mechanics is not required for this thesis. A review of modeling of cochlear mechanics can be found in [NEAT14].

The key point to understand is that the propagating fluid wave causes vibrations of the basilar membrane, which pushed the hair cells against the tectorial membrane. This causes gated ion channels on the hair-like processes to open. A sufficiently large influx of ions triggers a release of neurotransmitters in the basal region of the hair cells, which can initiate a neural impulse. The compositional difference between the peri- and endolymph creates a 140 mV potential difference between the apical and the basal portion of the hair cells residing in the organ of Corti. This differences plays a role in the dynamics of ion transport responsible for releasing the neurotransmitters. However, the intricate details of neural stimulation is beyond the scope of this chapter to describe. Further details can be found in for instance [KSJ$^+$12, SSP08]. It is sufficient to simply point out, that if the vibrations of the basilar membrane are strong enough, then it causes a electrical stimulation of the spiral ganglion nerve.

The tonotopic map: What really matters for understanding the auditory function is the elegant relationship between sound frequency, anatomical location and nerve impulse generation. The relationship is also known as the tonotopic map and is illustrated in Figure 2.3.

As stated previously the basilar membrane has varying properties along the length of the cochlear spiral. The changing width and stiffness means that each point of the basilar membrane will resonate at a specific wave-frequency. The cochlea works as a frequency analyzer, splitting up complex sound patterns consisting of many different frequencies into strong vibrations at specific spatial locations. Thus allowing frequency-based selective activation of neurons, that the brain is then able to interpret and make sense of.

2.1.3 Hearing loss and defects

A brief description of hearing loss types are given here to the extend required for understanding this thesis. Impairment of auditory function is divided into broad categories dependent on where in whole auditory system the problem resides,

- **Conductive:** If the loss of hearing is caused by sounds being prevented from reaching the inner ear, then it is categorized as a conductive. The issue is located in the outer or middle ear and is typically of a mild and temporary nature or at least amendable.
- **Sensorineural:** This type of hearing impairment originates from within the cochlea and is caused by a loss or degeneration of the hair cells. The severity varies greatly dependent on the amount of remaining and functional hair cells. Sensorineural hearing loss comes with aging, but also exposure to loud noises, hereditary conditions and even certain diseases such as meningitis can contribute to hair cell loss. When the hearing loss is still mild, a typical sufficient remedy is a standard hearing aid. If and when the impairment deteriorates to a severe or profound loss, then the CI may be a solution.
- **Partial:** A sensorineural hearing loss that only affects the high-frequencies is referred to as a partial hearing loss. This corresponds to the hair cells in the basal part of the cochlea being more degenerated than towards the apex. The distinction between partial and sensorineural hearing loss has implications for the choice of treatment strategy.
- **Neural:** Although a much rarer condition, neural hearing loss is caused by a damaged hearing nerve or even auditory cortex. Typical causes would be congenital malformations or head traumas.
- **Mixed:** A hearing loss can finally be categorized as mixed if two or more of the above categories apply.

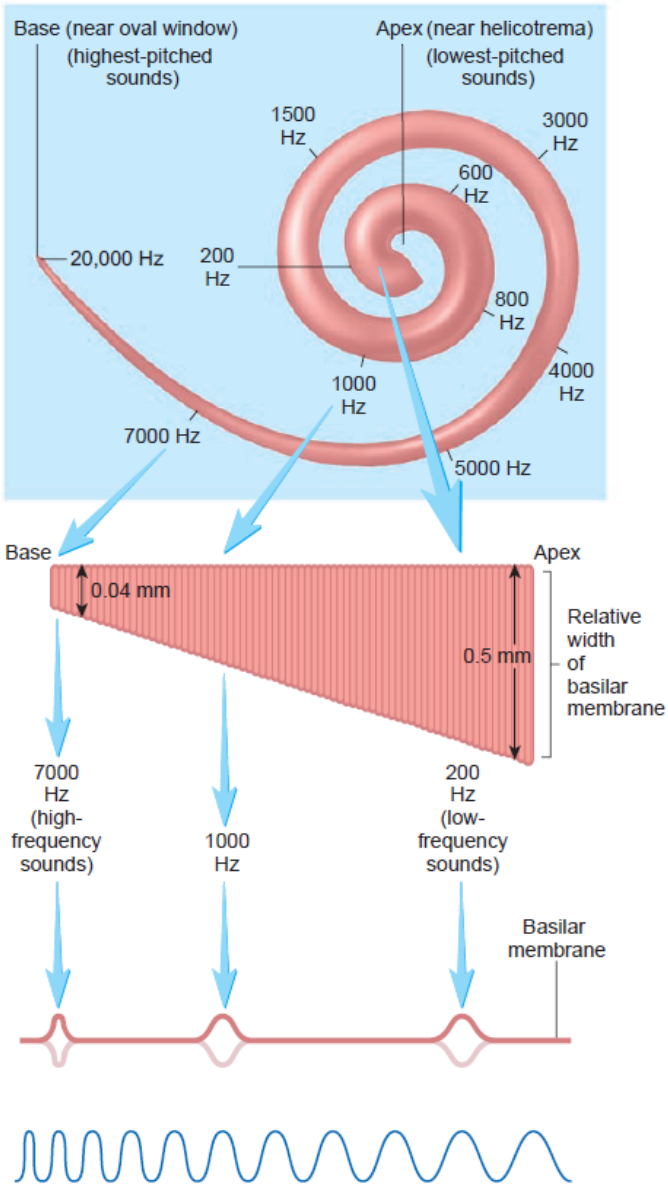


Figure 2.3: The tonotopic map. The variation in the basilar membrane width and stiffness causes it to resonate with a particular sound frequency at a specific location. Illustration from [SSP08].

2.2 The Anatomy and Physiology of Balance

Although not strictly required for the main line applications of this thesis, there are however a few points in providing a brief description of the sense of balance. One, it provides a more complete picture of the inner ear structure and function. Secondly, there may be some relevant and illustrative statistical shape modeling points in operating with the full inner ear, rather than considering the cochlea solely.

The sense of balance is divided anatomically and physiologically in two parts - the static- and the kinetic labyrinth.

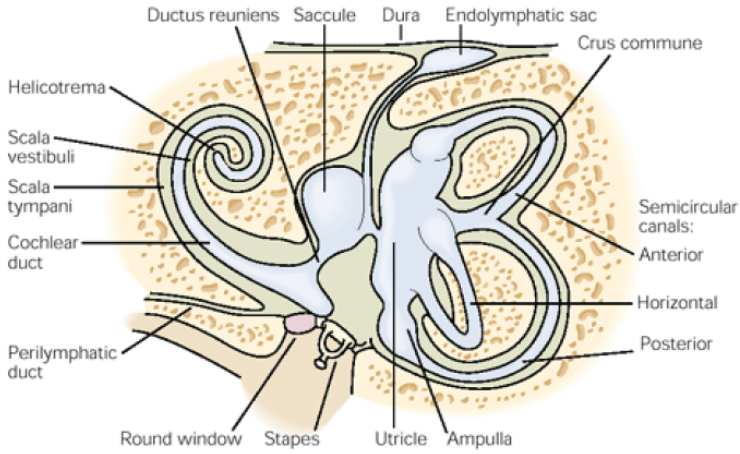


Figure 2.4: Overview of the membranous labyrinth. Illustration from [KSJ+12].

Static Balance: Within the vestibule the membranous labyrinth forms two oval chambers called the *utricle* and *saccule* (see Figure 2.4). Both contain a functional unit called the *macula*, which is a patch of epithelium bearing much resemblance to the organ of Corti. The macula contains cells with hair like processes called stereocilia, which are further embedded in a gelatinous mass called *otolith* (equivalent to the hair cells and tectorial membrane). The hair cells synapses with the vestibular nerve fibers. The maculae are oriented to sense linear acceleration in the horizontal (utricle) and vertical plane (saccule). The otoliths are weighted in such a way that a low level of stimuli is generated constantly. The brain is always being provided information about the head position and tilt in this way even when not in motion.

Kinetic Balance: The three semicircular canals (SCC) make up the kinect labyrinth. They are placed in a close to orthogonal configuration and are referred to as the anterior/superior- (aSCC), the posterior- (pSCC) and the lateral/horizontal semicircular canal (lSCC) (see Figure 2.4). One end of each canal expands into a spheric chamber called an *ampulla*. The non-ampullae ends of the aSCC and pSCC joins together in the *common crus* (crus commune). The functional unit resides within each ampullae and resembles the macula or the organ of Corti. A specialized sensory epithelium called the *crista ampullaris* contains hair cells that synapses with the vestibular nerve fibers. The hair like processes of the crista hair cells are embedded in a curved gelatinous mass called the *cupula*. During rotational movements of the head the angular accelerations causes the endolymph in the SCCs to move. This exerts a displacement of the cupula and triggers neural stimuli providing the brain with information about the rotations.

2.2.1 Balance loss

Just as age takes its toll on the hair cells in the organ of Corti, so does the hair cells of the saccule, utricle and ampullae degenerate with time. Balance impairment can arise from a variety of other sources, many of which are shared causes of hearing impairment - ranging from an infection (meningitis), ototoxic drugs and medicines, trauma, genetic or autoimmune conditions etc.

Balance impairment is generally associated with dizziness, vertigo, disorientation and nausea. When these things take place over brief periods and occurs infrequently, then the impact on life quality is minor. However, when conditions are more chronic and profound, balance impairment can prevent people from walking or even standing, and thereby be severely debilitating and arguably to a degree much greater than deafness.

The goal of this section was not to provide a comprehensive introduction to balance and balance disorders. The purpose was simply to point out the close relationship between the cochlea and the vestibular region both in anatomy and function. Building shape models of the vestibular anatomy and variability could have potential useful applications on its own.

2.3 Summary and Conclusion

The main points of the chapter are stated here,

- We have here provided an introduction to the anatomy and physiology of hearing and balance. A brief description of various anatomical terminology and concepts used throughout the thesis can be found here.
- The most important thing to understand about hearing with regards to Cochlear Implants is the elegant relationship between hearing anatomy and physiology. The spiral shaped cochlea is basically a sound frequency analyzer, due to the varying width and stiffness of the basilar membrane. This creates a very specific and unique functional relationship between an anatomical location and the brain's interpretation of a signal from the cochlear nerve. This relationship leads to a profound statement, essential for the approach taken with this thesis,
  - **The more accurately we can model the patient anatomy, the better we will be able to understand and optimize the functionality of the Cochlear Implant.**
- Characterization and modelling of the cochlear shape is independent of postnatal age, gender and whether it is left or right sided.
- A sensorineural or partial hearing loss is caused by the degeneration of hair cells within the organ of Corti. The loss is typically not uniformly distributed along the cochlear, and often most prominent in the basal turn (i.e. affecting the perception of high frequency sounds).
- A short introduction to the sense of balance was also given. The close anatomical and physiological relationship with the cochlea and hearing means that techniques and experiences that we develop for CI-applications could easily spill-over into development of vestibular implants.

Chapter 3

The Cochlear Implant

The chapter is meant to provide background knowledge about implantable hearing aids with a focus on the Cochlear Implant (CI). This should contextualize the challenges which we seek to solve. A reader with a familiarity to the topic can easily skip or skim through the sections of this chapter. The most important points are summarized in Section 3.7.

The articles of [Loi99, ZRH+08, WD08] provide further introduction to CIs. The information given here is mostly concerned with the biomedical and technological aspects of the CI, which has a relevance for the thesis.

3.1 Overview

A standard modern hearing aid basically consists of a microphone, a sound processing unit and a loudspeaker. It is a system designed to amplify (the desired parts of) the recorded sound signal, but otherwise complete rely on the normal hearing physiology to take care of the rest. Broadly speaking, the Cochlear Implant (CI) replaces the loudspeaker with an array of electrodes, which is surgically inserted into the cochlea, where it can activate the spiral ganglion nerve fibers with electric stimulation in accordance with the tonotopic map (see Section 2.1.2 and Figure 3.1). The brain will not be able to tell the difference whether a nerve signal was artificially stimulated or naturally generated. The CI is essentially a device made for exploiting the anatomical and physiological relationship of the cochlea.

While simple in theory, it is much more intricate to actually perform in practice. For each of the following components - making up the physical CI device and the related actors - there are inter-dependencies, variables and decisions that increases the complexity of the system,

- The external CI components
- The internal CI components
- The surgical procedure
- The anatomy and physiology
- The condition and culture of the recipient

Each of these points will be addressed separately below to highlight some of the challenges and limitations currently present. In the clinical practice there are some general aspects, which are not yet fully understood, and certain variables are unknown or uncertain for the specific case. This lack in knowledge (both general and case specific) results in a large variability in the hearing restoration outcomes among CI-recipients. A few users achieve near normal hearing (at least on some of the simpler standardized speech tests), while a few others experience little benefits of the having the device [WD08, PPM10].

The solution is naturally to increase the knowledge and our understanding of the system and its actors. In that regard, it is crucial to keep in mind the simple concept of the CI,

- **The more accurately one can exploit the tonotopic map, the better the potential is for improving hearing restoration.**

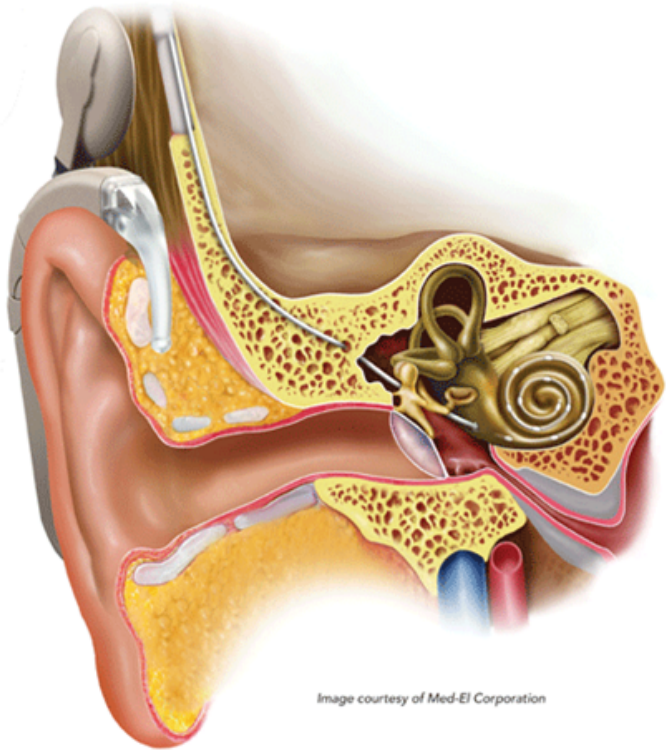


Figure 3.1: Overview of the human ear with CI. The external part consists of a microphone and a sound processor worn on the outer ear and the transmitter forming the connection to the internal part. The internal part consists of a receiver/stimulator unit (barely shown) connected to the electrode array inserted into the cochlea. Image courtesy of Med-EL Corporation.

3.2 External CI - Sound Processing

The external part of a CI bears much resemblance to that of a standard hearing aid. The sound signal is picked up by a microphone and is processed in the sound/speech processor unit. Where the output signal is normally sent to a loudspeaker, the CI sound processor ends up sending a set of instructions to the internal stimulator through the transmitter-receiver induction coils (Figure 3.2, left).

The sound processing should consider not only what the user wants to hear, but also what he/she is physiologically able to hear, and how the brain interprets a given stimuli. Generally speaking, only the part of the sound signal within the human frequency range (20 Hz to 20 kHz) is of interest. Noise should preferably be discarded or dampened. Dependent on the hearing impairment of the user, not all sound frequencies should be treated and amplified equally.

These considerations are very generic and applies to standard hearing aids and CIs alike. However, relating it to CIs specifically adds an extra layer of considerations. The output signal must be boiled down to a set of instructions to the electrodes. This is naturally heavily dependent on the amount of available electrodes and their location compared to the tonotopic map.

- Proper sound processing depends heavily on a patient specific model of the electrode placement in relation to the tonotopic map and knowledge of the hearing conditions of the recipient.

Figure 3.2: CI Components. (Left) The external sound processor and battery, connected to the transmitter pad. (Right) The internal receiver/stimulator and the electrode array. Images from `www.medel.com`.

3.3 Internal CI - Electrode Array Design

The implant consists of a receiver/stimulator unit and a connected array of electrodes embedded in silicone (Figure 3.2, right). The transcutaneous link between the internal (receiver) and external (transmitter) components is made by a pair of magnets to hold the transmitter pad in place and induction coils to transfer power and instructions as a radio frequency signal [ZRH+08]. The receiver unit resides in a drilled well in the mastoid part of the temporal bone. There are some design considerations regarding this unit. For instance concerning efficient power transfer across the skin and MRI safety. However, that plays little into the work of this thesis.

The design parameters of the electrode array on the other hand is of utmost relevance to discuss. Taking a look at the varying designs offered by manufacturers (Appendix K) provides little clarity as to what constitutes an efficient design. Variations in length, the number of electrodes, the electrode spacing etc. suggest that optimal design is one of the general aspects of CIs, which are not yet completely known and controlled.

Number of electrodes: Considering the core idea of being able to exploit the tonotopic map, it is reasonable to think that more electrodes is generally better, as it would allow a denser sampling of the sound frequency domain. However, that it is only true if the control of the electric stimulation is accurate. In reality, electrodes tend to stimulate a small region of nerves, which sometimes can cause stimulation of non-intended frequencies. This is known as cross-talk and it degrades the quality of the hearing restoration. An increasing number of closely spaced electrodes becomes more difficult to control, and are more likely to interfere with each others stimulation regions. Interestingly, some patients can recognize up to 20 different frequencies when stimulation is done one an individual electrode basis. However, when multiple of them are activated simultaneously in order to model complex sound patterns, then they can only have between 6-8 useful electrodes [WD08]. Having more electrodes in the array also means more wires and that changes the material stiffness properties of the array, which may have an impact on how the array behaves during the surgical insertion and the likelihood of causing trauma to intra-cochlear structures.

Array size: There is also the array length and the insertion depth to consider. The immediate obvious choice should favor a long electrode, which can cover the entire cochlear spiral (a deep insertion). The exact length would ideally depend and vary on the specific anatomy of the recipient. However, deep insertions are more difficult to achieve and are more likely to cause damage to internal cochlear structures [ZRH+08]. For recipients with a partial hearing loss (impaired in high frequencies), it is favored to do a shallow insertion and rely on the natural

hearing for the lower sound frequencies. Precisely because of this, an alternative to the CI is the so called Electro Acoustic Stimulation (EAS) approach. This device is basically a CI with a short electrode array combined with a standard hearing aid.

- In summary, a good electrode array design depends on the hearing conditions and anatomy of the recipient, material properties allowing for safe insertions, and the capabilities of which complex sounds can be represented.

3.4 Surgical Procedure

The typical surgical procedure starts with an incision behind the ear. An implant well is drilled in the mastoid bone to fit the receiver unit. A tunnel is drilled through the temporal bone to the middle ear cavity, providing direct access to the round window (RW). The RW membrane is cut open, and the array of electrodes is threaded into the scala tympani with little control of the implant placement from the surgeons side. The round window is covered up with bone residuals, which will solidify to a provide a seal of the cochlea.

This insertion through the round window is a standard approach, although there are plenty of varying ways, in which it can be done. The finer nuances of different surgical approaches is not within the scope of this thesis to discuss. It is sufficient to note, that the aim of a successful surgical procedure is to,

- Insert the array in the scala tympani to the desired insertion depth, preferably with the electrodes close to the spiral ganglion nerves.
- Avoid damage to intra-cochlear structures and preserve the natural hearing in cases where this is desired.

Achieving this is primarily challenged by the lack of control over the implant behavior during insertion. The surgeon can only gently push the array further into the cochlea from the RW entrance. Deeper insertions are less likely to obtain good implant placement. The array can fold and move in a backward direction, or be blocked or get stuck due to some anatomical malformation or irregularity. Forcing the electrode further into the cochlea increases the risk of damaging the spiral lamina, or rupturing through the basilar membrane to have the array pass into the scala vestibuli. The rigidity and stiffness properties of the array and the insertion angle at the RW conceivably influence how the array behaves during insertion, but this is generally not well understood. The lack of control means that the final position of the electrodes per default is unknown.

Simply stated, deciding on the surgical strategy is therefore a trade-off between the desired insertion depth (i.e. length of electrode array) and the risk of trauma. One opinion favors atraumatic insertion over better cochlear coverage. No one knows what treatments will become available in the future, and since many CI-recipients are children, doing no harm takes precedence.

Imaging: Although the use of imaging techniques does not provide additional control over implant behavior during insertion, the possibility of having knowledge about the recipient anatomy pre-operatively or being able to assess the implant placement post-operatively can still be a valuable aspect of the surgical procedure.

Pre-operative imaging is usually done with CT, but MRI could be used as well. No matter the choice of modality, the resolution of the images is low considering the small size of the inner ear. Only little anatomical information can be directly extracted from the data. Today, pre-op imaging is mostly used for assessing whether there should be any structural malformations influencing the drilling access and the cochlea insertion path. Certain pathological conditions can ossify the bony labyrinth, which obviously could change the surgical strategy.

Post-operative imaging is only feasible with CT, since the metallic parts of the CI are not really compatible with MRI imaging. Metal artefacts can also degrade the quality of a CT image, but to a much lesser extend. The images still have a low resolution, in the sense of observable inner ear anatomical information. However, since metal is strongly contrasted it is easy to identify the positions of the electrode contacts for validating the success of the implantation.

Note that CT imaging (whether pre- or post-op) is not standard practice everywhere. The use of ionizing radiation is administered with care, especially since many CI-recipients are children. The benefit of taking the images must be crystal clear, before the given radiation dose can be justified.

In summary,

- the lack of positional control of the array during insertion means that the electrode array behavior and final placement depends the recipient anatomy and the material properties of the electrode array.
- CT imaging is a valuable source of information for assessing both the pre- and post-operative aspects of the surgery. However, the immediate content about the inner ear anatomy in the data is limited.
- without post-op CT imaging the electrode positions can only be assumed known for use in the speech processing.

3.5 Recipient Anatomy and Physiology

The major dependency relating to anatomy and physiology is naturally the tonotopic map. The relationship was theoretically described in Chapter 2. The practical formulation of the link between an anatomical position, x, in the cochlea and the perceived frequency, f, is known as the Greenwood function [Gre90],

$$f = A\,(10^{ax} - K) \tag{3.1}$$

where x is the fractional length along the cochlear spiral starting from the apex and ending in the basal turn. A and K are species dependent constants, with $A = 165.4$ and $K = 1$ recommended for humans. The constant a is conserved across mammalian species after scaling with cochlear length, meaning that $a = 2.1$ if x is measured in relative length of the cochlea, or $a = 0.06$ if x is measured in mm from the apex.

Note from Eq. 3.1 that the only dependency is the cochlear spiral length. The number of turns and other cochlear shape characteristics are not included in the model. It is possible that more advanced functional models could exist, but this is beyond the scope of this thesis to investigate. Nevertheless, the placement of the electrode array in relation to anatomy is important. As stated above in Section 3.4 the exact anatomical details are not well known in the specific recipient case, and the surgeon lacks control over the finer placement of the electrodes. This information is therefore typically unknown (and assumed perfect). The use of imaging (pre- and/or post-operatively) is the obvious solution to remedy this situation. Being able to measure or estimate the cochlear length and find electrode contact positions can add a lot of valuable information for controlling and optimizing aspects of CI procedures.

- The tonotopic map and the Greenwood function are central aspects to take into consideration, and pre- and post-operative imaging is an obvious way for acquiring knowledge about this.

3.6 Recipient Conditions and Culture

The biggest variable for hearing restoration outcomes is arguably the individual CI-user. Since no two persons are identical, the CI design and actors should be uniquely defined in an ideal world. Some important and noteworthy specific factors are presented in the following.

Conditions: The age of the recipient is an immediate variable to consider. The risks of trauma during surgical insertion and the use of ionizing radiation in imaging can be assessed quite different whether the intervention is made for an infant or an elderly. Broadly speaking, when it comes to child recipients the earlier the intervention the better the hearing outcome and development of language and speech skills can be achieved [PPM10]. This explains why CI-surgery is regularly done in infants before the age of 1 year. Bilateral implantation generally results in better outcomes, but as it obviously involves two surgeries, twice the cost etc., this is not the automatic choice for a recipient of any age.

The absolute age is not the only time-related factor to have an influence. Longer periods of sensory deprivation has a negative impact on CI hearing restoration outcomes, most likely due to neural tissue degeneration or changes in nerve connections in the brain [WD08]. The duration of deafness before implantation therefore has an impact on what level of restoration that is achievable. Broadly speaking, post-lingually deafened persons recover better with a quick intervention, and pre-lingually deafened adults fare worse the older they are.

Finally, the type and severity of the hearing loss may favor a full insertion or a strategy aiming to preserve natural hearing in certain frequencies. As already stated, CI design and programming depend on this patient specific factor.

Culture: The ability to hear has a great impact on human communication and interaction. Arguably, this is the primary function of the hearing. However, it is important to note that language and speech is a product of culture. They are extremely difficult to master without the ability to hear, but hearing alone is not enough. For example, any person with a latin-based native language will find it difficult to learn and understand an oriental language like Japanese. Being hearing impaired will only make it more difficult.

The gained benefits of a CI is therefore heavily dependent on post-operative language and speech therapy. Essentially the brain needs to learn how to hear, interpret sounds and put it into a greater social/cultural context, and this is

naturally a hard and long training process. A key influence to hearing restoration outcomes is therefore the motivation of the recipient and the support from his/her family and immediate social network.

There is an additional curious note to make regarding CI efficacy dependency on culture. Some languages might be better suited for certain implant designs. There are languages relying heavily on using tones (Mandarin for example), i.e. where the sound pitch varies the meaning of words. Such a cultural aspect affects the way proper sound processing should be done for that population type [ZRH+08].

The major point from this section is that,

- In order to optimize the CI for a specific user, it is necessary to know the context and culture of the general population, as well as the user specific conditions.

3.7 Summary and Conclusion

Even though the idea of making a device to exploit the tonotopic structure of the cochlea seems simple and elegant, it is next to impossible to do perfectly. There is a large variation in hearing restoration outcomes, which can be attributed to the many different aspects and variables that are not controlled or known in the specific user case,

- During surgery the electrode array is blindly threaded into the cochlea with little control over the final placement.
- Electrode arrays designs are very generic, and the process of choosing a model suited for a particular user it not fully understood.
- Imaging is the source to provide more information about how to optimize certain aspects. However, the possibilities and potential lying within the data is not yet exploited.

Chapter 4

Image Data and Data Processing

The rest of the thesis builds upon images and data of the temporal bone and the inner ear. Several different types of datasets were acquired and made available throughout the project. They form the basis for our understanding of the anatomy, the reasoning for the methods that we developed to describe the anatomical shapes and the limitations in our built models. A chapter describing the nature of the datasets, their acquisition and processing is therefore a proper starting point.

First an overview of the datasets is provided in Section 4.1 for quick referencing. The imaging modality and sample preparation dictates what kind of anatomy that is observable in the data, which is described in Sections 4.2 and 4.3. Segmentation and extraction of surface models from the data is covered in Section 4.4 and 4.5. Finally, Sections 4.6 is dedicated to additional data processing steps.

4.1 Dataset Overview

The various datasets are summarized in Table 4.1 and a brief description given in the paragraphs below. Note that these datasets do not constitute all of the available data within the HEAR-EU project. There are additional scans of full heads and other specimen samples, both with and without implanted CI electrode arrays. However, as this thesis focuses on the process of building and validating statistical models of shape, this is merely how the data is bundled together and referenced meaningfully for the work presented.

Table 4.1: Overview of datasets

Dataset Name	**Dried**	**Thiel**	**TUM**
Anatomy	Temporal Bone	Inner Ear	Cochlear + Vestibule
Sample Prep.	Drying	Thiel-fixation	Freezing
Modality	'CT'/μCT	CT/μCT	μCT
No. Samples	17/18	14	1
Segmentations	yes	no	yes
Resolution μCT [μm]	16.3	7.6	5.9
Resolution CT [mm]	0.15	0.15	-
Type	Pre-op	Pre-op	Pre-op

Dried dataset: This dataset consist of 20 dried temporal bone samples from the anatomical collection of the University of Bern. They were scanned using μCT (Scanner type: μCT 100, Scanco Medical AG, Brüttisellen, Switzerland), and corresponding 'CT's were generated by downsampling the data to an appropriate voxel resolution.

One dataset was completely abnormal and not comparable to the remaining datasets. Another sample lacked almost all of the lamina spiralis and was therefore also excluded. Finally one data sample had a damaged posterior semicircular canal, and can therefore only be used in certain situations. This explains why Table 4.1 states 17/18 samples. The 18 usable samples contain a segmentation of the inner ear as one complete and closed object. Corresponding surface models were extracted (Section 4.5), and the samples were parameterized using landmarks (Section 4.6.3) and heat maps (Section 4.6.4).

Thiel dataset: This dataset comes from 8 unique cadaver heads preserved in a thiel solution. A region of the temporal bone around the inner ear was cut out and provided 14 usable samples, which were scanned both with CBCT and μCT (Scanner types respectively: ProMax 3D Max, Planmeca, Finland and μCT 100, Scanco Medical AG, Brüttisellen, Switzerland).

Some of the samples had fluid retained within the cochlea. This gives a significantly different appearance in the μCT images (See Figure 4.2). Dependent on the context in which the data is to be used, it could therefore be divided into subcategories.

TUM sample This dataset consists of a single human frozen sample, which was described in the work of Braun et al. [BBS12]. The region of the semicircular canals was not retained in the preparation process, in order to fit the specimen in a smaller container and scan the sample at a higher resolution (Scanner type: μCT 50, Scanco Medical AG, Brüttisellen, Switzerland).

The data has good contrast between certain intra-cochlear structures, which allows for a segmentation of the cochlea scalae, a cochlear partition (approximating the basilar membrane), the stapes and the surrounding bone. With these structures segmented, the dataset constitutes the basis for a detailed computational model of the cochlea anatomy. Both surface models and heat maps (Section 4.6.4) were calculated for the sample.

4.2 Modality

The anatomical information in an image depends heavily on the modality used. In this thesis only CT-based techniques are considered - namely CT/CBCT and μCT.

CT scanners are designed to image large structures, for example whole heads or bodies. A typical voxel size is in the order of 0.1-0.4 mm. The data is not always acquired isotropically, and the slice spacing can be significantly coarser than the in-plane resolution. For imaging of the inner ear, the clinical CT can provide a lot of useful information about the anatomy surrounding the bony labyrinth. However, the relative small size of the inner ear, means that only vague and rough information can be gained about the cochlea anatomy. No intra-cochlear details are present, and even the complete spiral characteristics can be difficult to recognize (See Figure 4.1). Post operative imaging with CT

is great for assessing electrode placement, but the metal artifacts only makes it more difficult to characterize the anatomy from the images.

μCT systems on the other hand are designed to scan small specimens or extremities, for example hands and feet. The achievable resolution of μCT depends on the size of the object and the allowed dosage (the latter only relevant when scanning living beings). Although not strictly defined, μCT images typically have voxel sizes in order of 5-50 μm. It is not possible to achieve this kind of resolution on whole human heads, so imaging of the inner ear has to be done on ex-vivo specimens of the whole temporal bone or samples cropped even more closely to the bony labyrinth. This way it becomes possible to observe intra-cochlear anatomical details, while still retaining a representation of the complete inner ear or cochlea. The thin bony processes of the interscalae septae and lamina spiralis are well contrasted even up to voxel sizes of approximately 50 μm. The basilar and Reissner membranes are normally only properly seen at resolutions below 10 μm, but this is also highly influenced by the sample preparation (see Section 4.3).

Note, that other modalities in principle can be used for imagning of the inner ear. MRI typically struggle with bone structures, but the fluid within the labyrinths and the nerve fibers can be imaged with good quality [LW10]. Techniques such as Scanning Electron Microscopy (SEM) allow for imaging of the cochlea on an even finer scale, such that a small region the organ of Corti and the hair-cells can be studied [RALE$^+$12]. However, imaging techniques often has a trade-off between the level of detail and the size of the imaged region. In regards to this project, we are aiming for modalities that can image the entire inner ear.

To summaries,

- Imaging of the inner ear with μCT provides data on a scale where it is possible to see intra-cochlear structures, while still being able to image the entire inner ear or cochlea. However, for humans it can only be acquired ex-vivo.
- Clinical CT/CBCT is the only commonly used modality to provide patient specific information about the inner ear or cochlea anatomy. However, the level of detail provided is very vague and coarse.

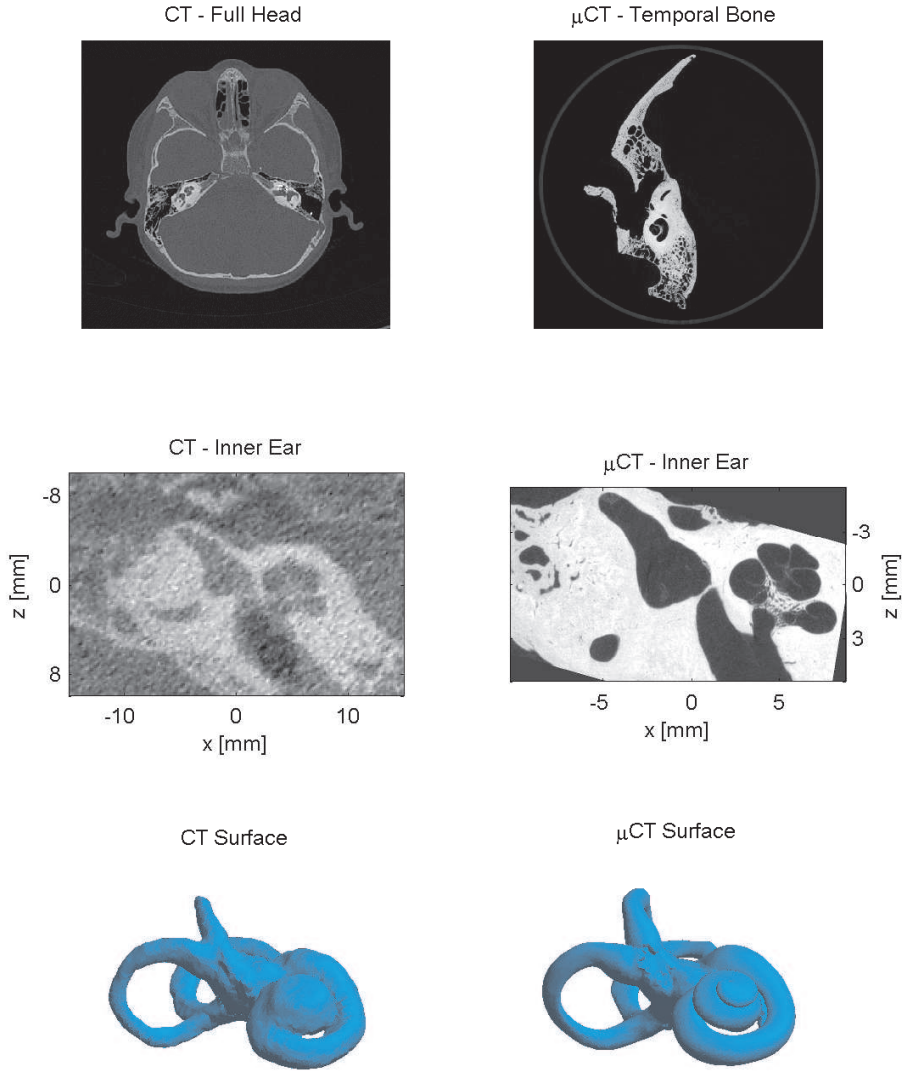


Figure 4.1: Imaging Modality. The image slices (middle row) showing the inner ear have respectively 0.15 mm and 46 μm voxelsizes. This corresponds to the approximately finest level of detail achievable with CT and a coarse μCT.

4.3 Sample Preparation

The preparation and processing of ex-vivo specimens also have an impact on the quality and usability of the obtained data.

First, there is the actual physical preparation to consider. A suitable region of interest has to be extracted. This is dependent on whether the intended application requires the full temporal bone to be imaged, or if the bony labyrinth is sufficient. Note that the inner ear is a structure embedded within a dense bone, so it is a tricky task to cut out a region around it.

Secondly, there is the choice of staining and fixation consider. It is not just a matter of preserving the sample without significantly altering and damaging the anatomy. Various types of staining can be used to enhance the contrast of structures. The specimens presented in this thesis were either dried, frozen or stained with a thiel-solution. The effect is illustrated in Figure 4.2, and the implications for the datasets are described below.

Dried: Bones can efficiently be preserved by drying. The technique is however rather destructive as it removes all soft tissues. For the temporal bones samples (when scanned with μCT) it is possible to clearly see the lamina spiralis and the interscalae septae. The data provides a detailed overall description of the bony labyrinth. The boundary between the middle ear cavity and the inner ear is non-existent, as the oval and round window membranes are not present.

Thiel: The samples from this dataset were stained in a thiel-solution, as this technique should preserve the mobility and flexibility of tissues. A region around the inner ear was extracted from the excised temporal bones. When imaged with μCT it is possible to see the basilar membrane and the spiral ligament. However, the Reissner membrane is not visible, but whether this is due to a lack in imaging resolution or the fixation is not known. Unfortunately, the majority (9 out of 14) of samples had retained fluid within the bony labyrinth disturbing the contrast of structures (see the comparison in Figure 4.2). The fluid-filled specimens resembles the dried samples to a large extent.

TUM: Only the cochlea and the majority of the vestibule was retained from the physical preparation of this specimen. No fixation technique beyond freezing was applied. The sample was thawed before imaging in a μCT scanner. Neither the spiral ligament nor the Reissner membrane was visible. However, a structure forming a partition between the cochlear outer wall and the spiral lamina was visible. This partition is obviously related to the basilar membrane, but it also suggests that some degeneration and structural alteration of the spiral ligament has happened in the preparation process (see [BBS12]).

In summary, the available ex-vivo specimens have the following characteristics when imaged with μCT,

- The data is generally able to give a full representation of the bony labyrinth of the entire inner ear.
- Only a few of the samples describe the appearance of the basilar membrane and the spiral ligament, and a distinction of the scala media/membranous labyrinth is not possible in any of the data.

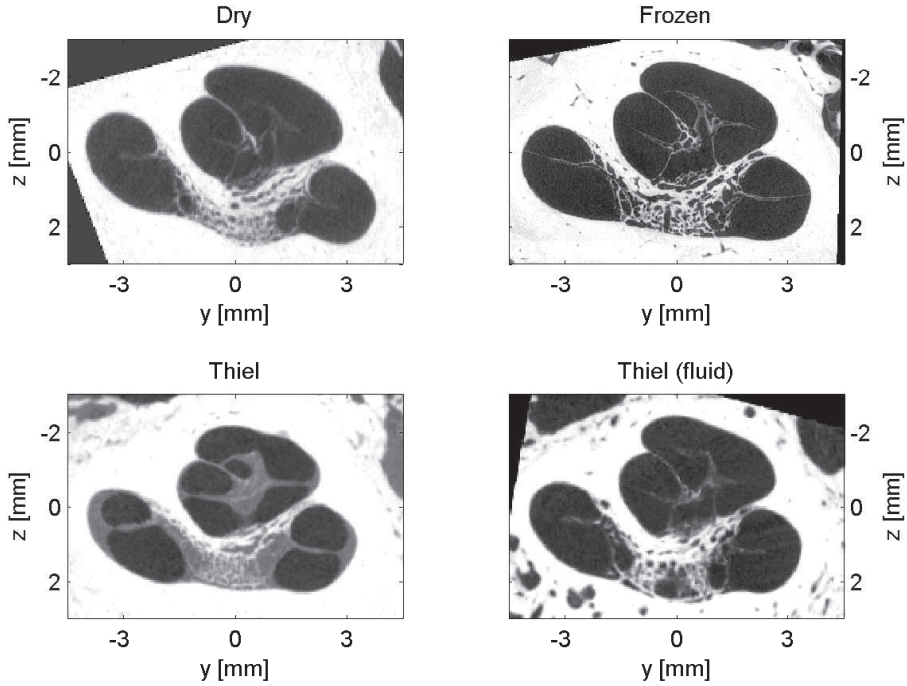


Figure 4.2: Sample preparation. Thiel-fixation provides excellent contrast of intra-cochlear anatomy, when the endo- and perilymph is drained out. Although difficult to see, the freezing procedure used for the TUM sample preserves a thin representation of a cochlear scala partition.

4.4 Segmentations

Image segmentation is the process of defining regions in the image data belonging to the same anatomy or structure. It typically involves a grouping of voxels sharing similar intensities, spatial location and connectivity.

Performing segmentations of the datasets serves multiple purposes. First, it facilitates simpler and better performance of other image processing and analysis tasks. Secondly, it becomes possible to extract 3D surfaces to represent the anatomical structures, and finally a proper segmentation can serve as a ground truth for evaluating the performance of other procedures.

Formally the task can be described as follows. Given a typical intensity volume, $\mathbf{D}$, where each of the N voxels are characterized by some value, d.

$$\mathbf{D} = (d_1, d_2, \ldots, d_N) \ , \ d_i \in \mathbb{R} \tag{4.1}$$

Segmentation is then the process where each voxel is assigned a label value, l, from a set of K possible classes corresponding to an anatomical structure,

$$\mathbf{L} = (l_1, l_2, \ldots, l_N) \ , \ l_i \in \{1, 2, \ldots, K\} \tag{4.2}$$

Image segmentation is a research field on its own. Establishing an automatic procedure is difficult, and in medical image analysis it is often necessary impose a priori anatomical knowledge in some way.

There is some published work on segmentations of the inner ear anatomy (see the review of Ferreira et al. [FGT14]), but these are mostly limited to clinical modalities. Very little work exists regarding segmentation of μCT data from ex-vivo specimens except for the work of Poznyakovskiy [PZK+11]. The approach taken is a contour segmentation of the cochlear scalae from 2D cross-sectional slices, which lies in planes perpendicular to the cochlear centerline direction. The anatomical preparation of the samples does not immediately correspond to our data. The differences in data quality and visible structures etc. makes it difficult to directly transfer what otherwise looks like a promising segmentation method to our data. In any case, within the work of this thesis, segmentation is seen more as a practical task to be solved, rather than being cast as a research problem. Development and validation of a fully automatic segmentation procedure would have been time-consuming, and as the μCT samples are limited in number, such a procedure was not a priority to acquire.

Fortunately, there are a variety of different existing software tools and packages that can readily be used to facilitate the segmentation process. The datasets in

this project were segmented using either ITK-SNAP [YPH+06], Seg3D [CIB14] or AMIRA. The former two are open-source packages and the latter a proprietary software. Because it is a time-consuming task, onlt the dried dataset and the TUM sample were segmented. The resulting data is illustrated in Figure 4.3 and additional details described below.

Dried: The segmentation of the dried samples contain only a single class, which represents the bony labyrinth. The rest is considered background. Note, that there is no separation between scala tympani and vestibuli.

The bulk of the segmentation was done with a semi-automatic level-set based method [OS88]. Manual corrections were required for handling obvious errors and undefined boundaries, such as the openings into the middle ear cavity. The use of a semi-automatic tool was critical for achieving smooth and rounded segmentations and for reducing the amount of manual work. Despite of that, a full segmentation could easily amount to approximately 12-15 h per dataset.

TUM: The segmentation process of this dataset is described in [BBS12]. It contains a label for the bony labyrinth (cochlea only), the cochlear partition, the surrounding bone, the round window membrane and the stapes. Some minor additional manual corrections were made to the segmentation. A small hole in the partition was closed, and some irregular noise was smoothened away.

Because of the segmented structures this sample is an excellent starting point for building a computational model of the cochlea.

To recap,

- The dried dataset and the TUM sample (Table 4.1) contain manual segmentations, **L**, of observable anatomical structures.
  - The dried samples were segmented as a single object of the bony labyrinth.
  - The TUM sample contain the cochlear part of the bony labyrinth and a cochlear partition. This makes it suitable for building an anatomical computational model.

4.5 Surface Models

The objective of this processing step is to extract surface models from the segmentations. Working with the data as volumes can be tedious. Visualization of

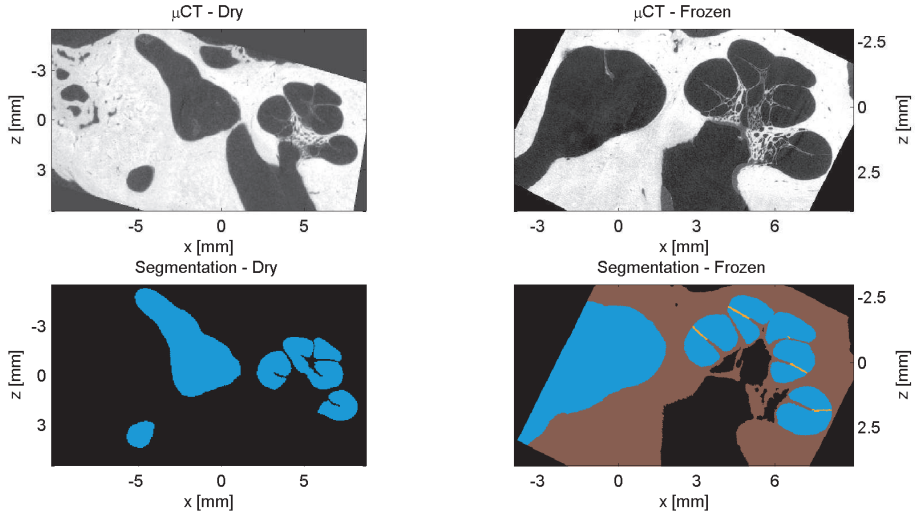


Figure 4.3: Data Sample Segmentation. The shown labels are: Bony labyrinth (blue), cochlear partition (orange) and bone (red).

3D structures is particularly difficult using single slices. Defining landmarks and measuring dimensions of objects is much easier when the true 3D representation of the object is considered.

Once a segmentation, $\mathbf{L}$, of a data sample has been obtained it is possible to extract a surface model, $\mathbb{S}$, corresponding to a specific segmentation label. The model consists of a set of points/vertices, $\mathcal{Z} \in \mathbb{R}^3$, and a structure connecting them in polygons. A triangle mesh is the only type of connectivity considered in this thesis, and the surface extraction is made in two steps, starting with an isosurface polygonization using Marching Cubes [LC87] followed by a Markov Random Field surface reconstruction [PBL10].

The Marching Cubes [LC87] algorithm is almost synonymous with a tool for polygonizing an object described by an implicit surface. However, applying it directly on a segmented voxel grid can result in a low quality triangular surface mesh. The rough mesh is however a good starting point for various surface reconstruction techniques, that can re-estimate the implicit surface better, and refine the polygonization to obtain a smoother mesh where the triangulated connectivity is more regular. Extracted surface models are shown in Figure 4.4.

The mesh vertex density and the mesh quality are the most important aspects of the surface models to consider. Enough vertices should be included to make a decent representation of even the small detailed structures like the lamina

spiralis, but having an excessive amount of vertices and triangles is mostly a computational burden. Since the applications of our work relates to computational models, the mesh should preferably have a good quality (i.e. no degenerate triangles), but any specific criteria would be application dependent.

In summary,

- The segmented datasets (Table 4.1) have corresponding 3D surface models, $\mathbb{S}$, representing the anatomical structures.

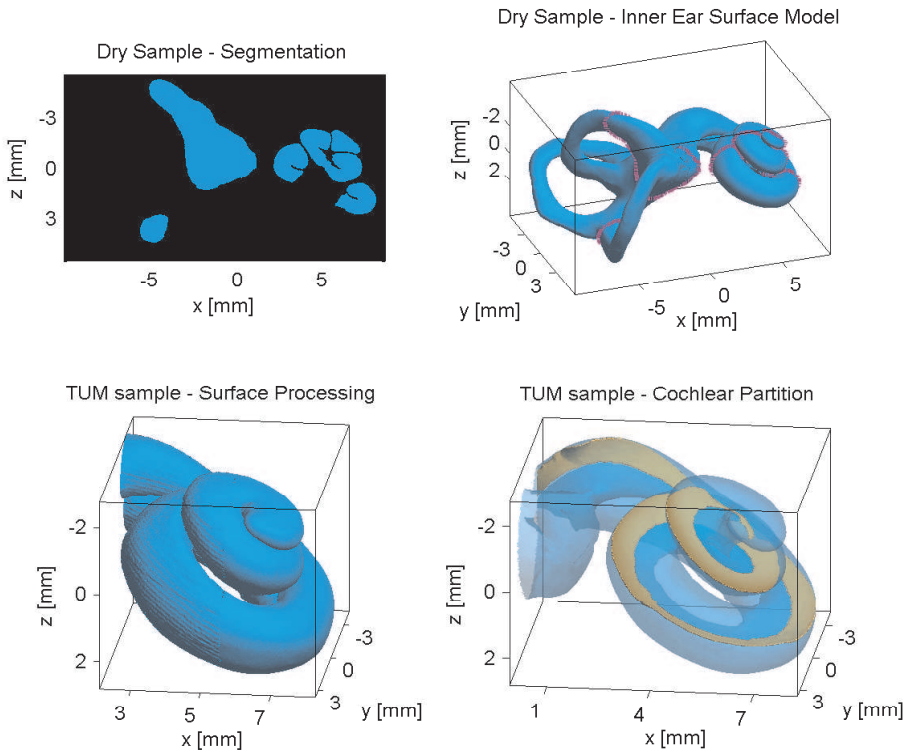


Figure 4.4: Surface Models Generation. Surfaces models are extracted from the segmented volumes. (Lower left) The initial mesh is rough and is therefore refined to obtain a smooth and well formed triangulated surface model. (Lower right) The TUM sample also contains a mesh of the cochlear partition.

4.6 Additional Data Processing

The data processing steps detailed in the preceding sections can be considered the core and minimal required steps. Based on those it would be possible to begin the process of building shape models, which is the topic of the subsequent Chapters 5 and 6.

However, instead of blindly tackling the shape modelling task, it can be beneficial to gain some a priori knowledge about the data. The content of the following subsections represent different ways of processing the data in order to make some sense and structure of it.

1. Initial rigid alignment of data.
2. Characterization of inner ear morphology
3. Cochlear skeletonization
4. Heat propagation through the cochlea.

4.6.1 Initial Rigid Alignment

The goal of this processing step is to remove translational and rotational differences between the samples through rigid transformations. During μCT scanning the specimens are placed in sample-holders, but the individual orientation is random and unknown. Many types of data processing are typically simplified by defining and placing the data in a consistent position and direction in a global coordinate system.

The approach taken here is also described in Contributions A and B. It is based on the extracted surface models, but can be applied to the volumetric data as well. The idea is simply to fit an ellipsoid to the points of each surface mesh and align the axises (the mathematical formulation is very similar to the principal component analysis described later in Section 6.1.1). In principle, any type of rigid alignment procedure could be used instead of the one suggested here.

The translational differences were removed by moving the inner ear center of mass to the origin of the coordinate system. This simple operation is accomplished by subtracting the mean vertex position from all vertices. Let $\mathbf{\Sigma}_i$ be the 3×3 covariance matrix of the mesh vertex positions of the i'th dataset (after the translation). The eigenvectors, $\mathbf{W}_i$, of $\mathbf{\Sigma}_i$ provides a rotation matrix, which when applied transforms the surface points to the principal component directions (equivalent to the major axises of the fitted ellipsoid). The procedure and result is illustrated in Figure 4.5.

This alignment procedure is robust due to the asymmetry of the inner ear shape. However, the sign of a principal direction among the datasets could potentially be opposite to each other. To handle this, one dataset is defined as the reference, and it should then be ensured that all other (moving) datasets are properly oriented compared to it by performing a simple check procedure. The bounding box (following the axis directions) of the reference and of the i'th moving point cloud is split up into a coarse $6 \times 2 \times 2$ grid of voxels. The vertex density, ρ_j^i, within the j'th voxel is calculated, as the number of vertices over the voxel volume. The sum of squared grid vertex-density differences can be used as a check metric,

$$\sum_j \left(\rho_j^i - \rho_j^{\mathrm{Ref}}\right)^2 \tag{4.3}$$

It is tested if an axis-flip of the moving dataset would result in a lower metric value, and if that is the case the flip is kept. While there is no guarantee for this to work in all cases, it has worked flawlessly for our limited number of μCT data samples.

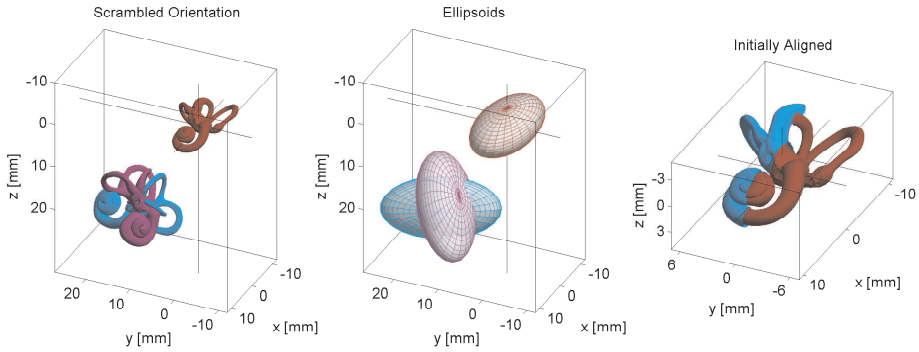


Figure 4.5: Initial Alignment Procedure. Ellipsoids are fitted to the surface models, which have random orientations and placement to begin with. Using the ellipsoid axises for aligning the inner ears works well due to the asymmetric shape.

4.6.1.1 Alternative Procedure

The above described initialization procedure cannot be used for all types of data and images, as it requires a segmentation (to obtain the surface model) representing the full inner ear.

We set-up and rely on the following alternative procedure, when the ellipsoid fitting is not a possibility. It is a two-step approach starting with a landmark-

based transformation followed by a rigid image registration (details about image registration models are given in Chapter 5). Four landmarks are manually placed as illustrated in Figure 4.6. There is a landmark at the round window, one at the modiolus center in the basal turn, one at the apical turn center (coinciding with the central axis) and finally the point on the cochlea inner wall intersecting the line from the round window to the basal turn modiolus center.

These landmarks can be reliable placed even in clinical CT images [WVW+14], which makes the procedure versatile and generically applicable. Accurate placement of the landmarks is not essential, as the follow-up refining rigid image registration can handle and compensate for smaller inaccuracies in the landmark positions.

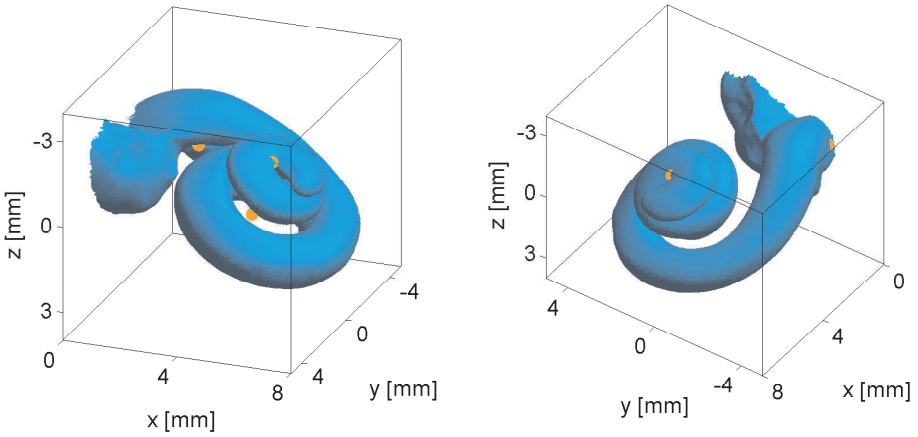


Figure 4.6: Landmark-based Initial Alignment. An alternative initialization procedure based on four manual landmarks (orange points) can be used in cases where the data sample is not segmented or if the semi-circular canals are not present.

To summarize, two rigid initialization procedures have been introduced,

- One for aligning segmented ex-vivo μCT samples of the inner ear, which is based on ellipsoid fitting.
- The other relies on manually placed landmarks and is applicable to the remaining types of CT/μCT images.

4.6.2 Morphological Characterization

Before attempting to do more advanced modelling of the shapes, it is sensible to establish a basic understanding of the inner ear anatomy and variability in the acquired data. It is important to consider if our data is representative of the normal anatomy. In order to explore this, we look into how the inner ear anatomy is normally characterized by clinicians and anatomists.

In Section 4.6.2.1 we present previous publications describing inner ear morphology using approaches and data similar to ours. We then measure the morphology of our data samples for comparison, treating the cochlear and semicircular canals (SCC) separately in Section 4.6.2.2 and 4.6.2.3 respectively.

4.6.2.1 Previous Work

The morphology of the inner ear has been studied for decades. With the development of better and more advanced imaging techniques the measurements have generally become more sophisticated, accurate and precise. Any CT-based measurement is simply much more uncertain than if measured on μCT. Morphological descriptions in older studies also tend to follow varying measurement definitions. The goal of Verbist et al. [VSC$^+$10] was to establish a consensus on how to define a cochlear coordinate system and have consistent morphological definitions to some extent. Studies published after 2010 therefore tend to define morphometrics more consistently, which makes the reported number easier to compare. Even though there are plenty of good studies of cochlear morphology in the literature, we confine ourselves within this thesis to focus only on newer publications where the imaging and data processing strategies have a degree of similarity to ours - i.e. measurements based on μCT scans of humans specimens.

Lee et al. and Shin et al., 2013 [LSK$^+$13, SLK$^+$13]: These two publications from a Korean research group study the morphology of respectively the cochlea and the semi-circular canals (SCC) based on 39/40 formalin-fixed human temporal bones scanned with μCT (voxel size of 35 μm). Intra-cochlear anatomical details were hardly visible in the images, but whether this is due to the formalin fixation or the imaging procedure is not clear. The data is processed to obtain 3D surface models - presumably using a threshold segmentation with some post-processing steps. Several different morphometrics are measured manually on the surfaces, and the numbers are given in Table 4.2 and 4.3 respectively. A statistical analysis comparing the cochlear sizes of male and female subjects showed no significant difference between genders.

Avci et al., 2014 [ANL$^+$14]: This study is based on 16 formaldehyde fixed samples, which were μCT scanned and reconstructed with varying isotropic voxel sizes ranging from 8-36 μm. The contrast between intra-cochlear structures were strongly diminished due to retained fluid in the cochlea, similar to our experiences with the thiel-fixed samples. The cochlear morphometrics were measured manually from the image data and the numbers are stated in Table 4.2.

4.6.2.2 Cochlear Morphology

From the consensus of [VSC$^+$10] the morpho-metrics describing the overall cochlear shape can be defined as follows (see Figure 4.7),

- **Cochlear Length** is the distance from the center of the round window membrane to the cochlea outer wall following a line tangential to basal turn plane going through the central axis.
- **Cochlear Width** is the outer wall to outer wall distance in the basal turn following a line orthogonal to both the length and central axis.
- **Cochlear Height** is the distance from the bottom of the basal turn to the center of the apical turn following the cochlea central axis.

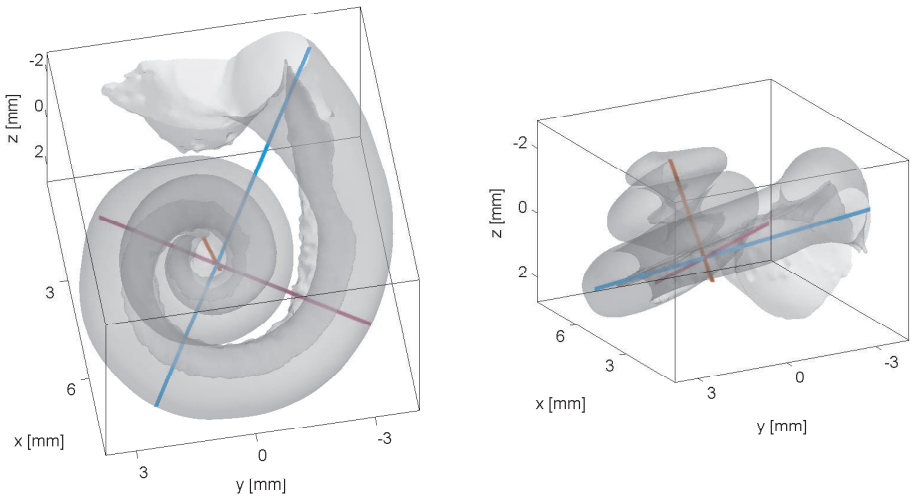


Figure 4.7: Measuring Cochlear Dimensions. The cochlea length (blue), width (purple) and height (red) is manually measured following the definitions of [VSC$^+$10].

We have manually placed the points allowing us to measure the cochlear morphometrics in both the dried and the thiel datasets. The metrics for the TUM sample were measured in [BBS12]. The first order statistics are given in Table 4.2. The measured morphometrics are in good agreement with the values from the literature.

Table 4.2: Comparison of Cochlear Morphology

Study	[SLK+13]	[ANL+14]	TUM	Dried	Thiel
Sample Size	39	16	1	18	14
Cochlear Length [mm]	9.7 ± 0.3	9.2 ± 0.4	-	9.6 ± 0.5	9.3 ± 0.4
Cochlear Width [mm]	7.0 ± 0.3	7.0 ± 0.3	7.7	7.0 ± 0.3	6.6 ± 0.3
Cochlear Height [mm]	3.8 ± 0.2	4.4 ± 0.3	4.1	4.0 ± 0.3	3.9 ± 0.2

4.6.2.3 Semicircular Canals Morphology

There is no consensus on how to define and measure the morphology of the semicircular canals (SCCs). Typically the radius, thickness and angles between the SCCs are reported.

Manually measuring the morphometrics is tedious work and the results inherently have some inaccuracy and noise. For the cochlear case described above we had no better solution. However, for the SCC morphology we have developed a more unbiased approach. Our strategy is to estimate the best torus to each canal, which is accomplished in two steps.

Step I: The region of each SCCs is marked in the surface models. Although this is manual work, it does not need to be very precise. The marked regions are shown in Figure 4.8. Note, that for simplicity we do not mark vertices in the common crus, and since the ampullae ends of the SCCs can be distinctively different from a torus-shape they are not marked either.

Step II: Given a 3D point cloud, we wish to estimate the parameters, μ, of the

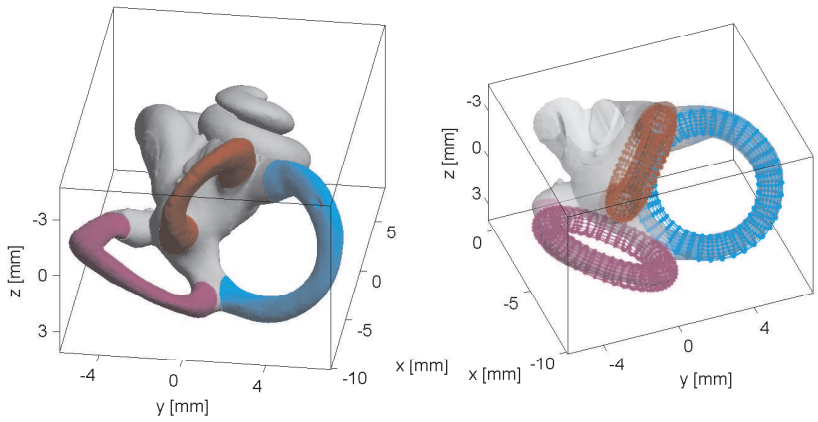


Figure 4.8: SCC Morphology by Torus Fitting. The vertices of the SCCs are marked manually (left) and the best torus fit is estimated (right).

torus that gives the closest fit in a least squares sense:

$$\arg\min_{\mu} \sum_{i} d(p_i, \mathbb{S}(\mu)) \tag{4.4}$$

where $d(p_i, \mathbb{S}(\mu))$ is the distance between the point, p_i, and the torus, $\mathbb{S}(\mu)$, described by the following parameter vector.

$$\mu = (x_0, y_0, z_0, n_x, n_y, n_z, R, s) \tag{4.5}$$

The generic torus shape is illustrated in Figure 4.9 and is defined by a center point, (x_0, y_0, z_0), a torus plane normal, (n_x, n_y, n_z), the torus plane radius, R, and finally the torus cross-sectional radius (thickness), s.

The mathematics of fitting a torus shape to a point cloud is described in [LMM98, Sha98]. The solution for Eq. 4.4 can be found numerically as an optimization problem. A good initial guess of the parameters, μ, is always required to ensure convergence to the correct solution.

Initial parameter guess: The torus size parameters (R and s) can be chosen as the average values from the literature on SCC anatomy [LSK$^+$13]. The positional parameters (the torus normal and center point) are more tricky to initialize, as they are dependent on the chosen coordinate system, position and orientation of the data. However, in our case we have already rigidly aligned all of our samples to a common space (see Section 4.6.1). The initial center

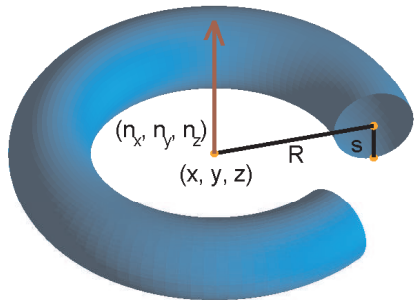


Figure 4.9: Torus definition. A torus described by two parameters controlling size and six for position and orientation.

point is simply estimated from the SCC center of the mass (COM). However, as the shape is not a complete torus, the COM is shifted away from the vestibule. We compensate for this by moving the COM closer to the coordinate system origin by a factor that we experimentally found to work well. Finally, we fit an ellipsoid to each SCC point cloud. The third and smallest axis gives a decent initial estimate for the torus plane normal.

The results of our SCC estimates are given in Table 4.4. Statistics can be made almost directly on the estimated tori parameters. To study the angular relationships between SCCs, we can simply calculate the angles between two tori normals,

$$\cos\theta = \vec{n_1} \cdot \vec{n_2} \tag{4.6}$$

where the vectors are assumed to have unit length.

The torus radius and the angles between the canals are in good agreement. The estimate of thickness seems consistently smaller than the values from the literature. This could simply be influenced by the differences in voxel resolution and the segmentations of the data.

In summary,

- We have made ground truth measurements of simple cochlear and SCC morphometrics for our datasets. These can be used for validations later on, and suggest that our data is morphologically normal.
- We have presented a novel and unbiased method for estimating SCC morphology based on torus fitting. This approach has to our knowledge not been attempted before.

Table 4.3: SCC morphometric statistics from [LSK+13]

Metric	Statistic	aSCC	lSCC	pSCC
Radius, R	$\mu \pm \sigma$ [mm]	3.6 ± 0.2	2.9 ± 0.3	3.6 ± 0.2
Thickness, s	$\mu \pm \sigma$ [mm]	1.0	1.5	1.2
Angle, θ	$\angle_{\text{ASCC}}$ $\mu \pm \sigma$ [°]	-	84.4 ± 4.2	92.1 ± 3.5
	$\angle_{\text{LSCC}}$ $\mu \pm \sigma$ [°]	84.4 ± 4.2	-	86.2 ± 2.4
	$\angle_{\text{PSCC}}$ $\mu \pm \sigma$ [°]	92.1 ± 3.5	86.2 ± 2.4	-

Table 4.4: SCC morphometric statistics of the dried samples

Metric	Statistic	aSCC	lSCC	pSCC
Radius, R	$\mu \pm \sigma$ [mm]	3.5 ± 0.3	2.6 ± 0.2	3.3 ± 0.4
Thickness, s	$\mu \pm \sigma$ [mm]	0.6 ± 0.1	0.8 ± <0.1	0.7 ± 0.1
Angle, θ	$\angle_{\text{ASCC}}$ $\mu \pm \sigma$ [°]	-	84.5 ± 5.8	90.6 ± 3.1
	$\angle_{\text{LSCC}}$ $\mu \pm \sigma$ [°]	84.5 ± 5.8	-	91.5 ± 4.7
	$\angle_{\text{PSCC}}$ $\mu \pm \sigma$ [°]	90.6 ± 3.1	91.5 ± 4.7	-

4.6.3 Cochlear Skeletonization

Comparing the shape of two cochlear spirals quantitatively is not immediately an easy task, and there is not necessarily one unique description or solution to use, as it depends on both the intended application and the desired accuracy. For example, in order to use the Greenwood function (Eq. 3.1) it is sufficient to describe and compare the cochlear using only a 1D space-curve representing the distance from the apex. However, when working with the design and function of the CI electrode array, the exact surface 3D geometry has to be considered.

In this section we briefly present one of our approaches for describing the shape of the cochlea. The methodology was one of our contributions presented in Paper A.

From the starting point it is assumed that a surface model, $\mathbb{S}$, of the bony labyrinth is given (see Section 4.5). Additionally we manually define the apex point, $\vec{A}$, the central axis direction, $\vec{n}$, and the edge of the lamina spiralis as a point set, $\mathcal{Z}^{\text{spiral}}$, consisting of approximately 100 points going in an ordered direction between the round window and the apex (see Figure 4.10, left). Using this information we can make a cochlea skeleton structure, from which we

can define a naive parameterization of the cochlear spiral described as a set of corresponding pseudo-landmarks, $\mathcal{Z}^{\mathrm{LM}}$.

For the i'th dataset we first create a parametric description of the cochlea skeleton using $\mathcal{Z}_i^{\mathrm{spiral}}$, from which we determine 37 corresponding positions on the skeleton with equal arc-length, $\mathcal{Z}_i^{\mathrm{S}}$. Secondly, we extract planar surface cross-section at each of the 37 points, p, in $\mathcal{Z}_i^{\mathrm{S}}$. The cross-section plane is determined by the tangent of the skeleton at p. Each cross-section of the surface mesh is then parameterized using 40 points. The central axis direction, $\vec{n}_i$, provides the reference for orientation and starting point in each cross-section parameterization. These cross-sectional points together with the apex landmark provides a set, $\mathcal{Z}_i^{\mathrm{LM}}$, of 1481 corresponding surface pseudo-landmarks. The process and the result is illustrated in Figure 4.10. Note that establishing pseudo landmarks in the apical region of the cochlear can lead to some conflicting ambiguity, as the cross-section can intersect one another. To avoid problems from this, the most apical part was not included.

There is a certain amount of noise and inaccuracy in this approach of modelling the cochlear shape due to the manually defined input. It is a rather ad-hoc procedure, developed as other approaches of skeletonization to our experience did not immediately work on our data samples. The biggest weakness is arguably the lack of pseudo-landmark placement in the apical region.

In summary,

- The cochlear spiral of the dried dataset samples (Table 4.1) have been parameterized with a point set, $\mathcal{Z}^{\mathrm{LM}}$, of 1481 corresponding pseudo-landmarks.

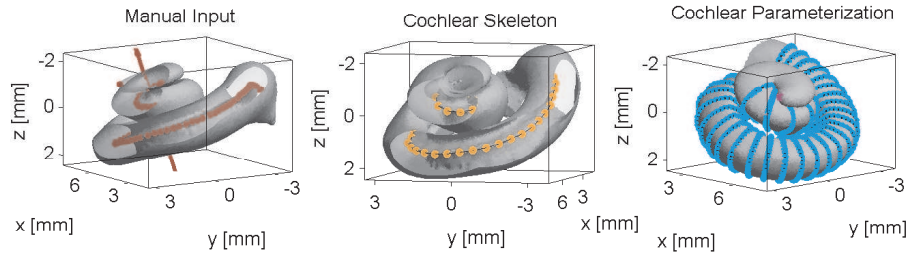


Figure 4.10: Cochlear Parameterization. Starting from a cochlear surface model with the apex, central axis and the lamina spiralis edge defined (left), a curved cochlear skeleton is extracted (middle). Cross-sections of the surface are parameterized from the skeleton, to provide a description of a cochlea shape based on corresponding pseudo-landmarks (right).

4.6.4 Heat Distribution of the Cochlear

There is another less naive way to define a parameterization of the cochlear spiral, than the skeletonization approach described above. The methodology is based on calculating the heat propagation through the cochlea. It was introduced as one of our contributions in Paper B and it is briefly recapped here.

Starting from a segmentation, $\mathbf{L}$, of the bony labyrinth (see Section 4.4), we wish to describe the heat propagation through the segmented object going from a source point manually defined in the cochlear apex to the sink placed in the coordinate system origin (0,0,0) (note, that the data was aligned in a common coordinate system as described in Section 4.6.1). Fixing the heat values at the two extremes, the heat distribution throughout the region can be solved as a steady state solution of a partial differential equation. A volume, $\mathbf{H}$, corresponding to the segmentation is generated, but where each voxel of the bony labyrinth class is assigned a heat value, h,

$$\mathbf{H} = (h_1, h_2, \ldots, h_N) \ , \{h_i \in \mathbb{R} \mid 0 \leq h_i \leq 1\} \tag{4.7}$$

Intuitively, the assigned heat intensity is related to the relative distance to the apex. Illustrations of the heat distributions in both the volumetric data and on the surface models are given in Figure 4.11.

This approach is arguably more versatile than the pseudo-landmark description based on skeletons (See Section 4.6.3). The main drawback of the procedure is that it cannot properly handle a region representing the entire inner ear. Propagating heat through the semi-circular canals has to be handled differently.

In summary,

- A map of the heat propagation, $\mathbf{H}$, through the cochlear spiral has been calculated for the samples belonging to the dried dataset and the TUM sample (Table 4.1).

4.7 Summary and Conclusion

Scanning samples with μCT is a great source of detailed anatomical information. Acquiring the samples is however a long and hard process. The number of available samples is limited, and it is therefore crucial to make the most of it. We have in this chapter presented a variety processing steps, so that we have high quality representations of the data available.

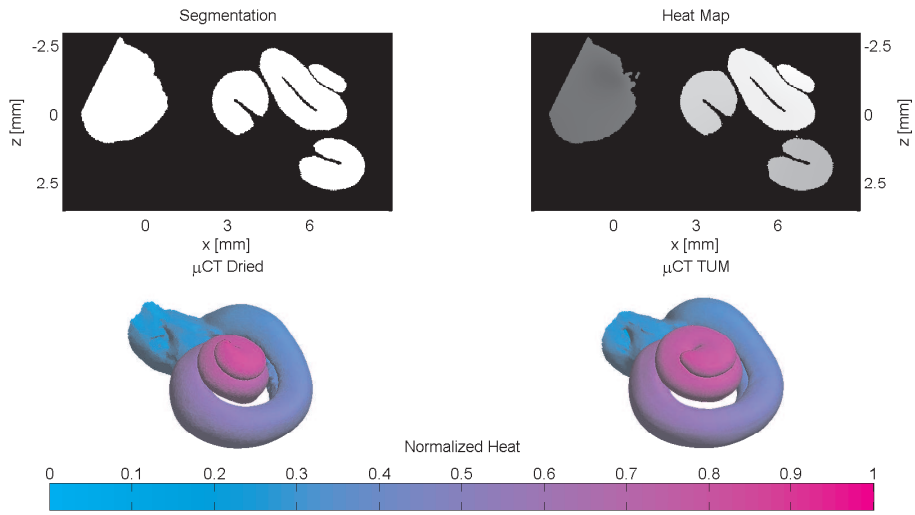


Figure 4.11: Cochlear Heat Maps. The segmentation of the bony labyrinth (upper left) is enhanced with an intensity value relating to the distance from the apex (upper right). Two example heat distributions shown on surface models (bottom).

- Table 4.1 provides an overview of the available datasets used in the subsequent chapters.
- Clinical CT images provide a rough representation of the inner ear, but it is the only data source for living CI-recipients.
- μCT images of the inner ear can only be acquired ex-vivo, but contain valuable anatomical information, such as intra-cochlear structures.
- All datasets have been rigidly aligned in a common global coordinate system.
- Segmentations, $\mathbf{L}$, of the dried dataset and the TUM sample have been obtained, and corresponding 3D triangulated surface models, $\mathbb{S}$, extracted. The cochlear of these samples have been naively parameterized using a set of pseudo-landmarks, $\mathcal{Z}^{\mathrm{LM}}$, and using a heat distribution, $\mathbf{H}$.
- The morphometrics of the cochlear and the semicircular canals were measured.
  - The measurements were manually made for the cochlear legth, width and height.
  - The SCCs dimensions were estimated using a novel torus fitting procedure

Chapter 5

Image Registration

Image registration plays a major role in statistical shape modeling with its' ability to create dense correspondences between volumetric data samples in a meaningful way. For this thesis, an effort has been put into the area and almost all of the scientific contributions contain aspects of image registration to some extent. A separate chapter is therefore devoted to introducing the topic and making the connections between the contributions.

The chapter contains two parts. First, a part that covers the underlying basic image registration theory and terminology. The notation introduced here follows the framework of `elastix` [KSM+10] and will appear in other chapters of the thesis. A reader familiar with image registration concepts can lightly read through this initial part, as there is no novel registration theory being presented.

The second part of this chapter introduces why image registration of the inner ear anatomy is challenging and the efforts taken to solve the problem. A description, comparison and evaluation of the different image registration models used for building statistical shape models is given, in order to provide some additional context to the contributions of Papers A, B, C and I.

5.1 Theoretical Framework

The theory provided here is not meant to fully cover all aspects of the broad topic that is image registration. Comprehensive reviews and sources for understanding of the theory and applications could for instance be the following references [MV98, HBHH01, RSH+99]. As the formulations and terminology used here closely follow that of `elastix` [KSM+10], the manual to that software package is naturally also a good source for further information.

Image registration is the task of finding the spatial relationship between two datasets. The registration of the moving dataset, $\mathbf{I_M}$, towards the fixed, $\mathbf{I_F}$, is formulated as a (parametric) transformation, T_μ, where the vector, μ, containing the p-parameters of the transformation model is found as an optimization of the cost function, $\mathcal{C}$, potentially in multiple stages.

$$\hat{\mu} = \arg\min_{\mu} \mathcal{C}(T_\mu, \mathbf{I_F}, \mathbf{I_M}) \tag{5.1}$$

An image, volume or dataset should in this context be considered one and the same thing. That is a 3D voxel grid defined in a common coordinate system by an origin, a voxel spacing and a grid size (Figure 5.1), and where each voxel is associated with an intensity value. Unless clearly specified, it can be assumed that everything in the following concerns 3D data.

As hinted above, a registration model is composed of several main components:

- Transformation Model
- Cost Function
- Multi-resolution Scheme
- Optimization

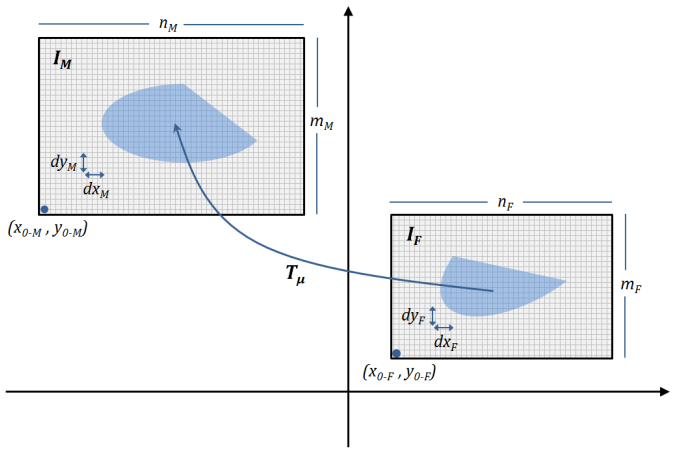


Figure 5.1: Sketch of the basic image registration. Each image is defined by an origin (x_0, y_0), a pixel/voxel spacing (dx, dy) and a size (m, n). The registration model is an optimization problem that finds the transformation that aligns the fixed image to the moving image $(T_\mu : \mathbf{I_F} \rightarrow \mathbf{I_M})$.

5.1.1 Transformation Model

The transformation model, T_μ, defines the type of movement or deformation that is possible and it can either be global or local.

Global: These transformations are characterized by having relatively few parameters, p, they are invertible, and when applied they affect the entire image in a consistent manner. The most basic is the pure *translation* ($p = 3$). Adding rotations yields the *rigid* transform ($p = 6$). The *similarity* transformation ($p = 7$) further allows for isotropic scaling. Finally the *affine* transform ($p = 12$) that also includes shearing.

Local: These transformations are also called deformable, non-rigid or free-from deformations. The main point is that they can affect one region/voxel of the image independently from one another. Different types of local transforms exist, but the thesis is only concerned with the *B-spline* transformation model (illustrated in Figure 5.2). A grid of nodes is defined in the domain of the reference image. Each node has 3 degrees of freedom (3 parameters) to specify the local deformation.

There are plenty of aspects to consider with such a transformation model. First, is the choice of the B-spline grid spacing. A distance too large can yield a poor result, simply by having too few parameters to adapt the structures in the desired manner. If the grid spacing becomes too fine, there is the risk of over-fitting to noise in the data, but the main concern is the huge amount of parameters (especially in 3D) and its effect on the optimization complexity. It increases computational time and perhaps more importantly also the amount of local minima in the cost function, making it difficult to reach the desired solution.

Secondly, as these transforms only have very local support, it is important that the fixed and moving voxel grids are closely aligned to begin with. The desired correspondences or solution should not be too far away. The typical strategy is to initialize the registration with a global transform (see Section 4.6.1) before applying a local non-rigid one.

Thirdly, many local transformations are not invertible, which can have some considerable implications when the registration is part of larger pipeline or framework. In these cases, the choice of fixed and moving image can be important to consider.

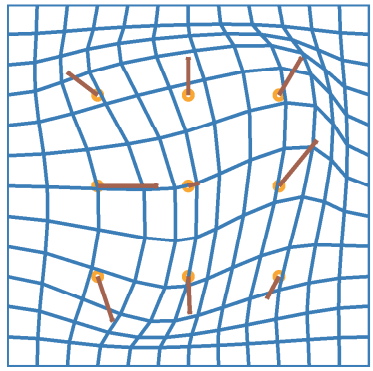


Figure 5.2: The B-spline transformation model. Consider having a deformation vector placed in each of the grid node points, which can vary independently and thereby model local image variation.

5.1.2 Cost function

A typical cost-function can be stated in the following manner,

$$\mathcal{C} = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, \mathbf{I_F}, \mathbf{I_M}) + (1 - \alpha) \cdot \mathcal{P}(\mu) \tag{5.2}$$

where α is a weighting factor between an image similarity term $\mathcal{S}_{\text{Sim}}$ and a regularization term $\mathcal{P}$. Additional terms can be added to the cost function together with extra weighting factors $(\beta, \gamma, ...)$ if appropriate for the problem.

Similarity Metric: The image similarity term is a measure of the goodness of the fit between $\mathbf{I_F}$ and the transformed moving image $\mathbf{I_M}(T_\mu)$. The metric is to be chosen based on the nature of the input images and their histograms (illustrated in Figure 5.3).

The most common and simple metric is the *Sum of Squared Differences* (SSD). It assumes that the image intensities compares directly. The SSD can be used for mono-modal data, but unless the same scanner and acquisition settings were used, the intensities of the images might first require a normalization of some kind. This can generally be handled by using the *Normalized Correlation Coefficient* (NCC) instead. This metric assumes only a linear relationship between intensities, making it less strict and generally suitable for images that have the same visual appearance (for instance CT-μCT registration). In multi-modal cases where the intensity distributions and the image appearances are not immediately similar, one should use a more general applicable similarity metric based on *Mutual Information* (MI) [MCV+97]. Finally, in the special case of a registration between binary volumes (segmentations) one can use the *Kappa Statistic* (KS) instead of the SSD.

Regularization: Regularizing terms are added to the cost-function in order to constrain or guide the optimization and can be based on physical constraints [RMBJ03], shape priors [BvdHL+13] or basically anything that fits the problem. Regularization is rarely required for global transformation models, but it is very common with local transforms. The high number of parameters typical for a deformable transformation provides so much freedom, that some restriction is often needed to efficiently find a good solution.

Two types of regularization play an important role in the work of this thesis. First, there is the general applicable Bending Energy regularization, $\mathcal{P}_{\text{BE}}(\mu)$, which restricts the transformation parameters to prevent fold-over and extreme local deformations [RSH+99]. Secondly, there is the use of anatomical landmarks, $\mathcal{S}_{\text{CP}}(\mu, \mathcal{Z}_F, \mathcal{Z}_M)$, where corresponding points sets are provided for both

the fixed, $\mathcal{Z}_F$, and moving, $\mathcal{Z}_M$, datasets [BSL+11]. The Euclidean distance between two landmarks is sought to be minimized, thereby acting as a regularization or similarity term to the registration.

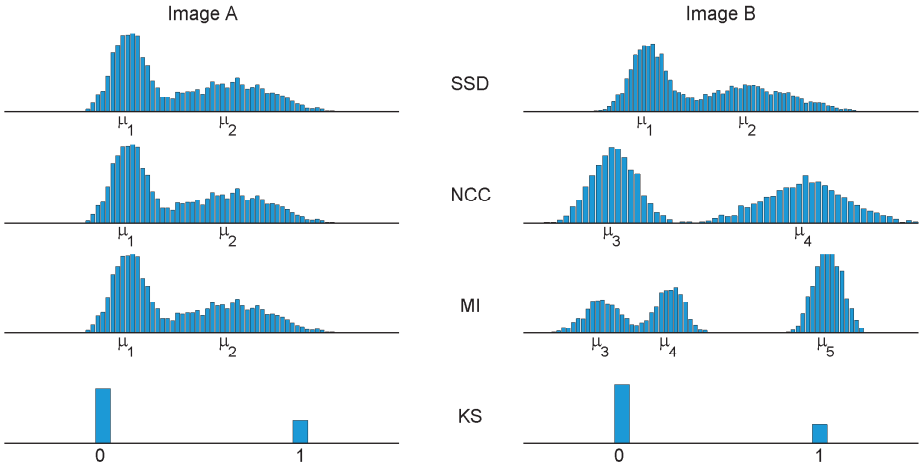


Figure 5.3: Image Histograms and Similarity Metrics. The appropriate similarity metric depends on the histograms of the data. (Top row) Image intensities of A and B compare directly. (2nd row) Linear relationship between the intensity distributions. (3rd row) No apparent relation between distributions can be seen. (4th row) Special case with binary volumes - i.e. segmentations or labeled images.

5.1.3 Multi Resolution Scheme

The point of the performing registrations in multiple stages is to start coarsely and then gradually refine the procedure. This is sometimes referred to as a pyramid. It aids the optimization both with regards to the amount of parameters included in the transforms (transformation complexity), and with regards to the level of detail in the images themselves (data complexity). It is a valuable strategy for avoiding local minima in the cost function, and for having a more computationally efficient registration [LA99].

Transformation complexity: The idea is to gradually increase the amount of transform parameters for each level. The first thing to do is initialization with a global transform. However, this will in some cases not be considered a

stage of the pyramid, but rather a stand-alone registration (see Section 4.6.1). In any case, to reduce the transformation complexity for a B-spline model the grid spacing should gradually be reduced for each resolution instead of running with the finest spacing all the time.

Data complexity: By using downsampling and/or smoothing of the images certain high frequency features are blurred away. This makes it easier to establish an optimum of the similarity metric. It is most commonly done with a Gaussian kernel where the width is gradually decreased for each resolution.

Note, that within the multi-resolution component it is also possible to change other parts of the registration model for each resolution. This could for instance be the weightings of the cost function terms (Equation 5.2), so a stronger emphasis is put on a certain behavior at the different stages.

5.1.4 Optimization

This component is concerned with the task of finding the minima of the cost function. Optimization is a research field in it-self. The challenge is often finding a balance between speed and accuracy.

There are many different solvers, which can have various settings to fine-tune in order to get a good result [KSP07]. To avoid this additional complexity, all registration models in the context of this thesis are solved with the Adaptive Stochastic Gradient Descent [KPSV09]. This is the recommended solver within the `elastix` framework, and it has a few additional settings that control the convergence rate. However, there is an option provided for automatic estimation of these settings, which was experienced to work well.

However, some parts of the optimization is worth controlling and tweaking. This would be the maximum number of iterations per resolution, the subsampling strategies and the use of sampling masks as these can notably affect the computation times and the accuracy of the optimization. It is not required that the cost function is evaluated for all voxels for each iteration. It is sufficient and faster to rely on a subset of samples, that can be randomly re-drawn every iteration. Further, it is possible to spatially focus the drawn samples to specific regions of the volume by supplying a binary mask of the fixed image.

5.1.5 Transformation Direction

There is one additional aspect worth noting. In the `elastix` framework a transformation is defined as going from the fixed image domain to the moving ($T_\mu : \mathbf{I_F} \rightarrow \mathbf{I_M}$).

This might sound counter intuitive, as the tentative description of a registration model is a procedure, which aligns or deforms the moving image to the fixed image. The reason for this choice of direction, is to be found in the operation of applying a transformation to the moving image. To generate $\mathbf{I_M}(T_\mu)$ a new 'empty' voxel grid is initialized with (typically, but not required!) the dimensions of the fixed image. Each voxel, x_i, should then be assigned an intensity value. The coordinate of the voxel is transformed ($x_i(T_\mu)$) and if it lands up in the domain of the moving image, then it can be assigned an intensity value based on an interpolation of the surrounding voxels. The moving image is in practice *re-sampled* in the domain of the fixed image, and it therefore has a computational advantage to have the transformation direction specified the way it is.

For global transformation models this notion of transform direction hardly matters as they are invertible. However, the implications are rather notable for the non-diffeomorphic B-spline model and how the transformations play into the larger framework of shape modeling (addressed in Chapter 6).

Allowed operations: The following B-spline transformations are well defined once the registration problem (Eq. 5.1) has been solved,

- Calculating $\mathbf{I_M}(T_\mu)$: It is straightforward to deform the moving image and/or the corresponding segmentation. This is simply the re-sampling of $\mathbf{I_M}$ in the domain of the fixed image.
- Transforming $\mathcal{Z}_F(T_\mu)$ or $\mathbb{S}_F(T_\mu)$: Let $\mathcal{Z}_F$ and $\mathbb{S}_F$ be respectively a point set and a surface model defined in the reference domain (see Section 4.5). Applying the resulting transform moves or deforms the points or vertices directly to the domain of the moving image. It is the same underlying process as for calculating $\mathbf{I_M}(T_\mu)$, except that the points are not necessarily voxel centers, and the part about interpolating values and the re-sampling of a new image is skipped.

Problematic operations: The following transformations can however not immediately be carried out,

- Calculating $\mathbf{I_F}(T_\mu^{-1})$: This is to be understood as the operation of re-sampling the fixed image in the domain of the moving image. However, there is a potential workaround for doing the operation on a segmentation

or labeled volume, $\mathbf{L}$ (see Section 4.4). Having extracted the surface model, $\mathbb{S}_F$, of the fixed dataset, this can be transformed instead (see above). It is then possible to make a discretization of the closed surface model, to turn it into a binary volume in the desired moving voxel-grid domain. However, it is still not possible to project the image intensity values in this manner.

- Transforming $\mathcal{Z}_M(T_\mu^{-1})$: It is not immediately possible to deform a point set $\mathcal{Z}_M$ nor a surface model defined in the moving image to the fixed image domain. The operation also has a workaround based on applying the transformation to a segmentation or binary image $\mathbf{L_M}$ (see above). After the deformation the surface model can be reconstructed (see Section 4.4). The draw-back is that the surface mesh model is regenerated with a different mesh structure and with new vertices. Any potential point correspondence to other surface models is therefore lost in the process.

5.2 Registration Models and the Inner Ear

During the course of the project several different registration models were developed and used for working with the anatomy of the inner ear/cochlea. Most of these models are presented separately in different contributions, which makes them cumbersome to compare and blurs out the nuances of the differences between them. This part of the chapter will summaries and clarify on the dissimilarities between the models and provide insights and discussions about how and why they came to be as they are.

The development of the registration models contains an evolutionary aspect. After having built a starting basic model which serves as a baseline, then several iterations of improvements were made to overcome certain limitations and issues or to simply just achieve better results based on the acquired experiences.

Before comparing the registration models in Section 5.2.3, it is however sensible to set the stage first. Section 5.2.1 reviews previous work concerning registration of the inner ear. Secondly, setting up proper registration models for inner ear and cochlea anatomy from μCT data is not straightforward. Section 5.2.2 provides details about the issues, and illustrates why the problem is worth researching more thoroughly.

5.2.1 Previous Work

There is an immense body of literature concerning image registration in general. However, very little work concentrates specifically on the inner ear or cochlea.

There are examples of some simpler rigid registrations based on manual landmark transforms or manually defined orientations [LSK+13, ANL+14, Wan13] to study cochlear morphometrics from μCT data. However, neither of these studies compare to the non-rigid registration models in focus here.

Noble et al., 2011 [NLMD11]: This study presents one comparative registration procedure. With the goal of building a statistical shape model (SSM) from 6 μCT datasets (36 micron resolution) segmentations of the scala- tympani and vestibuli were obtained to represent a model of the cochlear anatomy. To create point correspondences between the structures a registration model was set-up consisting of an affine initialization plus a non-rigid transform based on the adaptive-bases algorithm [RAD03]. However, no additional settings nor

evaluation of the registration accuracy (μCT-μCT) were presented. Further details concerning this study is more relevant to cover in Chapter 6.

The main point to extract from the presented studies on image registration of the inner ear or cochlea,

- Symptomatic for all of them, the registration is mostly used as a tool in order to get to the intended applications. At least, very little consideration is presented concerning what constitutes a good registration model for this particular the anatomy, and how the (in)accuracy of the procedure might affect the later obtained results.

5.2.2 General Consideration and Challenges

The present section is meant to convey some insights for setting up good registration models, based on the available data and acquired experiences during the project.

Samples and Image Quality: First and foremost it is quite important to understand and consider the type and quality of the image data. The registration models that we develop is based on the dried dataset and the TUM sample (see Section 4.1 and Table 4.1). These images have various characteristics due to either abnormal anatomy, poor sample preparation or imaging, which need to be considered:

- The openings into the RW and OW are open. The soft tissue membranes are simply not present in dried samples, meaning that there is no boundary distinction between the middle ear cavity and the inner ear (illustrated in Figure 5.4, left).
- Boundaries at the cochlear interface to the modiolus or internal auditory canal are in some cases very weak or non-existent (Figure 5.4, right)
- The gradient strength of the spiral lamina boundary can vary significantly. In some samples and regions the interface to the scalae is not even visible, but whether this is due to abnormal anatomy, sample preparation or imaging is not known.
- There are occasional cases of artefacts in the data. An example of this is illustrated in Figure 5.4, left, where a high intensity artefact (the stapes!) can be seen in the vestibule.

The above mentioned issues have the thing in common, that they can be handled by working on the segmented volumes rather than the intensity data. In the segmentation process the a priori knowledge about what is (ab)normal and interesting is in some sense put into the data. Using binary segmentations is a great way of simplifying the registration problem. Weak image gradients are made strong, noise in the images is discarded to a large extent and the registration is focused on the anatomical region of interest (see Figure 5.4). For instance, by registering segmentations the anatomical differences in surrounding bone is ignored.

The drawback is the demand for segmentations of good and consistent quality, which can be very time-consuming to acquire for μCT samples. When the intended application of the registrations is to build shape models, then the large time overhead can be justified. Having the best possible training data and registrations will arguably result in more accurate shape models.

In fact, the registrations presented in this chapter are meant for building shape models, and the gray-scale intensity μCT data is never involved in this process.

Normal Anatomy: It is not only the quality of the samples and images that affect the registration models. The normal anatomy of the inner ear it-self gives rise to certain considerations.

The samples contain irregular or noisy anatomical features, which are not particular interesting to include in the shape models. Examples of this could be dissimilarities in the vascular entrances to the inner ear or an uneven bone interface in the SCCs (Figure 5.5). Some irregularities could be handled by segmentation as explained above. In other cases, some image smoothing and making sure that B-spline grid spacing does not become too fine is sufficient to prevent an effect on the registrations.

The anatomy of the inner ear generally lacks distinct features and is highly self-similar. Throughout the basal- and middle turn in particular, the cochlear cross-sections looks almost the same. Similarly, any cross-section of a SCC is approximately ellipsoidal. It is therefore easily imaginable that the registration cost function is riddled with local minima, as any two small regions can easily match almost perfectly with little deformation. A good registration model should therefore rely on a good multi-resolution scheme to ensure that correct regions are matched properly.

Staying on the topic of the multi-resolution component. The μCT data contain some very thin and fine structures - namely the lamina spiralis and the inter-

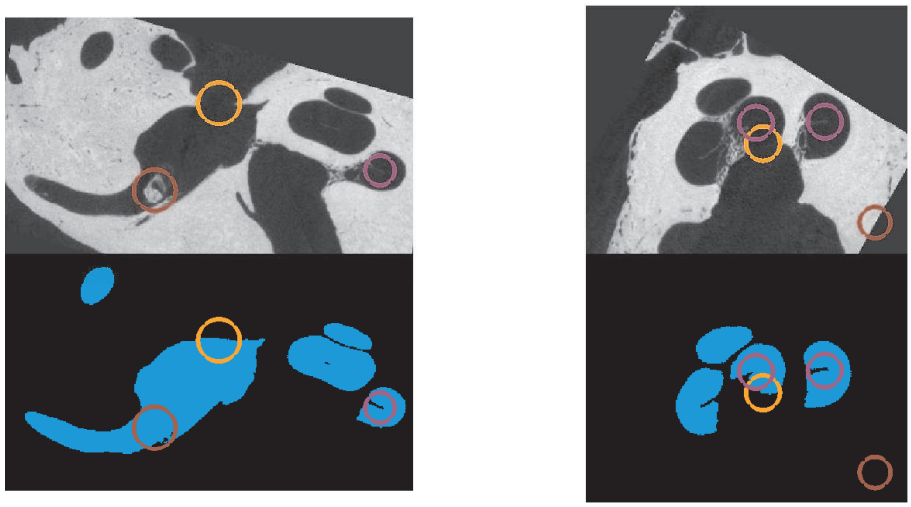


Figure 5.4: Data Quality Issues Concerning Image Registration The dried μCT specimens are difficult datasets to register. Many issues stemming from 'poor' data quality can be ignored with the segmentations of the samples. Orange annotations show lacking anatomical borders, purple circles the varying degree of contrast of spiral lamina (from decent to not present) and the red annotations artefacts that normally would not be there.

Figure 5.5: Variations in the Normal Anatomy. Some dissimilarities between samples are difficult to handle by segmentation and are not worth modeling.

scalae septa. Using smoothing or downsampling in the different resolutions in order to reduce the data complexity should be thought through. These small structures can easily be blurred away. Along the same line of thinking, it is important to consider what voxel resolution is even required to represent these structures properly.

The anatomical variability of the inner ear also plays a crucial part. The semi-circular canals can vary quite extensively. Further, there is the variability in the cochlear turning to consider - in particular with regards to the apical turn. It requires rather extensive deformations to cope with the differences in the apical region, but only selectively following the path of spiral. Deformations across the interscala septa is a no go. The first thing to consider is whether the B-spline transformation is the best choice. The cubic grid of the model has certain limitations with regard to the rotations and the twisting needed for this anatomy. Nevertheless, even with B-spline model it becomes important to consider the cost function and how much regularization is required. Generally a higher α (Equation 5.2) will increase the flexibility required for capturing turning and SCC variability. However, too much flexibility can get the optimization stuck in some undesirable local minima (exemplified in Figure 5.6). Considering the trade-off between flexibility and regularization in conjunction with multi-resolution component becomes very important for the registration model.

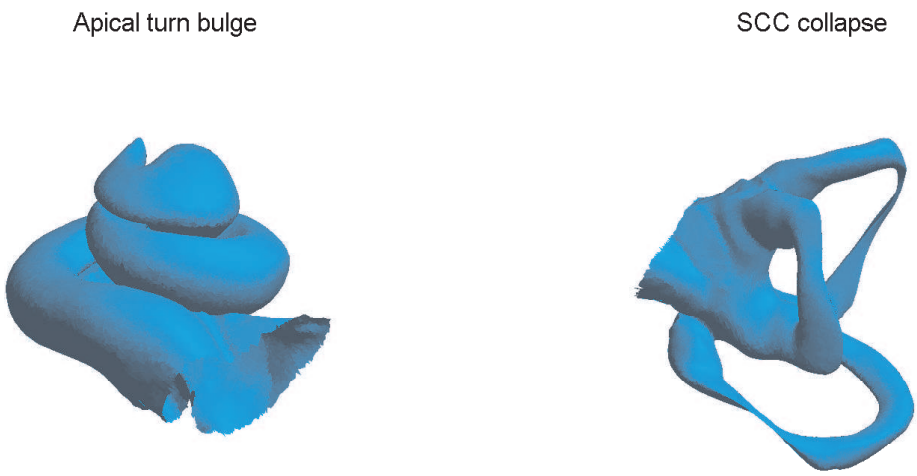


Figure 5.6: Parameter Tuning of α-weight. Too much flexibility (or too little regularization) in a registration model can have unforeseen consequences.

Finally there is the challenge of the compact cochlear spiral. Before applying a registration with a local transformation model, the fixed and moving image should preferably have a good alignment, which is typically based on a global transform. However, the standard global models tend to struggle with the inner ear anatomy due to the compacted spiral shape (illustrated in Figure 5.7). Two inner ears of roughly the same overall size are shown, but where one has <2.5 turns the other spins ≈ 2.9 turns. This is a large difference in the global cochlear shape, and therefore something that is normally assumed to be handled by a global transform. Unsurprisingly and as observed, neither of the standard methods can really cope with the turning discrepancy and fail to bring to the two shapes closer together.

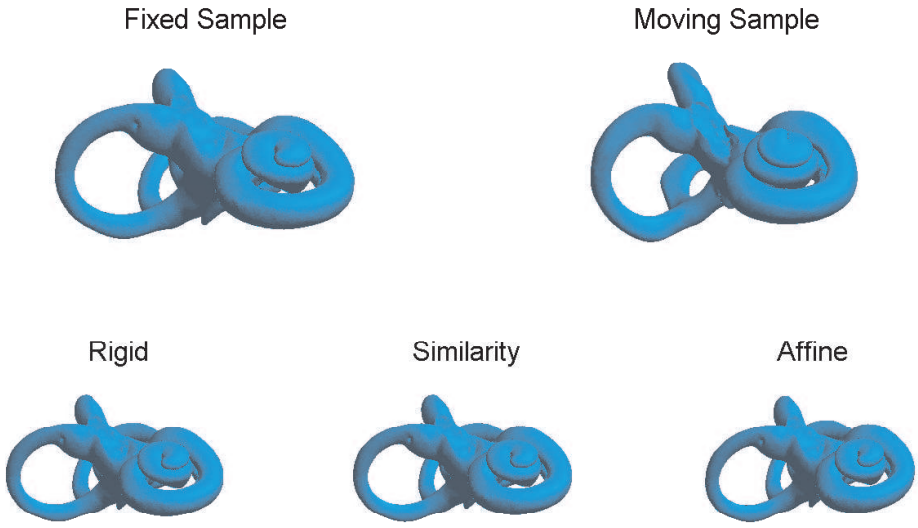


Figure 5.7: Global Transform and the Cochlear Spiral. Global transformations are too global to handle differences in the cochlear spirals.

Performance Validation: Characterization of the accuracy and precision is naturally an important part of any registration model. The validation depends on what type of ground truth knowledge is available. In the registration scenarios presented here we have both a ground truth segmentation, $\mathbf{L}$, and surface model,$\mathbb{S}$, of each dataset. It is then possible to calculate a Dice Score [Dic45] from the segmentations and also the mean and maximum (Hausdorff distance) surface errors. As these scores are very generic we further supplement with a metric based on how accurately the apex is registered.

The definitions of the evaluation metrics are stated here for the sake of clarity

and easy referencing: The Dice Score is calculated as the overlap of two the binary segmentations $\mathbf{L_F}$ and $\mathbf{L_M}(T_\mu)$,

$$\text{DSC} = \frac{2 \cdot |\mathbf{I_F} \bigcap \mathbf{I_M}(T_\mu)|}{|\mathbf{I_F}| + |\mathbf{I_M}(T_\mu)|} \tag{5.3}$$

Note, that it is actually the ground truth segmentation from each of the data samples, which is deformed to the domain of the chosen reference sample. This is merely a consequence of the direction of the transformation as described in Section 5.1.5.

The mean and maximum (Hausdorff distance) surface errors are calculated as well. Let $\mathbb{S}_F(T_\mu)$ be the deformed surface model of the reference sample. There is no direct point correspondence to the ground truth surfaces, $\mathbb{S}_M$, and they each contain a varying number of vertices - respectively N_F and N_M. The metrics are therefore based on the closest points, i.e. the minimum Euclidean distance from a point, p, to any of the points, q, in the other surface, $\mathbb{S}$:

$$d(p, \mathbb{S}) = \min_{\forall q \in \mathbb{S}} \left(||p - q||_2 \right) \tag{5.4}$$

The mean surface error, $d_{\bar{s}}$, of each sample registration is defined as the average of all the closest point distances:

$$d_{\bar{s}} = \frac{1}{N_F + N_M} \left(\sum_{\forall p \in \mathbb{S}_F(T_\mu)} d(p, \mathbb{S}_M) + \sum_{\forall p \in \mathbb{S}_M} d(p, \mathbb{S}_F(T_\mu)) \right) \tag{5.5}$$

The Hausdorff distance, d_H, is the maximum of all the closest point distances:

$$d_H = \max \left\{ \max_{\forall p \in \mathbb{S}_F(T_\mu)} d(p, \mathbb{S}_M), \max_{\forall p \in \mathbb{S}_M} d(p, \mathbb{S}_F(T_\mu)) \right\} \tag{5.6}$$

Finally. we define the apex error as the Euclidean distance between manually placed apex landmarks, $\vec{A}$, in the surface models of each data sample.

$$d_{\vec{A}} = \left|\left| \vec{A}_F(T_\mu) - \vec{A}_M \right|\right|_2 \tag{5.7}$$

5.2.3 Contributions

Several different registration models were developed during the project. A detailed summary and some additional material is provided before discussing each of the contributions further.

Since registration models consist of many different components, the important and differing aspects have been collected in an overview in Table 5.1. It contains some high-level details - such as the anatomical region being registered, the number of data samples included and at which image resolution. A few actual registration model settings are provided as well, but the full set of details and settings are kept in their respective papers. Initialization is always rigid, and the deformable transform is in all cases a B-spline model. The different cost functions are given below for easy referencing and comparison. As the validation of the registration accuracy and precision is a rather important aspect of the contributions, a comparative summary of the different evaluation metrics is given in Figure 5.8.

The first registration model was built using the bare minimum of settings,

$$\mathcal{C}_1 = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, \mathbf{L_F}, \mathbf{L_M}) + (1 - \alpha) \cdot \mathcal{P}_{\text{BE}}(\mu) \tag{5.8}$$

Whenever a more 'advanced' registration model is used, this cost function serves as a good baseline for comparison of the performance. Therefore, different versions of this cost function were seen in Paper A, B and C.

The skeleton-similarity registration model,

$$\mathcal{C}_2 = \alpha \cdot \mathcal{S}_{\text{Sim}}\left(\mu, \mathbf{L_F}, \mathbf{L_M}\right) + \beta \cdot \mathcal{S}_{\text{CP}}\left(\mu, \mathcal{Z}_F, \mathcal{Z}_M\right) + (1 - \alpha - \beta) \cdot \mathcal{P}_{\text{BE}}(\mu) \tag{5.9}$$

was presented in Paper A, and use the surface landmarks extracted from a cochlear skeleton as a global guidance of the registration procedure (see Section 4.6.3).

Finally the heat distribution registration model,

$$\mathcal{C}_3 = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, \mathbf{H_F}, \mathbf{H_M}) + (1 - \alpha) \cdot \mathcal{P}_{\text{BE}}(\mu) \tag{5.10}$$

which was introduced in Paper B and used with some modifications in Paper I. The main thing to note is that the similarity term is based on the heat distribution volumes, $\mathbf{H_i}$, rather than the segmentations (see Section 4.6.4).

Table 5.1: Comparison of Registration Models
The models are ordered in accordance with chronological development.
IE = inner ear, C = cochlear only, *= different reference

Model No.	**1**	**2**	**3**	**4**	**5**	**6**
Anatomy	IE	IE	C	C	C*	C*
No. dataset	17	17	17	17	19	19
Paper	C/A	A	B	B	-	I
Voxel size [μm]	24.5	24.5	49	49	24.5	24.5
Cost Function	Eq. 5.8	Eq. 5.9	Eq. 5.8	Eq. 5.10	Eq. 5.8	Eq. 5.10
Similarity Metric	SSD/ KS	SSD/ KS	SSD/ KS	SSD	SSD/ KS	SSD
α	0.9	0.8	0.9	0.1	0.95	0.9
β	-	0.11	-	-	-	-
Multi-Res. Stages	9	9	3	3	3	3
Spatial Samples	2^{14}	2^{14}	2^{15}	2^{15}	2^{18}	2^{18}
Mask	no	no	no	no	yes	yes

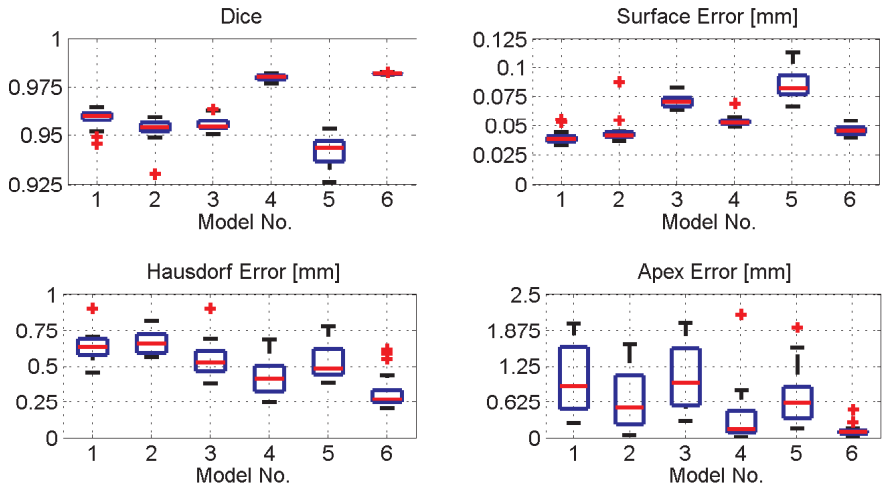


Figure 5.8: Comparison of Registration Model Performance. The accuracy and precision of the different registrations models (see Table 5.1 for corresponding model numbers) are compared. Assuming that errors are normally distributed, an outlier is here defined as values outside $\pm 2.7\sigma$ from the mean.

Discussion: The evaluation of the registration models is an important aspect to discuss. Especially, if the image registration as here is intended for building shape models (see Chapter 6). In such cases, the transformation is implicitly the modeling force of the shape variation. The (in)ability to align two samples accurately will potentially translate into a lack in the shape models' ability to represent the anatomy of the training data.

However, the comparison of registration model accuracies is not that straightforward. The apex error is the only metric score of the ones presented here, which can be compared on the absolute values. The other metrics can only be fairly compared in pairs of two (Model [1,2], [3,4] and [5,6]) as done in the respective papers. The absolute values of these scores are influenced by the vertex densities of the surface models, the voxel resolution, the reference dataset and finally of the anatomy being registered (inner ear versus cochlear) and does therefore not compare directly.

There is still some meaningful information to take away though. Concerning the apex error, the baseline (Model 1, 3, 5) tends to perform worse, and there is a trend showing that the error decreases with the later generations of registration models. Another notable improvement is the precision that the heat-based cost function (Model 4, 6) provides. The results are generally much more consistent. This simply suggests that the experiences gained in the development process was gradually used to refine and build better registration models.

This process is also qualitatively illustrated in Figure 5.9. Even though the skeleton model improves the registration compared to the baseline model, the results with heat distributions are simply better. Comparing the registrations using the baseline models, it should be observed from the figure that model 5 has a better alignment than model 1. This can simply be attributed to a smaller starting turning discrepancy between the target and the TUM sample.

Paper C: The aim of this study was simply to investigate the feasibility of registering inner ear μCT datasets with a basic B-spline based registration model for the purpose of building statistical shape models. Although not perfect, the obtained registration accuracy was sufficiently encouraging in order for us to move forward with this registration model as a base.

Paper A: The primary goal of this publication was to improve upon the basic registration model with a special attention to capturing of the cochlear turning.

The cost function (Equation 5.9) was expanded with landmarks based on a skeleton (see Section 4.6.3). As seen from Figure 5.7, the standard global initialization transformations do not function well with the cochlear spiral. The approach taken in this paper was our first attempt to incorporate a global sim-

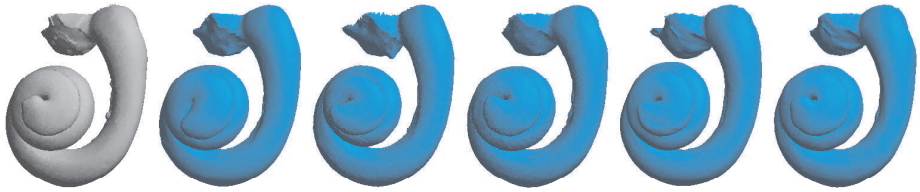


Figure 5.9: Image Registration Models and Cochlea Turning. Given a target shape (grey), a comparison of the resulting deformed surfaces (blue) from the different registration models are shown. The ability to capture the cochlear turning seems to improve with the later registration models. From left to right the model numbers are 1, 2, 4, 5 and 6 respectively.

ilarity or shape regularization into the local non-rigid transformation model. The registration did improve in its ability to capture turning (the apex error was reduced), but lost some of its local adaptability (seen by the minor decrease in Dice Score, surface- and Hausdorff error). This can likely be attributed to the strictness of having one-to-one correspondences with the landmarks, considering that their placement and extraction have some inherent noise. Further, there were still some issues with apical region where landmarks where excluded due to difficulty in consistent placement. In summary, the presented idea was sound, but the execution a bit wanting.

This paper also contributes with the concept of evaluating the local 'torque' on the surface meshes, as a way of describing the registration models' ability to turn the cochlear. Most of all, it was a neat tool for visualizing differences in the resulting transformations, although there are other imaginable uses.

Paper B: The purpose of this manuscript was to explore another way of incorporating a global shape similarity in the local transformation model, this time using heat distributions (Equation 5.10) instead of the landmarks.

The rationale behind the idea was that the heat maps are less strict, in the sense that they do not rely on the given one-to-one correspondence of the landmarks. Adding further to that, with the heat maps it was easy to cover the entire apical turn. Unfortunately, propagating heat through an object like the semi-circular canals could easily lead to some new issues. The region of the registration was therefore reduced to only contain the cochlear and a part of the vestibule. This makes it a very different anatomical model, which of course has some implications for the succeeding shape modelling tasks. This also drastically change the registration model set-up. Fewer resolutions of the multi-resolution scheme

were needed to obtain good alignments. Generally the process of setting up the model parameters was markedly easier. The convergence of the optimization was improved, which probably can be attributed to a simplified cost function, now that spatial position to some extent is encoded in the data. It is at least very plausible that the self-similarity issues explained earlier are removed to a larger degree.

The paper was meant as an exploratory study, and we therefore worked on downsampled data to reduce the complexity and the computational burden. The results were very promising with good increases in both accuracy and precision, although the procedure had some issues with stability. Note for instance the very low α-value required to enforce strong regularization (Table 5.1), and even still there were cases with too extreme deformations present.

Paper I: The focus and contribution of this paper is more on the shape modeling part (see Chapter 6 and 7), but since changes were made to the registration model it is briefly included here for completeness.

The registration procedure builds directly upon our work from Paper B but with the following modifications. The reference dataset was changed to the TUM dataset for reasons relating to the shape modelling. Since it is a cochlea only model, we could include one additional dried dataset, which was previously exempted due to a hole in the posterior SCC (see Section 4.1). The above mentioned stability issues of the heat-based cost function were addressed by registering the data in finer voxel sizes, plus some general debugging and tweaking of the set-up. Backed by the observed general increase in precision and improvement in apex error, we believe that we managed to set-up a promising registration model.

In conclusion, the best of the developed models is the one presented in Paper I. We therefore did the registrations once more in even higher settings (increased number of spatial samples and inclusion of a registration mask - see Table 5.1) to ensure having the 'maximum' possible registration accuracy.

5.3 Summary and Conclusion

In this chapter we have,

- Introduced general image registration theory and terminology. One detail to take note of is the specification of the transformation direction ($T_\mu : \mathbf{I}_F \rightarrow \mathbf{I}_M$) in the `elastix` framework. It goes from the fixed to the moving image, and this has some implications for building and using shape models detailed in the succeeding chapter.
- Given some general considerations regarding image registration of the inner anatomy from μCT data based on our experiences, that may be of relevant use for other researchers.
  - The most notable consideration is the issue of global initialization. We tried to solve it by including global shape based similarity into our non-rigid registration models. This is the main contributions of the papers concerning image registration.
  - A second take home message, is that registration of the entire inner ear is challenging. It may very well be worthwhile to make a split and register the cochlear and the semicircular canal regions separately. We had good success with registration of the cochlear using a calculated heat distribution model for global image similarity. More work could easily be done to establish registration models for the semi-circular canals, but because we are interested in CI-applications our focus went to the cochlea.
- Finally, fulfilling one of the thesis goals, we have here developed and presented a registration model, that by far performs better in aligning cochlear anatomy from μCT than a standard procedure does.

Chapter 6

Statistical Shape Modeling of the Inner Ear

Statistical Shape Modeling (SSM) is a popular technique for working with biological shapes. The form and variability of natural shapes means that simple and perfect geometry models are fundamentally limited in their descriptive accuracy. The amount of parameters or morphometrics required to adequately represent more complex shapes quickly grows to sizes which are difficult to handle manually. For these reasons it makes sense to use computational models of anatomy.

The chapter is structured in a similar manner to Chapter 5. There is an initial part describing the general theory and background deemed necessary for following the remainder of the thesis. This part can be lightly read by a reader already familiar with the topic.

The second part introduces shape modeling of the inner ear anatomy and the contributions of the thesis in this regard. This includes a review of relevant literature and a summary and discussion of the Papers C, D, I and H.

6.1 Theory & Background

An introduction to the underlying multivariate statistics will be given first, and then descriptions of various relevant SSM concepts.

6.1.1 Principal Component Analysis

The PCA is the main driving force of most SSM approaches and originates from multivariate statistics. More rigorous details can therefore be found in [HTF01] for example.

The PCA is an unsupervised approach for defining a structure in observed data in a way that best explains its' variance. Assume having N observations each describing the same p-variables:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,p} \\ x_{2,1} & x_{2,2} & \dots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,p} \end{bmatrix} \tag{6.1}$$

A structure can be imposed on the data by fitting a p-dimensional hyper-ellipsoid to the observations in a least squares sense. Letting $\mathbf{\Sigma}$ be the $p \times p$ covariance matrix (assuming the mean value of the observed variables to be 0) of $\mathbf{X}$, then the operation is the mathematical equivalent of making an eigendecomposition of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \mathbf{X}^T\mathbf{X} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T \tag{6.2}$$

where $\mathbf{\Lambda}$ is a $p \times p$ diagonal matrix containing the sorted eigenvalues, λ_i, in order of decreasing magnitude. The orthogonal $p \times p$ matrix $\mathbf{W}$ contains the corresponding orthonormal eigenvectors, v_i, such that

$$\mathbf{W} = (v_1, v_2, \dots, v_p) \tag{6.3}$$

The eigenvectors form a basis for a space where the explained variance of the projected observations is equal to λ_i. The direction of the first eigenvector describes the most variance and is called the first principal component (PC). The

second eigenvector explains the 2nd most variance, given that it is orthogonal to the first, and is called the second PC and so on. In other words, the principal components form the axes of the hyper-ellipsoid, and the larger PCs will be in the directions of strongly correlated variables. This is illustrated in Figure 6.1. The population of pyramids are generated randomly using a normal distribution for one side length, $a \in \mathcal{N}(20, 5)$ and an independent normal distribution for the height, $h \in \mathcal{N}(12, 4)$. The second side length, b, is generated from a with some added Gaussian noise, to ensure a strong correlation between the two variables.

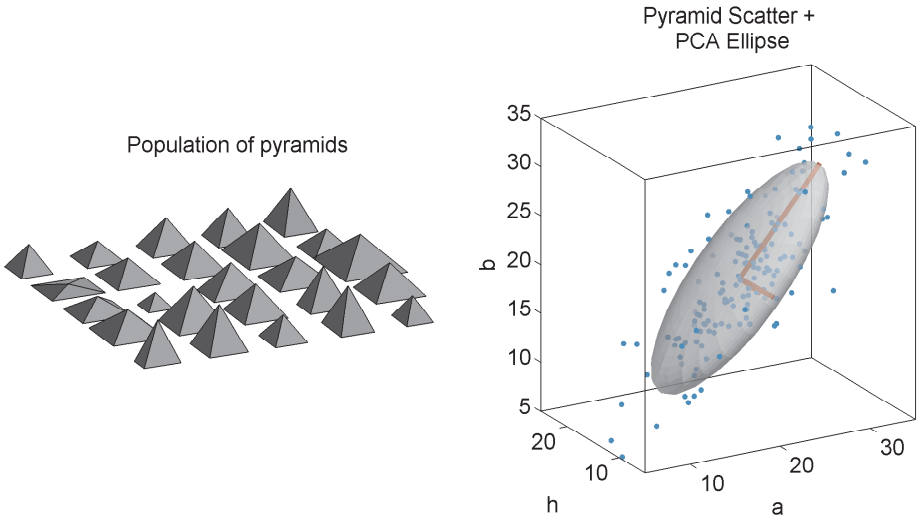


Figure 6.1: PCA, Ellipsoid Data Fitting. Principal Component Analysis is performed on a population of $N = 150$ pyramids described by $p = 3$ variables of height, h, and side lengths a, b. The pyramid bases are mostly square, which implies the correlation between a and b. (Left) A random subset of the pyramids are shown to demonstrate the variability. (Right) Each pyramid plotted as a point in the p-dimensional space, with the PCA ellipse superimposed.

The first key feature of the PCA is the ability to reduce dimensionality of the data. The idea is that PCs with sufficiently low λ can be ignored, because they describe an insignificant amount of variance (assumed to be noise). Using the $M < p$ principal components makes it possible to project the observed data into a space of reduced dimensions:

$$\mathbf{Y} = \mathbf{X}\,(v_1, v_2, \ldots, v_M) \tag{6.4}$$

The more and the stronger the correlations are between the variables, the fewer PCs are needed to describe the variance and the greater the reduction of dimensionality can be achieved.

A second important aspect of the PCA is the fact, that it can be used as a generative model. Any of the data observations, x, can be approximated by adding a linear combination of the M-principal weighted components to the mean observation, $\bar{x}$, i.e.

$$x \approx \bar{x} + \mathbf{W}\mathbf{b}^T \tag{6.5}$$

where the weights of $b = (b_1, b_2, \ldots, b_M, 0, 0, \ldots)$. Note that it is possible to choose the weights in the reduced space and still generate samples in the full dimensionality. Any weighting can in principle be chosen, thereby providing a generative property to the PCA. Generally it is assumed that the PCs are actually a description of a multi-dimensional Gaussian distribution. Constraining the weights to be a maximum of 2.5-3 standard deviations from the mean will therefore usually ensure that instantiated samples are not too far from the observed, i.e.

$$|b_i| \leq \pm 3\sqrt{\lambda_i} \tag{6.6}$$

The two properties of the PCA are illustrated in Figure 6.2, based on the population of pyramids from Figure 6.1. A subspace of two variables explains $> 97\%$ of the observed variance. The first PC explains the (co-dependent) changes in side lengths, while the second PC explains the variability in pyramid height.

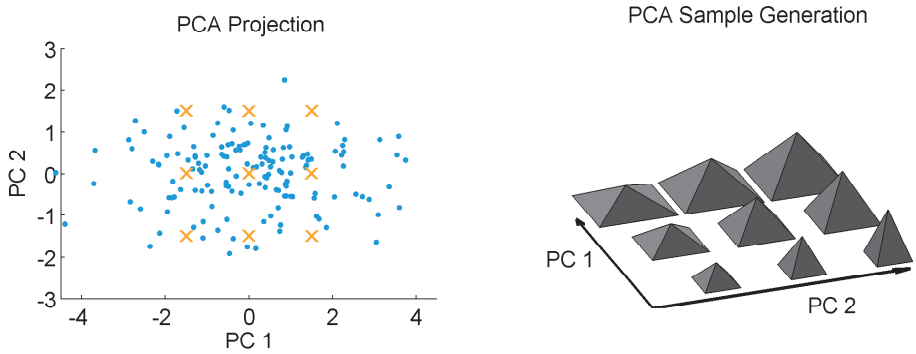


Figure 6.2: PCA Properties. (Left) The population of pyramids from Figure 6.1 projected to the subspace of the first and second PC. (Right) Instantiated pyramid samples corresponding to the orange marks in the PCA projection.

PCA Special Case: The scenario where $N < p$ is typically encountered when working with SSMs and that has some implications for the PCA. It is often the case that the amount of variables or parameters, p, used to describe the anatomy is greater than the number of observed samples, N.

Mathematically, when trying to do the eigendecomposition of $\mathbf{\Sigma}$, this scenario will result in a $p \times p$ $\mathbf{\Lambda}$-matrix where the $p - N$ last eigenvalues are zero. It is therefore more computationally efficient to solve the eigendecomposition for the smaller $N \times N$ covariance matrix defined as,

$$\mathbf{\Sigma} = \mathbf{X}\mathbf{X}^T \tag{6.7}$$

There some additional details on how to operate with the eigenvectors of this reduced matrix, detailed in for example [CTCG95]. However, in principle the PCA works in the same manner, so the mathematical details are the lesser of the consequences.

Having only a few PCs (at most $N - 1$) to describe many parameters, means that the each PC becomes even more holistic or global in its description. The low number of observations increases the risk of not being able to identify all dependencies between variables. Some variables will appear to be correlated in the principal components, but in reality they could be decoupled. There is just not enough data or observations to support this.

6.1.2 General SSM Concepts

The SSM framework is an extensively researched technique, which offers meaningful input to versatile applications in image analysis. It is for instance a valuable tool for conducting shape analysis, and it can serve as a shape prior or regularization in image segmentations and registrations.

The concept of an SSM should be understood in a larger context. It can handle a variety of shapes representations. There are various methods for more or less automatically establishing point correspondences across datasets and ways of fitting the models to new datasets. It has been applied to numerous different anatomical areas, and with varying applications in mind. A thorough review of the SSM methodology is found in the paper from Heimann et al. [HM09].

It is a sane and obvious choice of methodology for working with the inner ear anatomy, and this section will introduce the following key concepts relevant for this thesis,

- Point Distribution Model (PDM)
- Active Shape Model (ASM)
- Statistical Deformation Model (SDM)
- SSM Evaluation

All handling and interaction with SSMs within this thesis is done via the Statismo [LBA+12] framework. It is an open source library made for handling all types of shape models, and aims to do it in a standardized manner.

6.1.2.1 PDM - Point Distribution Model

The 'classic' formulation of a SSM is the PDM. It is explained in the work of Cootes et al. [CTCG95], and is the obvious starting point for a SSM description.

The principal concept is that any shape can be described by a sufficient number of landmarks or points. Typically the landmarks lie on the surface of an object or on the interface with another structure. The points does not need to be anatomical landmarks, but could be any type of pseudo-landmarks.

Mathematically each instance of a population of objects can be described by a point set, $\mathcal{Z}_i$, consisting of p-corresponding points:

$$\mathcal{Z}_i = (\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_p, y_p\}) \tag{6.8}$$

Note that the notation used here is for 2D, in order to make the equations more compact, but it should generally be thought of as extendable to a higher dimensionality (typically 3D).

The trick of the PDM is to realize that each point set can be folded out and be treated as an observation of a multi-dimensional variable. The data matrix (Eq. 6.1) thus becomes,

$$\mathbf{X}_{\text{PDM}} = \begin{bmatrix} \mathcal{Z}_i \\ \mathcal{Z}_{i+1} \\ \vdots \\ \mathcal{Z}_N \end{bmatrix} = \begin{bmatrix} x_{i,1} & \ldots & x_{i,p} & y_{i,1} & \ldots & y_{i,p} \\ x_{i+1,1} & \ldots & x_{i+1,p} & y_{i+1,1} & \ldots & y_{i+1,p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & \ldots & x_{N,p} & y_{N,1} & \ldots & y_{N,p} \end{bmatrix} \tag{6.9}$$

Applying the PCA to this structure provides the PDM, which qualitatively is a description of how the landmarks co-vary using a certain number of principal components (also called modes of variation). However, before performing a PCA on a distribution of point coordinates described here, it is generally advisable to remove differences in translation, rotation and often in scaling as well. Otherwise these effects will contribute with so much of the total variance, that actual differences in shape are drowned out.

The typical procedure for doing this is the Generalized Procrustes Analysis (GPA). Having shapes with point correspondences makes it possible to calculate

global transforms that minimizes the mean squared distance between two shapes analytically. Using this in an iterative process makes it possible to align a group shapes to their unknown mean shape easily. Further details can be found in for instance [CTCG95].

PDM and Image Registration: The main challenge of the PDM approach is to describe all the shapes of the training set using the exact same p-points or landmarks. Manual placement of points, although classic, is time-consuming and it scales badly with the number of landmarks and is especially challenging when operating with 3D objects.

One solution, as introduced in the work of Frangi et al. [FRSN02], is to establish point correspondences via image registration techniques. One image is chosen as the reference, $\mathbf{I}_{\text{Ref}}$. All other training images, $\mathbf{I}_i$, are registered to this reference. Having defined a set of landmarks, $\mathcal{Z}_{\text{Ref}}$, or a surface-model, $\mathbb{S}_{\text{Ref}}$, in the reference domain, it can now be projected unto a training sample using the resulting transformation. Assuming a perfect registration, it now resembles the anatomy of the training data sample, but in the point or surface structure of the reference. Let T_μ^i be the transformation $T_\mu^i : \mathbf{I}_{Ref} \rightarrow \mathbf{I}_i$ resulting from application of a certain image registration model. Then,

$$\mathcal{Z}_i \approx \mathcal{Z}_{Ref}(T_\mu^i) \tag{6.10}$$

The concept is illustrated in Figure 6.3. With this approach it becomes possible to represent each of the training samples with point-to-point correspondence. The data matrix (Eq. 6.9), for which PCA is intended, can now be formulated as,

$$\mathbf{X}_{\text{PDM}} = \begin{bmatrix} \mathcal{Z}_{\text{Ref}}(T_\mu^i) \\ \mathcal{Z}_{\text{Ref}}(T_\mu^{i+1}) \\ \vdots \\ \mathcal{Z}_{\text{Ref}}(T_\mu^{N-1}) \end{bmatrix} \tag{6.11}$$

It is generally good sense not to include the chosen reference dataset into the data matrix. Ideally the chosen reference shape should either be accurately representative of the true population, or the mean shape of the training samples can be used for this purpose.

Note, that any kind of image registration model with any kind of transformation model could be used. It is even possible to use a different registration for each training sample. With this approach all variability between shapes is modeled by the transformation, so naturally the approximation (Eq. 6.10) should be as accurate as possible. All considerations of a good image registration model (see Chapter 5) are therefore relevant.

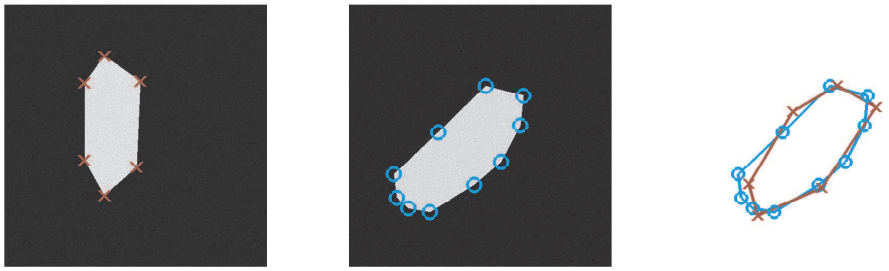


Figure 6.3: Point Correspondence from Image Registration. Given a reference image and model (left), and a target image (middle). The target shape can be approximated by applying the resulting transformation of an image registration to the reference model (right).

6.1.2.2 ASM - Active Shape Model

The ASM can essentially be regarded as a PDM-to-image fitting procedure. The process tries to solve the problem of describing a previously unseen sample (i.e. a test data set) in the definition of the PDM.

In its basic form it is an iterative optimization approach,

- The mean shape of the PDM, $\bar{x}$, is placed in proximity of the target shape in an image.
- A displacement vector, dx, is estimated for all points in the model.
  - The calculation can be done in numerous ways, but in the original formulation it is implemented by finding a strong image gradient in the direction of the normal vector going from a model vertex.
  - In a first step a similarity transform is calculated to adjust for the pose change of the model (translation, rotation and scaling) that can be explained by dx.
  - The residual displacement error after the similarity transform can only be described by adjusting the shape. The optimal weights, b, of the principal components can be calculated, and forced to be within the allowable shape space (Eq. 6.6).
- The shape estimate is updated with the resulting pose and shape changes, and the process is iterated until convergence, which is specified by some tolerance criteria.

This is only a rough outline of the ASM. A more detailed description is not

required, as the ASM is not an approach used in this thesis. More rigorous details and extensions to the method can be found in [CTCG95, HM09].

However, the main point of the procedure is the strong regularizing effect of forcing the current shape estimate into the learned space of allowable shapes. Assuming the PDM is trained well, it can become a very robust procedure.

6.1.2.3 SDM - Statistical Deformation Model

The PDM can be closely connected to an image registration model for solving the point correspondence problem between the training data. The most consistent approach is naturally to use the same registration model with the same p-parameters in the transformation. Essentially realizing that the transform parameters can be considered corresponding variables leads to the idea of the Statistical Deformation Model (SDM).

The concept is introduced and explained in detail in the work of Rueckert et al. [RFS01] and Frangi et. al. [FRSN02], where the non-rigid transformations were based on a B-spline model. As shown in Figure 5.2, this transform be described as a deformation field from a fixed grid of nodes - and the approach is therefore aptly named SDM. The following descriptions assumes that a B-spline model is used, although strictly speaking this concept is not limited to that transformation type.

Let the k'th B-spline node be associated with the three deformation parameters $(v_{k,x}, v_{k,y}, v_{k,z})$. The total $3p$-parameters of the B-spline transformation model can then be stacked in a vector, similar to the way it was done for the landmarks of the PDM,

$$T_\mu = \begin{bmatrix} v_{1,x} & \dots & v_{p,x} & v_{1,y} & \dots & v_{p,y} & v_{1,z} & \dots & v_{p,z} \end{bmatrix} \tag{6.12}$$

Using the same notation as previously, $T_\mu^i : \mathbf{I}_{Ref} \rightarrow \mathbf{I}_i$ is the resulting transformation of an image registration between a chosen reference image and the i'th sample of the training dataset. Ensuring that the same number of nodes are placed consistently for all registrations provides correspondence between the deformation vectors (parameters), and the data matrix thus becomes,

$$\mathbf{X}_{\text{SDM}} = \begin{bmatrix} T_\mu^i \\ T_\mu^{i+1} \\ \vdots \\ T_\mu^{N-1} \end{bmatrix} \tag{6.13}$$

Applying the PCA to this structure provides the SDM, which qualitatively then is a description of how the chosen reference domain statistically deforms to fit other shapes using a certain number of modes of variation.

Comparison to PDM: By comparing Eq. 6.11 and 6.13 it can be realized that all of the modeled anatomical variability comes directly from the image registration model, underlining the importance of having accurate registrations. In the typical scenario, where the same registration model is used for all training samples, it is possible to build both a SDM and PDM from the same transformations. In such a case, the two models should describe the same anatomical variability just using different shape representations.

The PDM may conceptually be easier to understand than the SDM. Instantiating a point set or surface mesh from the PDM is easy to visualize and interpret. However the SDM has some good properties. An instantiated deformation field, while not very useful on its own (and certainly difficult to visualize properly in 3D), can however be readily applied to anything defined within the domain of the reference, which makes it very versatile. Auxiliary points or anatomical structures can be introduced, and deformed without having to be explicitly included in the statistical model. This facilitates a description of intra- and inter-structure variability not immediately available with the PDM.

SDM-regularized Image Registration: Similarly to the ASM, there are ways of using the SDM in a fitting procedure to describe new and previously unseen samples. Since the SDM is basically a statistical model stemming from a series of image registration transformations, then it seems logical to incorporate the model into a new image registration model.

A realization of this requires that a SDM built with Statismo [LBA+12] can be integrated into the registration framework of `elastix` [KSM+10]. Such a software integration package exists already - aptly named Statismo-`elastix`. Using that tool the following type of image registration model is enabled,

$$\hat{\mu} = \arg\min_{\mu} \mathcal{C}(T_{\mu}^{\mathrm{SDM}}, \mathbf{I}_{\mathrm{Ref}}, \mathbf{I}_{M}) \tag{6.14}$$

where a new type of transformation, T_{μ}^{SDM}, based on the SDM is introduced. The fixed image should correspond to the reference sample that was used to build the SDM, to maintain correspondence between the image and SDM domain.

By putting it into the frame of an image registration the whole optimization procedure is taken care of, and it is in principle possible to take advantage of

the standard image registration components as described in Chapter 5. Different types of images can be used (gray-scale/binary, CT/μCT etc.) by adjusting the similarity metric accordingly. In principle, it is also possible to add additional regularizations in the cost function. However, the main point of incorporating the SDM is to have the shape regularizing property, which is inherently built into the transformations (assuming a well trained model). Referring to this transformation type as a *SDM-regularized B-spline transformation* may therefore be an appropriate name.

The amount of parameters to optimize for corresponds to the M-retained modes of variation of the PCA. Typically, the observed number of samples, N, is much smaller than the $3p$ parameters of the B-spline model (assuming p nodes). The normal scenario will have $M < N < 3p$, and it is therefore a highly limited parameter space to optimize compared to a normal local non-rigid registration. While this is efficient and fast for an optimization point of view, it is also highly regularized as only the transformations within the learned space of allowable deformations can be accepted.

6.1.2.4 SSM Evaluation

A substantial challenge when working with statistical shape models is the validation. Characterizing the accuracy and precision of a model is naturally important, but unfortunately not an easy task and there is not a standardized approach. Consider that in some cases the SSMs are meant as a tool for further image- or data-processing via the above mentioned fitting procedures. This makes it possible to evaluate the ability of the model to perform the intended task, rather than the model itself. It is implicitly assumed that if the SSM is built to more accurately represent the population, then it will also perform better at the given task. However, when SSMs are built for analyzing shapes and how the morphology can be related to other factors, then the model itself needs a validation.

As stated there is no standardized approach, but four possible ways are described in the following paragraphs.

Qualitative: Model evaluation should preferably be quantitative. However, when it comes to shape models a visualization and description of the modes of variation is quite sensible and typical to perform. When the parameters of the model (i.e. landmark coordinates or transformation parameters etc.) are stacked into a vector of a very large dimension, the immediate ability to conceptually

understand the shape is lost. A visualization of the modes of variation can therefore be very helpful.

A qualitative evaluation consists of simply sampling the allowable shape space along a certain principal component and visualizing the changes of the shape, as illustrated in Figure 6.2.

The underlying anatomical model: As stated earlier (Eq. 6.11 and 6.13), the variability modeled in a SSM is in some cases directly dependent on an underlying registration model. It can to some extent be assumed, that the accuracy of the registration model is directly related to the accurateness with which the SSM describes the population variability.

However, as presented and discussed in Chapter 5 evaluation and comparison of registration models can also a difficult task to perform.

Internal model validation: This type of evaluation is only possible when all samples are already defined within a specific domain or representation. The following descriptions assume a PDM type, but image-based equivalents could be imagined.

When shapes have point-to-point correspondence, the evaluation of a shape model is actually rather straightforward. The following three metrics, proposed and detailed in [Dav02, SRN+03], makes for sensible criteria to measure:

- Generalization
- Specificity
- Compactness

Generalization is the ability of the model to represent unseen or new samples. If the set of training data is not representative for the true variation in the population, or if the model is overfitted to the training samples, then the generalization capability will be poor.

The property is measured in a leave-one-out cross-validation. The shape model is rebuild with sample x_i excluded. Since point-to-point correspondence is still assumed known, the model parameters, b_i, providing the closest fit can be estimated analytically using Equation 6.5. The property can be evaluated as a function of the number of retained modes of variation, M. For each of the N training samples the sum of squares approximation error is calculated, and

finally the generalization ability, $G(M)$, is defined as the mean squared error,

$$G(M) = \frac{1}{N} \sum_{i=1}^{N} \left| x_i - (\bar{x} + \sum_{m=1}^{M} v_i^m b_i^m) \right|^2 \tag{6.15}$$

The standard error of $G(M)$ can be calculated as

$$\sigma_{G(M)} = \frac{\sigma}{\sqrt{N-1}} \tag{6.16}$$

where sigma is the sample standard deviation of $G(M)$.

Specificity is the ability of the model to generate samples that are actually represented in the training data.

This is typically measured by instantiating N random shapes from the allowable shape space with M retained components, $x_j(M)$. The closest of the training samples, x'_j, is found as the sample with smallest sum of squares error. Finally the specificity, $S(M)$, is defined as the mean squared error over the instantiated population,

$$S(M) = \frac{1}{N} \sum_{i=1}^{N} \left| x_j(M) - x'_j \right|^2 \tag{6.17}$$

The standard error of $S(M)$ can be calculated as

$$\sigma_{S(M)} = \frac{\sigma}{\sqrt{N-1}} \tag{6.18}$$

where sigma is the sample standard deviation of $S(M)$.

The sampling process used to instantiate random samples can in principle be anything, but using a uniform distribution ensures that the limits of the allowable shape space is reached more easily, compared to sampling from a Guassian-distribution.

Finally, the **compactness** of a model can be used as an evaluation criteria. The simpler the model (i.e. the fewer principal components required to model a certain amount of data variance) the better.

The cumulative variance as a function of retained modes is therefore defined as the model compactness, $C(M)$.

$$C(M) = \sum_{m=1}^{M} \lambda_m \tag{6.19}$$

The standard error of $C(M)$ can be calculated using,

$$\sigma_{C(M)} = \sum_{m=1}^{M} \sqrt{\frac{2}{N}} \lambda_m \tag{6.20}$$

The most notable and important issue with internal model validation is that point correspondence is required! In most scenarios it is actually difficult to obtain the correspondences without some kind of error, both when building the model and when applying or fitting it to new samples. This form of model validation, is therefore idealized in some sense.

External Model Validation: Finally, there is the possibility of testing the SSM outside of the situation where point-to-point correspondence is assumed known. I.e. testing the ability of the shape model to perform a task on some test data. This would typically be fitting the model uto unseen data using the ASM/SDM procedures described above.

Note that is possible to use a training dataset as a test sample in a leave-one-out cross-validation set-up. One training dataset is removed, the reduced SSM is rebuilt and the task is performed. In some sense, this procedure is closely related to the generalization ability described above, except that the correspondence between the excluded training data and the model should not be assumed known. The evaluation should therefore rely on other metrics for measuring accuracy and precision. These scores are further influenced by the inaccuracies and uncertainties stemming from the performance of the task. In principle a shape model could be built to perfectly represent its population, but still perform poorly at the given task.

6.2 Inner Ear Statistical Shape Models

One of the primary thesis goals is to build shape models of the inner ear or cochlea. Having introduced the general theory and concepts, this section will describe the different SSMs, which have been built throughout this project and that appear in the literature. Some of our models are presented in different contributions. Other parts of our modeling efforts were not published for various reasons. To facilitate an easier and clearer comparison, an overview and summary of the models are provided followed by a discussion of the results and experiences gained in the process.

One of the key applications of our work is to use the built SSMs in fitting procedures to CT images. Due to the low resolution and lack of details in CT, a fitting procedure has to rely on a strong shape prior in order to be successful. We are most interested in exploring how well this can be done. Most of our validation is therefore an external test of the SDM models' ability to fit in unseen samples.

6.2.1 Previous Work

There is a large body of work concerning statistical shape modeling of numerous different anatomies. However, applying SSMs to the cochlear and inner ear anatomy is still sparsely researched, and the most relevant of those studies are presented below.

Wannaprasert, 2013 [Wan13]: Although the thesis of Wannaprasert has a completely different scope for analyzing inner ear anatomy, there are a few interesting points to take away from it.

Scope: The study is based on μCT scans of inner ear specimens from a variety of mammalian species. The scope is to investigate cochlea morphology and its connection to eco-behavioral traits and evolutionary aspects.

Model building: 25 corresponding cochlear 'skeleton' landmarks are extracted from manual placement of centerline points in cross-sectional slices taken with equi-angular distance around a manually defined cochlear central axis. (There is some degree of similarity to our approach used in Paper A). The corresponding cochlear skeleton points are then aligned using GPA followed by PCA, so it is in principle a PDM type, although the terminology from SSM is never used.

Results: Several different statistical analyzes are done, but the findings are generally not very conclusive, and presented more as indicative of certain connections between eco-beavioral traits and cochlear shape.

The key things to note from this study,

- Building correspondences of the cochlea anatomy is not straightforward, and the study presents a much more simple approach than us. The study rely only on a relatively low number of centerline-skeleton points as being representative for the general cochlear morphology. This demonstrates an important shape modeling point, namely that there is no right or wrong way of building a SSM. The shape model should not be made more complex than the application demands for.
- The lack of statistical significant results may in some cases be due to the lack of data, but it also conceivable that the skeleton modeling choice affects the statistical analyzes. However, the study clearly lacks a sound evaluation of the built PDM, underlining the importance in performing a model validation.

Noble et al. [NLMD11, NLGD13]: The most notable and directly relevant work are the publications from a American research group, whose strategy is also to use SSMs of cochlear structures to aid in various CI-applications.

Scope: In these particular papers (out of a longer series) their μCT-based PDM is presented along with their ASM procedure to segment and predict intra-cochlear anatomy in pre-operative patient CT datasets.

Model building: The model is based on six μCTs (36 micron voxel-size), where scala tympani and scala vestibuli are manually segmented and a surface model extracted. Point correspondences are built using an image registration procedure, although the accuracies of the alignments are never stated. In a later publication [NLGD13], the PDM is further expanded to include a model of the spiral ganglion. The PDM is built as one global model. Further all vertices of the model are marked as either internal or external points, such that if a point faces the outer cochlear wall it is marked as external.

Model fitting to CT: A ASM type procedure is employed to fit the PDM to low resolution CT data. As stated previously, the ASM works by estimating a displacement of each model vertex. The lack of image contrast in intra-cochlear anatomy in CT means that all the internal vertices of the PDM will never be able to find a correct point to displace to. This issue is handled by making a weighted ASM (wASM), simply meaning that displacements of all internal

vertices are ignored. The fitting relies solely on the contrasted cochlear outer wall.

Validation: The ASM fitting is evaluated (externally) in a leave-one-out cross-validation set-up with the training samples. The evaluation is made using Dice Score, mean and maximum surface error, which are presented in Table 6.1.

Results: Combining their fitted patient specific cochlea model with a segmentation of the post-operative placement of CI electrodes, they are able to make corrections to the CI-programming that results in improved hearing restoration.

The main things to note about this work,

- The PDM is based on a limited data foundation. Six samples are hardly able to represent to true population variability.
- The shape modeling approaches are rather straightforward and standard, i.e. a single global PDM and a tweaked basic ASM procedure. However, even with this and a limited amount of training samples they are able to optimize CI-programming.

6.2.2 Contributions

Combining all of the different registration models presented in Chapter 5 with all of the various possible takes on shape modeling and evaluation would quickly become overwhelmingly confusing. Presented and compared here are therefore a selection of four models. The goal is to convey the most illustrative shape modeling points and the experiences gained in the process. The four models are called:

- Inner Ear Baseline Model (B.)
- Inner Ear Skeleton Model (S.)
- Cochlea Heat TUM Model (H.)
- Inner Ear Multi-Region Model

Generally speaking, the point correspondences for all of these models were built using image registrations. This enables a construction of both a PDM and a SDM version. It is simple to sample the PDM along certain principal components and visualize the result, in order to make a qualitative validation of the model. The SDM version is used for fitting to unseen samples using the SDM-regularized B-Spline transformation model (Eq. 6.14). Because the fitting procedure is just another image registration, it is possible to use the exact same

evaluation metrics in the same manner as previously, i.e. Dice Score (Eq. 5.3), mean surface- (Eq. 5.5), Hausdorff- (Eq. 5.6) and apex error (Eq. 5.7).

An overview of the models and their fitting accuracy are given in Table 6.1, Figure 6.4 and Figure 6.5. For the validation presented here, the SDM is fitted to CT and/or μCT training samples in leave-one-out cross-validation (LOOCV) set-up. This results in a more idealized validation procedure, than if 'real' test samples had been used, as the initial position and quality of the training data makes the fitting problem easier to solve.

Table 6.1: Evaluation of SDM fitting accuracy
Models were fitted to training samples in a LOOCV set-up, and the mean value of the performance metrics are given. The performance of the wASM presented in [NLMD11] is also included.

Model	**B.**	**S.**	**H.**	**S.** (CT)	**H.** (CT)	**wASM** (CT)
Anatomy	IE	IE	C	IE	C	C
No. dataset	17	17	19	17	19	6
Dice	0.83	0.83	0.90	0.81	0.88	0.75
Surface Err. [mm]	0.15	0.15	0.10	0.17	0.11	0.2
Hausdorff Err. [mm]	0.88	0.90	0.56	1.01	0.58	0.8
Apex Err. [mm]	0.98	0.77	0.85	0.80	0.84	-

Inner Ear Baseline Model: This was the first iteration of the shape models, built from the 'basic' image registrations (Eq. 5.8), and serves the purpose of having a baseline performance. Both a PDM and SDM was built, and the PDM was included in Paper C.

From a qualitative evaluation (Figure 6.5) and from the observed apex error it is concluded that little cochlear turning is present in the modes of variation. This explains our need for building better models - primarily by improving the underlying registrations.

Inner Ear Skeleton Model: This was the second generation of shape models, based on the registrations using skeleton landmarks (Eq. 5.9) to improve

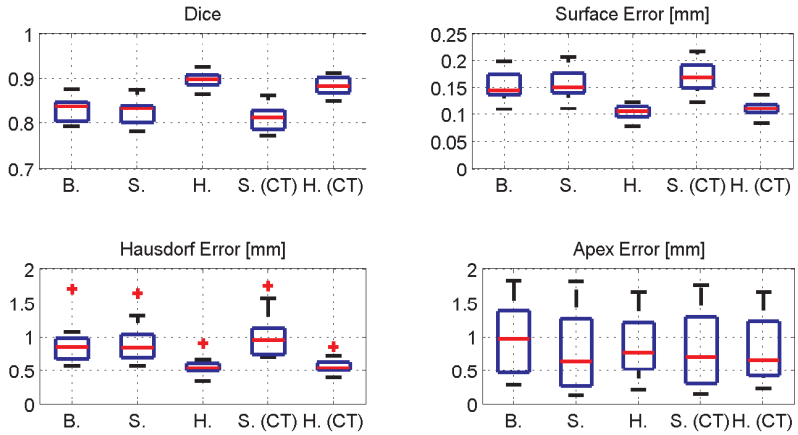


Figure 6.4: Validation of SDMs. Boxplots showing the performance of the SDM fitting with models corresponding to Table 6.1.

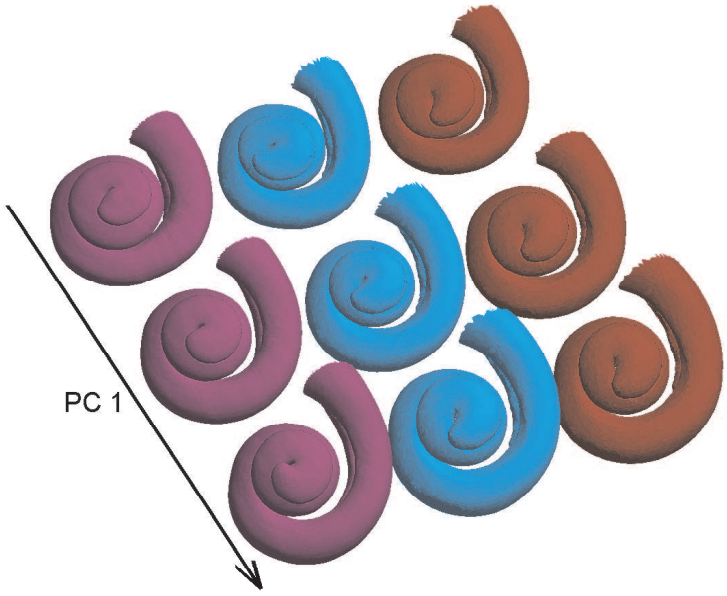


Figure 6.5: Qualitative Comparison of PDMs. The variance explained in the 1st mode of variation for the PDMs using heat similarity (purple), skeleton similarity (blue) and baseline registration (red).

cochlear turning. Both a PDM and SDM were constructed, and the SDM was the underlying model used in Paper D.

The SDM was fitted to both μCT and CT training data. Evaluating against the baseline it hardly makes a difference whether one compares the registration models (Figure 5.8) or the SDM fitting to μCT data (Figure 6.4), as they both show the same trends. An actual improvement was only visible in the apex error. While the idea of including a skeleton in principle was good, the effect was weaker than we had hoped.

Comparing the SDM fitting to the wASM results of Noble et al. [NLMD11], it can be seen that our results are slightly better. However, since these models cannot be compared on a absolute scale, then it is difficult to conclusively state anything. The studies represent two differing anatomical regions and two different fitting procedures. While both the SDM and wASM solves the same task, the approach of keeping everything as image registrations has certain advantages. For instance, we don't have to explicitly handle internal/external cochlear differences.

Returning to the shape model itself, it was noted to have a few issues. Even though the cochlear turning was improved, the model still could not generalize very well to samples with the largest of the possible cochlear turns (the apex error went down, but the standard deviation was still remarkably large compared to the baseline model).

Secondly, as illustrated in Figure 6.6, the poorest fit was typically observed in the SCC region, contributing with an increase in surface- and Hausdorff error. Note that the image registration of the whole inner ear proved difficult to do properly (see Chapter 5). We achieved an average mean surface error in the order of 0.05 mm for the entire inner ear during image registration (Figure 5.8). It is therefore a reasonable assumption, that the underlying registrations captures the SCC variability to the extent possible. However, when fitting the inner ear shape model the performance was observed to drop to 0.15/0.17 mm. This raises an important modelling point. The limited amount of data unfortunately means that the shape model is not able to represent the true variability of the vestibular system. When this is combined into a global fitting procedure, the optimization must compromise between a decent vestibular and cochlear fit, which translates into an overall poorer performance. Arguably, it is then a non-optimal choice to work with the entire inner ear as a single global shape model.

To summaries the main points of the study of this model,

- One global shape model of both cochlea and vestibular region is sub-optimal. A solution is either to acquire a lot more data, or make changes

to the modeling approach.
- Fitting of apical turn does still not generalize to the extent of the population variability.
- Performance seems comparable to the work of Noble et al.[NLMD11].

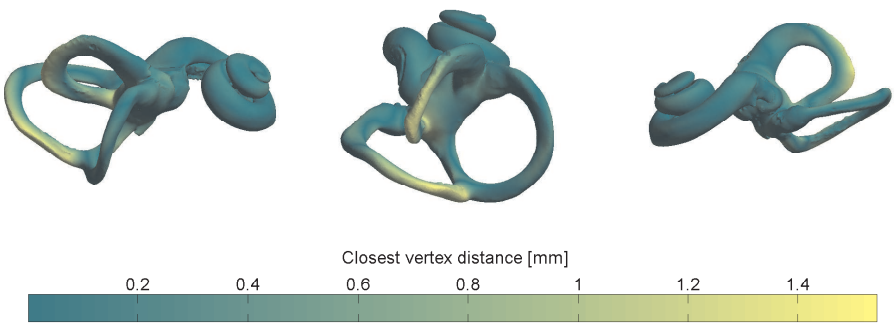


Figure 6.6: SDM Fitting Accuracy. A target surface model is shown from multiple views, colored with the distance to the closest vertex of the fitted reference surface.

Cochlea Heat TUM Model: Two independent changes constitutes the step to the third generation shape model.

- One change, motivated by the prospect of having a shape model that captures the population variability more precisely, was switching the underlying registration models to the one based on heat similarity (Eq. 5.10).
- Secondly the TUM sample was used as the reference, which provides two distinct advantages.
  - It becomes a shape model of the cochlea only. This should hopefully allow us to better capture the cochlea variability in the PCA modes, and remove the influence of the vestibular region during the SDM-fitting.
  - This data sample contains a model of the cochlear partition. Some of our intended applications can benefit substantially by having an anatomical model with an actual separation between the scala tympani and scala vestibuli.

Building the SDM based on the registration model with heat similarity should arguably yield a better shape model considering the accuracy of the registrations. Fortunately it is also observed to translate into an improvement of the

SDM fitting performance (increased Dice score, and lowered average surface- and Hausdorff error - Figure 6.4), both considering the fit to μCT and CT data. Although the numbers do not compare absolutely, the apparent increase in precision speaks of an improved procedure.

The average apex error stays approximately the same for all shape models. It would seem that the ability to adjust or fit to the true cochlear turning is problematic no matter which shape model we use. The issue is not necessarily the SDM, but the fitting procedure and/or a lack of data. The apical region is a very small one (especially in CT images), meaning that a poor fit here is not seen as a big error in the optimization procedure. The limited amount of data further means that there are no PCA modes to independently adjust the apical region, without also making sizable changes elsewhere in the model.

Changing the reference dataset from a dried specimen to the TUM sample allowed us to have an anatomical model with a cochlear partition (CP). As the CP is not present in any of the dried samples, the registrations were made disregarding this anatomical structure, and it is therefore not possible to incorporate the CP variability in the shape model. Further, since the basilar membrane is not present in CT data, there would be no information to actually adjust the CP position during SDM-fitting.

Having everything defined in terms of deformation fields (or transformation parameters) is one of the advantages of using an SDM approach. The CP does not have to be actively modelled. Any instantiated deformation can simply also be applied to the CP mesh, thereby having it passively following any other corresponding deformations. While it is a neat feature of the SDM, there is absolutely no guarantee for accurate and precise placement of the CP during SDM-fitting procedures. We simply do not know whether the current CP model is in any way representative for the real basilar membrane position.

This SDM is to be considered the final model of this project, and as the one that will be most useful for our intended CI-related applications (see Chapter 7). The model was presented in Paper I, with the purpose of making a summarizing demonstration and validation of the SDM model building and fitting process. This SDM can be seen as one of the key contributions of this thesis, as it serves a fulfillment of one of the main thesis goals.

With this study we have,

- Shown that the image registration efforts of Chapter 5 also translates into a better shape model.
- Built a detailed anatomical cochlea model and a description of its' statistical shape variability.

- Demonstrated the key component for our procedure for patient specific anatomical modelling. The SDM fitting was initially tested on the training samples in a leave-one-out cross-validation study, and was shown to perform 'at par' if not better than other published methods.

Inner Ear Multi-Region Model: The improvements of the previous shape model was primarily achieved by simplifying the problem - i.e. disregarding the vestibular region in both the image registration and the shape model building process.

Resorting to a cochlea model is not the only possible solution. An alternative, explored here and in Paper H, is to simply construct the shape model differently, trying to build the model with an independence between the functional regions of the inner ear (illustrated in Figure 6.7).

Approach: First, point correspondences are established with a local non-rigid image registration procedure between the 17 dried specimens. The chosen reference surface model is deformed (Eq. 6.10) to each of training samples to obtain surface models with point correspondences. We set-up procedures for splitting and merging regions of the meshes, allowing the creation of regional PDMs that can behave independently, instead of building one global PDM.

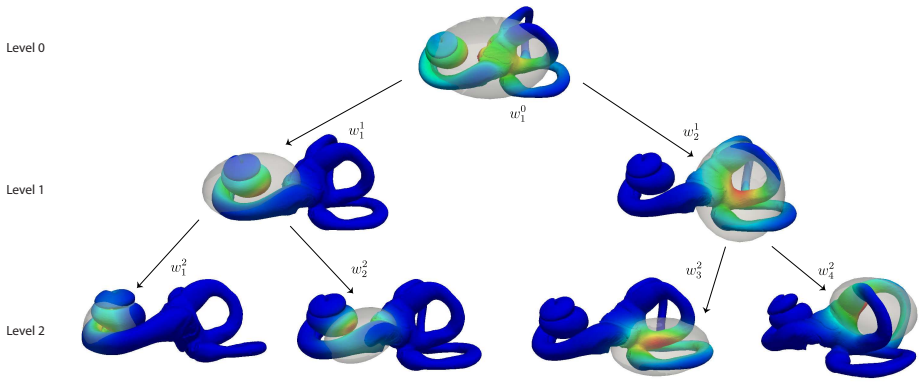


Figure 6.7: Multi-region PDM. The inner ear surface model can be split into functional subregions, whose variability is then modeled independently.

A validation of the generalization ability of the multi-regional PDM showed a promising potential of the approach. However, some additional work would be

required for using this idea in a fitting procedure.

The main contribution of the study relating to this thesis,

- It supports the idea of modeling regions separately, considering the limited amount of data available and the large variability of the inner ear anatomy.

6.3 Summary and Conclusion

In this chapter we have,

- Introduced some general statistical shape modeling methods and terminology, in particular with regards to the interplay with deformable image registration and shape model fitting.
- Given some general considerations regarding statistical shape modeling of the inner anatomy from μCT data based on our experiences, that may be of relevant use for other researchers.
  - A noteworthy experience concerns modeling of the full inner ear vs. the cochlea only. Given the scarce availability of μCT, some caution should be taken when trying to build a model of the full inner ear. Simplifying the problem and building regions separately, potentially using some form of multi-region modeling techniques, could be an adequate solution.
- We have obtained a detailed anatomical cochlea model and a description of its statistical shape variability, which completes one of the thesis goals. We can show that the efforts made to improve the image registration translates into a more accurate and precise shape model. There is still room for improvement, but with the limited number of training samples and the limitations of the data, we believe that we have made the most of it.

Chapter 7

Statistical Shape Models and CI-Applications

The final thesis objectives concern the application of the developed methods and shape models. This chapter focuses on the aspects where the efforts from the previous chapters can be used in relation to CIs,

1. Cochlea shape prediction from pre-op CT
2. Electrode position from post-op CT
3. Simulation of Electrode Array Insertion
4. Electrical Stimulation Models

The first point relates to how the statistical shape model can be used to generate a patient specific model of the cochlea anatomy. Having such a model has certain applications on its own, but arguably the truly interesting potential emerges once the anatomy is related to a model of the inserted CI electrode array. The electrode placement can be estimated from post-operative imaging (point 2) or predicted using a virtual simulation (point 3). Having a the complete model of the hearing system opens up for a range of applications based on computational models (point 4). The above mentioned points will be addressed in more detail in following subsections, and each point largely corresponds to a specific publication. Most of the work presented here is primarily driven by colleagues and collaborators from the research project.

7.1 Patient Specific Anatomical Modelling from CT Images

This is one of the primary applications of our work concerning SSMs,

- The goal is to infer the detailed anatomy of a μCT from a clinical CT image to obtain a patient specific anatomical model.

Essentially, if the clinical CT image is seen as a very sparse representation of an underlying μCT image, then we intend to recover the full information to the best of our ability. One naive strategy would be to make a deformable registration between the target CT image and a suited μCT. However, due to the large difference in resolution, the result of such a procedure would be highly unpredictable and not well behaved (see Figure 7.1, top right). Some type of regularization is required, and including a shape prior in the form of a statistical shape model is an obvious approach to explore, as this more or less guarantees that the output result has a plausible anatomical shape (assuming that the shape model was trained properly)(see Figure 7.1, bottom right).

The main component for realization of this strategy is the SDM fitting procedure. It was introduced and initially evaluated in Section 6.1.2.3 and 6.2.2, but only using cross-validation on the training data, which is a slightly simpler problem. When the strategy is applied to test- and clinical CT data two additional influencing aspects has to be considered,

1. The initial image alignment
2. Performance evaluation

With the training data we have the ground truth segmentations for use in performance evaluation, and the data is located exactly in the position in which the shape model was trained.

7.1.1 The pipeline

The procedure for making a prediction of patient anatomy using the SDM given a pre-operative CT image was proposed in Contribution D, but is also restated here and illustrated in Figure 7.2.

The pipeline can be summarized in three steps. First an initialization based on a landmark-based transformation (T_I) followed by a refining image regis-

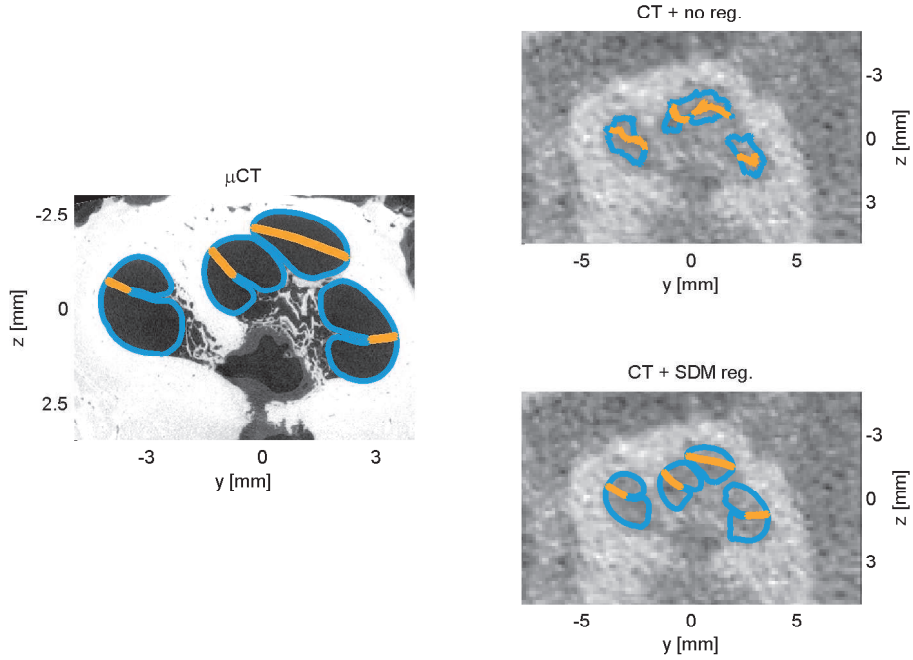


Figure 7.1: The Patient Specific Anatomy Estimation Problem in a Nutshell. The intra-cochlear structures visible in μCT data (left) are difficult to transfer to a clinical CT image (upper right) unless some form of regularization is used (lower right).

tration (T_{II}) (suggested in Section 4.6.1) is calculated to bring the CT image into the coordinate system of μCT and the SDM. Secondly, a SDM-regularized image registration (T_{SDM}) is performed with the reference μCT (presented in Section 6.1.2.3). The third and final step is a projection of the deformed surface model back to the coordinate system of the CT data by simply inverting the rigid transforms.

Note, that because of the transformation direction specified in `elastix` (see Section 5.1.5), it is immediately possible to apply the SDM deformation to the anatomical reference surface model (i.e. $\mathbb{S}(T_{\text{SDM}})$) and transform it into the space of the CT image.

Ideally, in an imagined real clinical scenario, an automatic initialization procedure would be preferred instead of the landmark-based approach. This could for example be accomplished by an image registration between full head CT images similar to [NLMD11]. However, in the frame of this thesis, where the amount of test data is limited, the manual initialization works fine.

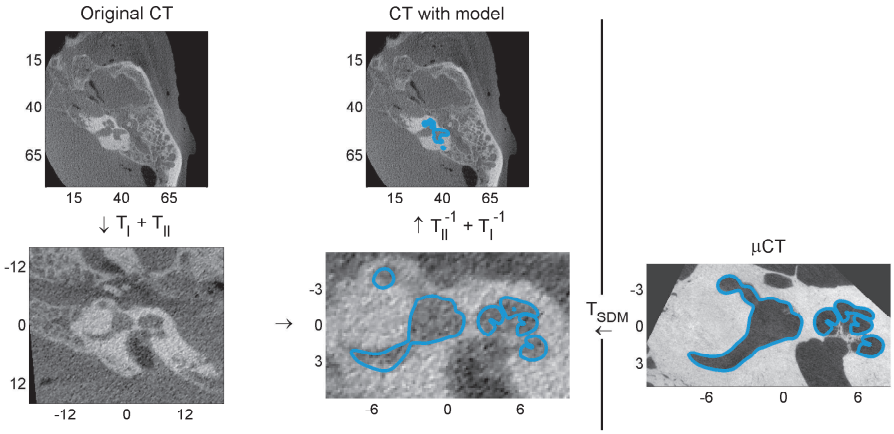


Figure 7.2: Approach for Modelling Patient Specific Anatomy. The three transformations steps of the proposed pipeline are shown. The full head clinical CT image is initially transformed to the space and orientation of the μCT reference (left column). The detailed anatomical shape of the CT image is estimated using a SDM-regularized B-spline transformation (bottom row). The predicted anatomy is projected back to the original CT image (middle column).

7.1.2 The evaluation

The key problem with regards to the performance evaluation in the clinical scenario, is the fact that evaluating a prediction of something in high-resolution from data in low resolution is inherently not well defined. The clinical CT simply cannot be considered ground truth for prediction of intra-cochlear anatomy.

For this reason we include a final test of the SDM fitting using the thiel dataset (see Table 4.1) as a set of 14 test images. We perform the SDM fitting on the CT scans of the samples. The μCT scans are co-registered to the CT data using the landmark-based alignment approach described previously (see Section 4.6.1). Assuming a perfect co-registration the fitted surface model can be transferred to the μCT data for validation. Since the thiel-fixed samples do not contain full ground truth segmentations, we can only base our validation on the apex error and the simpler morphometric errors (see Section 4.6.2.2). We manually measure the cochlear length, width and height of the fitted anatomical model and compare it against the measured ground truth. The average cochlear length error, $\bar{d_L}$, is simply defined as,

$$\bar{d_L} = \frac{1}{N} \sum_{i=1}^{N} L^i_{\text{true}} - L^i_{\text{fit}} \tag{7.1}$$

where L^i is the measured cochlear length of the i'th test sample. The error for cochlear width and height use the same equation.

This evaluation is also described in Paper I, where it is the finalizing contribution with regards to the shape modelling effort we have made. The results are also given here in Table 7.1 and illustrated in Figure 7.3.

Qualitatively, the cochlear partition does not seem to fit very accurately, which is not surprising. First, because this structure was not modelled in the SSM (see Section 6.2.2), and secondly because there is no information in the CT image to drive the fitting of the partition. The resulting fit is merely a reflection of how the partition was seen in the reference dataset, and that may not be very representative for all other samples.

The results show the same trends as for the previous cross-validation evaluation of the training data. The apical region is still problematic to fit properly, but generally the basal and middle turn are modelled with good accuracy. The precision of fitting is quite low. However, it should be noted that these metrics were measured manually, so the numbers themselves have a degree of uncertainty. In any case, having more training data would be valuable and increase the predictive strength of the SDM and the fitting.

Table 7.1: Statistics of test data SDM fitting performance. Reported as the mean +/- 1 std.

Type	**Apex Err.** [mm]	**Cochlear Length Err.** [mm]	**Cochlear Width Err.** [mm]	**Cochlear Height** [mm]
Test Data	1.2 ± 0.4	-0.05 ± 0.4	-0.03 ± 0.3	0.2 ± 0.2

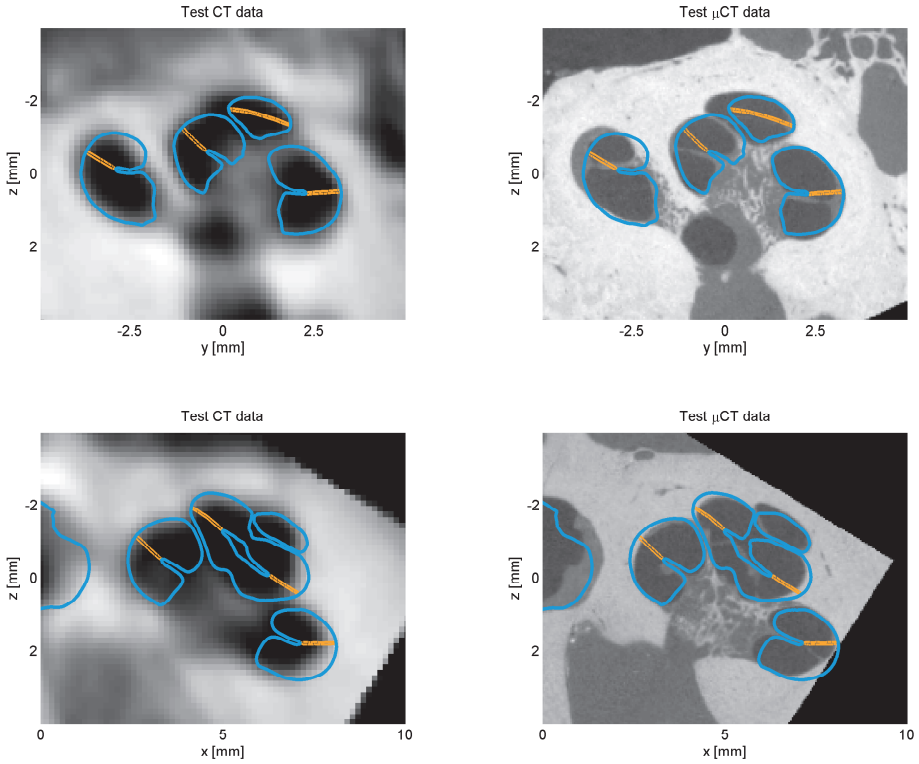


Figure 7.3: Qualitative Validation of SDM Fitting to Test Data. The predicted anatomical shape from two different cases (rows) is shown on top of both the CT and co-registered μCT (columns).

7.1.3 Summary

With the application presented here, we have shown that we can make an estimation of the cochlea shape including intra-cochlear details. Generally the fitted shape is accurate in the first two cochlea turns. The prediction of the cochlear partition placement is not guaranteed to be accurate, as it is a passive structure in the model. The fit in the apical turn is also found a bit wanting.

Despite of these limitations the predicted shape could still be good enough for it to be useful in a range of applications. The usability depends heavily on the required accuracy, which we do not know. Only by testing our procedure in clinical scenarios will we be able to determine and say more about the effectiveness of our model. Such tests are however beyond the scope of this thesis to perform. But given how little anatomical information that is present in a clinical CT scan, then our predicted model must be seen as adding valuable information for understanding the patient specific hearing system.

7.2 Patient Specific CI Programming Model

For many CI-related applications it is generally not enough to only have a model of the patient specific anatomy. Knowledge about the CI placement should naturally also be taken into account.

- The goal is to locate the positions of the CI contact electrodes post-operatively and combine it with a model of the detailed cochlea anatomy.

This combination could be called a patient specific CI programming model. With this, it is possible to calculate at which sound frequency a given electrode is actually placed according to the Greenwood function (Eq. 3.1). This opens up the possibility of modifying CI stimulation strategies and improve the hearing restoration as demonstrated in [NLGD13].

The described procedure for estimating the anatomy was only developed to handle pre-operative images. An actual measurement of the electrode contact positions can obviously only be done with post-operative imaging. So far we have assumed an ideal clinical scenario, where both a pre- and post-op CT image was taken. This is not always the case, and following scenarios can be imagined,

- **Case - Both pre- and post-op CT**
  In this ideal scenario, the anatomy is estimated on the pre-op CT and

transferred to the post-op CT using a rigid transformation [NLMD11].

- **Case - Only post-op CT, but unilateral implantation**
  In this case the anatomy is predicted on the non-operated ear. Assuming a perfect left-right cochlea symmetry, the anatomical model is transferred to the operated ear using a rigid transformation [RML$^+$14].
- **Case - Only post-op CT**
  An alternative approach is to locate the electrode positions in the post-op CT and remove the metallic intensities and artefacts from the image. In this way the 'pre-op' CT is simulated, and the anatomical fit can be made.

No matter the scenario, the electrode positions should be estimated. While it is possible to do manually, we have developed an automatic approach, which is the contribution of Paper J. The positions of the electrode contacts for various MED-EL arrays are shown to be estimated with a mean error of 0.15 ± 0.08 mm.

In order to fit the anatomical model to the data, we had to simulate the pre-op CT. A simple threshold segmentation was used to detect metal voxels, which were then assigned a bulk intensity value corresponding to the intensity in the scalae. This simple processing was found sufficient to allow us to fit the anatomical model to image using the SDM regularized registration.

The accuracy of the electrode position estimation is illustrated in Figure 7.4. This combination is an example of what a patient specific CI programming model would look like. From this it could be determined how the electrodes are placed compared to the Greenwood function (Eq.3.1), which allow for informed modifications to the CI speech programming.

7.3 Simulation of Electrode Array Insertion

There is one scenario that was left out in the previous section. It is possible that no post-op CT was ever taken, which makes it impossible to actually measure the electrode contact positions. However, knowing that during surgery the array insertion behavior is dominated by the geometry and anatomy of the cochlea, it is possible to simulate the insertion, if the anatomy is known (or estimated).

- Based on an predicted cochlea model, the goal is to simulate the surgical insertion and predict the placement of the electrode contacts.

This application was briefly introduced as a contribution of Paper F and the idea of the approach is illustrated in Figure 7.5. Some validation has since then been

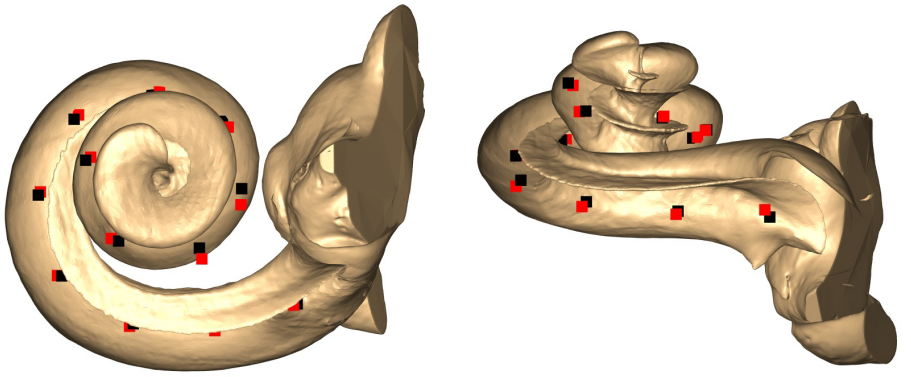


Figure 7.4: Electrode Position Estimation. The fitted anatomical model is shown together with the true (black) and estimated (red) electrode positions.

done, and the preliminary results show that the electrode contact positions can be estimated with an error of 1.8 ± 1.1 mm. The performance of the procedure is naturally heavily dependent on how accurately the anatomy has been predicted. Intra-cochlear structures like the lamina spiralis and the basilar membrane are as a minimum required before the simulation has any resemblance to the reality.

Note, that this simulation can be done purely virtual. Any generated cochlea anatomy can be used, and it is possible to insert electrodes with different properties and sizes.

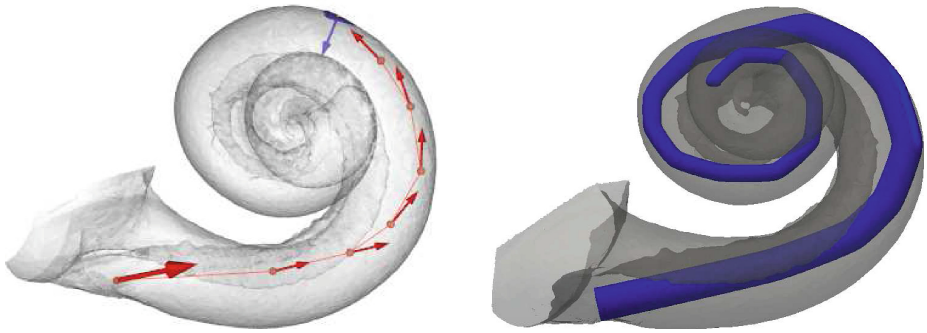


Figure 7.5: Virtual Electrode Insertion. Based on the estimated cochlea anatomy the insertion of the electrode array can be simulated.

7.4 Computational Models for Electrode Stimulation

As described in the previous sections we have procedures for generating accurate models of the cochlea anatomy and the inserted electrode array. An interesting application is to build this into realistic computational models for further virtual simulations. This could for instance be simulations of the cochlea fluid dynamics, but here we focus on the electrical stimulation of the nerves.

- Based on the predicted model of the cochlea anatomy, combined with a model of the electrode array, the goal of this application is to set-up a pipeline for virtual simulations of electrode stimulation.

The generation of the computational model and electrode stimulation protocols are the main contributions presented in Papers E and G. Before the model is transformed into a volumetric mesh, nerve fibers are inserted from the edge of the lamina spiralis, and the cochlea is embedded in a box representing the surrounding bone (Figure 7.6).

Running a defined stimulation protocol for all electrodes can for instance reveal which them are prone to generating cross-talk. Ideally, an electrode should only activate a few nerves to generate one particular sound frequency.

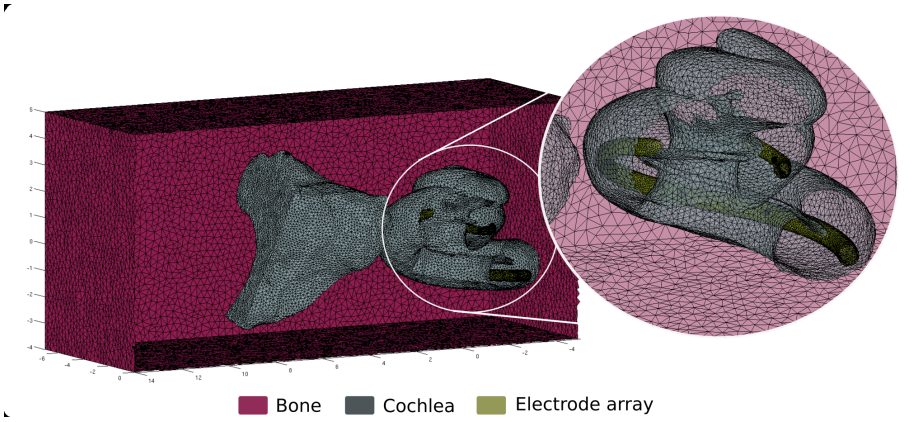


Figure 7.6: Computational CI Model. An instance from any of the developed shape models, can be made into volumetric a computational model.

An important thing to note is, that the computational models can be generated based on patient-specific information, but can also be completely virtual. Be-

cause of the generative property of the PCA (see Section 6.1.1), it is perfectly possible to sample anatomical instances with the SDM and perform computational studies on the population described in the shape model.

Patient specific level: Given a patient specific situation, the computational model can for instance be used to identify likely electrodes with cross-talk zones, and modify the CI programming accordingly. Cross-talk is known to degrade the hearing restoration outcome, but without a patient specific computational model like this, it is hardly possible to deal with in a sensible manner. These models therefore have the great potential for improving hearing restoration for the individual CI-user.

Population level: By generating a population of computational cochlea models, it becomes possible to investigate for instance the designs of the electrode arrays. Whether to make them longer or shorter, change the electrode spacing etc. have been difficult questions to answer previously. The use of a computational model like this could be valuable for the CI manufacturers for virtual product development and testing.

7.5 Summary and Conclusion

With the CI applications presented here, we have given a perspective on what our generated statistical shape model can be used for. This accomplishes the final thesis goal.

Each of the applications helps to increase our understanding of the variables governing the hearing system and the CI-device.

- On a patient specific level, we can predict the anatomy of the cochlea with certain limitations. Combining it with either a prediction (virtual insertion) or measurement of electrode array positions, provides a patient specific CI programming model.
- On a population level. Anatomical shapes representing the population variability can be sampled from the shape model. With virtual computational models is becomes possible to investigate electrode design parameters, electric stimulation protocols, asserting the risk of causing trauma during surgical insertion etc.

These applications show a promising potential for us to better understand, control and optimize the CI. This should ultimately result in an improved hearing restoration for CI-users.

Chapter 8

Conclusions

The fundamental goal of the thesis concerns improvements of hearing restoration based on Cochlear Implants (CI). Closing the gap between natural- and artificial hearing would improve the quality of life for a large number of people and it would be a great feat for biomedical engineering.

In theory, the concept of the CI is very simple. The inner ear is a remarkable structure, which is basically a sound frequency analyzer. There is a direct relationship between anatomical position within the cochlear and the perceived frequency. The CI device is designed to exploit this relationship, by placing a number of electrodes within the cochlear. Each electrode encode for a specific frequency based on its position and is activated accordingly, thereby artificially generating the sensation of hearing.

In practice, the system it a lot more complicated. From an engineering point of view it is a system riddled with unknowns and uncontrolled variables spanning multiple disciplines. There is the design and programming of the CI device, the surgical insertion procedure and the patient specific anatomy and physiology. An understanding of each them and their interdependencies is required before the system can be fully controlled. In the current clinical practice, it is remarkable how little system information is actually used to make decisions.

The thesis goal can be achieved by providing relevant and usable knowledge

to allow for informed decisions. The source of this information is imaging. To control the patient variables it is crucial to know the detailed patient specific anatomy. The only available data source for this is CT images, which in themselves carry very little useful anatomical information. We therefore have to develop methods to infer and predict the shape of the cochlear from the low resolution images. To control design and implant variables we need to have realistic and detailed computational models of the cochlea including the population variability. This can then be used for virtual product design and testing either on patient specific cases or more population based. The data source for realizing this, is μCT imaging of ex-vivo samples.

Statistical shape models offers the possibility to address both concerns within the same methodology. The practical thesis goal is therefore to construct the best possible shape models of the cochlea or the inner ear. This is broadly speaking dependent on three different aspects, which we deal with to varying extends in this thesis. One is the data quality and quantity. Secondly, there is the process of constructing the shape model. Third and finally there is the model itself and how you can and intend to use it.

Since μCT data is a limited source, it is crucial to make the most of it. We have made accurate segmentations and surface models to have good representations of the anatomy in the data. We have characterized the shape and developed methods for parameterizing it to provide high quality inputs to the model building process. The methods are not necessarily perfect, but we have demonstrated that they could significantly facilitate and improve our image registration models.

The primary challenge concerning construction of shape models is the methodology for creating point correspondences between data, and it is an area that was explored extensively. We have managed to identify and address different concerns, and have in the thesis shared our experiences. The biggest challenge of registering cochlea shapes is arguably the difficulty in providing a proper global initialization. We have made two research contributions exploring how to incorporate global structure similarity into a non-rigid registration model. We have demonstrated that our efforts improves both accuracy and precision of our registrations and our constructed shape models. Further research could most likely improve the model even more, but given the limitations in quality and quantity of the data, we believe that we have managed to make the most of it.

We have developed a procedure for predicting the patient specific anatomy from CT images, using our shape model to provide the required regularization to solve the problem. We have further identified potential and beneficial applications of our models. We have yet to show if and how much we can improve the restoration of hearing from a CI. However, we can now provide a detailed anatomical

computational model with a description of the population variability. It may not be perfect, but it allow us to do realistic studies of population anatomy and opens up for virtual product design and testing. We can estimate intra-cochlear details from clinical CT scans, thereby adding valuable information and knowledge to the patient specific system of the CI-hearing. It has certain limitations, but that will always be the case, so only further clinical tests will be able to determine the usability.

The tools, methodology and shape models that we have developed constitutes a promising step in the direction of controlling different variables of the hearing system, which ultimately could improve the restored hearing from Cochlear Implants.

Appendix A

Free-form Image Registration of Human Cochlear μCT Data Using Skeleton Similarity as Anatomical Prior

Pattern Recognition Letters 000 (2015) 1–7

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Free-form image registration of human cochlear μCT data using skeleton similarity as anatomical prior☆

H. Martin Kjer[a,*], Jens Fagertun[a], Sergio Vera[b,c], Debora Gil[c,f], Miguel Ángel González Ballester[d,e], Rasmus R. Paulsen[a]

[a] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Copenhagen, Denmark
[b] Alma IT Systems, Barcelona, Spain
[c] Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain
[d] Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
[e] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
[f] Serra Húnter fellow

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Image registration
Regularization
Skeleton
Cochlea
Inner ear

ABSTRACT

Better understanding of the anatomical variability of the human cochlear is important for the design and function of Cochlear Implants. Proper non-rigid alignment of high-resolution cochlear μCT data is a challenge for the typical cubic B-spline registration model. In this paper we study one way of incorporating skeleton-based similarity as an anatomical registration prior. We extract a centerline skeleton of the cochlear spiral, and generate corresponding parametric pseudo-landmarks between samples. These correspondences are included in the cost function of a typical cubic B-spline registration model to provide a more global guidance of the alignment. The resulting registrations are evaluated using different metrics for accuracy and model behavior, and compared to the results of a registration without the prior.

1. Introduction

Image registration and establishment of correspondences between data is a common challenge in biomedical image analysis. The best registration model is largely case-dependent, influenced by the anatomy, the involved imaging modalities, the desired end-goal, etc. [11,17,18]. In cases that require large and complex deformations finding the optimal registration procedure becomes a challenging task. As the amount of parameters in the transformation model increases it becomes more and more difficult for the optimization to avoid local minima. In these cases, it is often required to include some additional prior knowledge or regularization/constraints to efficiently solve the registration.

The challenging case presented in this paper is the task of registering data of the (human) inner ear. This structure controls the sensation of hearing and balance, and an understanding of the anatomy and anatomical variability plays an important part in utilizing the full potential of Cochlear Implants [30]. Detailed anatomical models have interesting patient-specific applications as they can provide information about the type of electrode design that suits the anatomy of the user [29], or by allowing improvements to the implant programming based on simulations mimicking the actual anatomical and physiological situation [6].

The anatomy of the inner ear is composed of the vestibular system and the cochlea. Image registration of the latter is challenging for a couple of reasons, and will be the focus of this work. The human cochlea is a spiral structure with outer dimensions of approximately $10\times8\times4$ mm. The size and the shape of the spiral can vary extensively. On average, the cochlea winds 2.6 turns [9] but can approach up to three full turns – corresponding to a difference in the order of 1–2 mm following the path of the spiral. The separation between the cochlear turns is typically one order of magnitude smaller. Specially deformations to properly align the most apical region of spiral have been difficult to model to our experience. Further, the whole spiral is a tube-like structure (see Fig. 1, right) with a large degree of self-similarity in the cross-sections. This lack of distinct features makes it difficult to identify corresponding anatomical positions across samples.

The desired registration model should not just expand or compress the apical part of the spiral to align two samples, but rather

☆ This paper has been recommended for acceptance by Punam Kumar Saha.

* Corresponding author. Tel.: +4545255228.
E-mail address: hmkj@dtu.dk (H.M. Kjer).

http://dx.doi.org/10.1016/j.patrec.2015.07.017

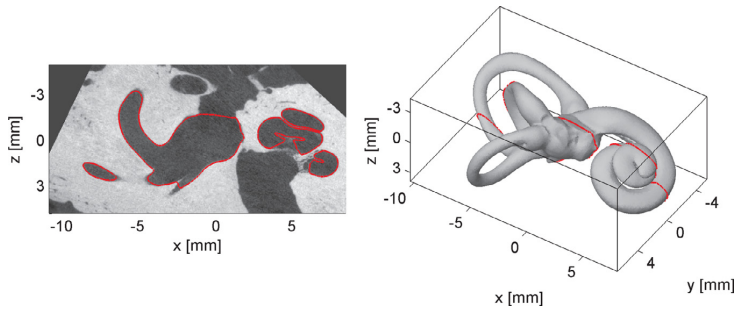


Fig. 1. Left: Impression of the μCT data and segmentation (red border). Notice the small spacing separating the cochlear turns (right side of CT image), the weak contrast towards internal cochlea borders, and the opening into the middle ear cavity (middle of the image). Right: The corresponding surface model provides an overview of the inner ear topology. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

model a change along the entire spiral. Essentially the model should be able to handle very local deformations while still adhering to the global structure of the samples. This type of behavior is usually not native to non-rigid registration models without some kind of prior or regularization included.

Modifications to a registration model to include such prior knowledge have been studied previously. A way of introducing anatomical shape priors is the use of a statistical shape model [4,10]. However, building statistical shape models is in itself a labor intensive task rivaling if not surpassing the task of the registration, as the prerequisite for building the model is data that is already registered to have correspondences.

A multitude of physical constraints have also been proposed as regularizations. For example, local tissue rigidity can be enforced in specified areas [23], or conditions of incompressibility or volume-preservation can be applied [20]. However, finding the suitable physical constraint for a registration task is not straightforward, as this is case- and application dependent.

In the work of [3] an articulated skeleton model was pre-registered to intra-mouse data studies in order to recover large pose-differences between data acquisitions. The presented application is narrow in its scope, but the registration methodology of using landmark correspondences as regularization is more generally applicable, thus we adopt this approach for this work.

In this paper we explore the potential of using the skeleton of a surface object as an anatomical prior in free-form registrations using a B-spline transformation model.

Skeletonization of a volume or surface is a research field in itself [7,22]. The skeleton provides a global description of shape in a simplified and structured form. Matching based on skeleton similarity could provide a global anatomical guidance or regularization to a locally defined free-form image registration procedure with a high resistance to noise compared to using only the image intensity similarity. The use of skeleton similarity in image registrations should be applicable to many different problems and there are many published methods and approaches for finding and matching the skeletons for differing types of data and geometries [24,26]. Skeleton correspondence has been seen in image registration tasks before, relating to for instance 2D/3D multi-modal registration [15] and matching of vessels in time-series angiography data [27]. More related to our approach is the work of [25], where multiple different shape features were calculated from surface objects and transformed into vector-valued 2D feature images, which were aligned with a classic image registration formulation. Skeleton features were used for global alignment in the coarser levels of the registration. Our strategy is similar although the prior will be included into the registration model differently.

The purpose of this study is to test and evaluate deformable registration using a B-spline transformation model on a series of inner ear datasets with/without the use of skeleton-based similarity in the registration model.

The rest of the paper is structured as follows. Section 2.1 provides a description of the data and the processing, and Section 2.2 contains the procedure for finding skeletons and their similarity across datasets. The registration models and their evaluation are detailed in Sections 2.3 and 2.4 respectively. The results are presented in Section 3 and discussed in Section 4.

2. Material and methods

2.1. Data and processing

A collection of 17 dried temporal bones from the University of Bern were prepared and scanned with a Scanco Medical μCT100 system. The data was reconstructed and processed to obtain image volumes of 24 micron isotropic voxel-sizes containing the inner ear (Fig. 1, left).

Image segmentation: The border of the inner ear was segmented in all datasets semi-automatically using ITK-SNAP [31]. On standard CT images the cochlear will appear to have a circular cross-section. Due to the higher resolution of μCT and the sample preparation it becomes possible to see the *lamina spiralis*. It is a bony ridge structure that traverses the entire cochlea from the spiral central direction, partially separating the cochlear into two chambers (*scala tympani* and *scala vestibuli*) and creating a 'U'-shaped cross-section (see Fig. 1, left). The semi-automatic tool in the segmentation software was critical for achieving smooth and rounded segmentations in data with that kind of resolution, and for reducing the amount of manual work. The images contain some openings, less well-defined regions and non-anatomical artifacts that had to be manually handled to obtain comparable segmentations across datasets. For this reason a segmentation of one dataset easily amounts to 12–15 h, but it is a requirement for having a ground truth and a correct representation of the object. A surface model was generated for each dataset using Marching Cubes [16] followed by a surface reconstruction [19] to obtain a well-formed triangular mesh (Fig. 1, right).

2.2. Skeletonization

Implementing and comparing skeletonization methodologies is not the scope of this work. Our aim is to find a simple method for generating skeleton correspondence between samples that can easily be included in a registration model thus allowing us to explore the potential of the approach. The object topology is an important consideration when working with skeletons, as this poses a restriction for certain methods. To avoid working with a genus 3 surface, we exclude the vestibular system and focus only on a skeleton of the spiral shaped cochlea.

Please cite this article as: H.M. Kjer et al., Free-form image registration of human cochlear μCT data using skeleton similarity as anatomical prior, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.07.017

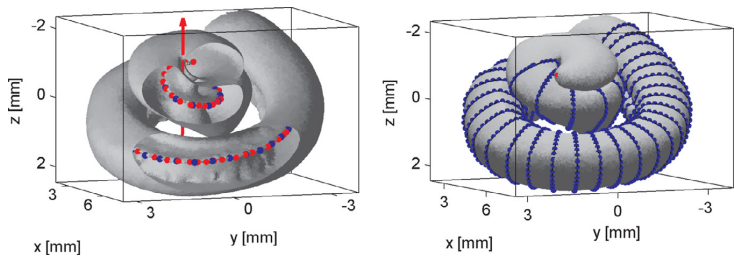


Fig. 2. The cochlear skeletonization. Red annotations are manually determined information (Left: $\mathcal{Z}^{\text{Man}}$ and central axis, Right: cochlear apex). Blue annotations represent parametric pseudo-landmarks (Left: $\mathcal{Z}^{S}$, Right: $\mathcal{Z}^{\text{LM}}$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Initial experimentation showed that methods for automatically finding the curved skeleton [7], medial curve or centerlines [5] tended to be attracted to the larger of the two scalae and thus creating inconsistent skeletons across samples. Medial surface representations [12,28] were also found challenging to generate with desirable topological consistency and smoothness. To keep things simple we propose to use a set of corresponding pseudo-landmarks, $\mathcal{Z}^{\text{LM}}$, of the cochlear surfaces obtained from a parametric 'curved skeleton' following the *lamina spiralis* ridge.

We manually define the following information (marked red on Fig. 2) in each dataset to determine our corresponding pseudo-landmarks: The cochlear apex landmark (A_i), the central axis of the cochlear spiral (defined by a unit direction vector, $\vec{n}_i$, and an axis point, C_i), and a set of points (approx. 100 per dataset) following the ridge line of the *lamina spiralis* from the beginning of the cochlear base to the end in the apical turn ($\mathcal{Z}_i^{\text{Man}}$).

From the above information, we generate a naive parametric model of the cochlea. First, we create a parametric description of the cochlea skeleton using $\mathcal{Z}_i^{\text{Man}}$, from which we determine 37 corresponding positions on the skeleton with equal arc-length ($\mathcal{Z}_i^{\text{S}}$). Secondly, we extract planar surface cross-section at each of the points, p, in $\mathcal{Z}_i^{S}$. The cross-section plane is determined by the tangent of the skeleton at p. Each cross-section of the surface mesh is then parameterized using 40 points ($\vec{n}_i$ provides the reference for orientation and starting point in each cross-section parameterization). These cross-sectional points together with the apex landmark provides a set, $\mathcal{Z}_i^{\text{LM}}$, of 1481 corresponding surface pseudo-landmarks (Fig. 2, right) to be included in a registration model. Finding the cochlea cross-section in the apical region of the cochlear can potentially lead to some ambiguity, as they could intersect with themselves. To avoid this the skeleton cross-sections in the apical turn were not included.

2.3. Image registration

The registration procedure follows a common work-flow. One dataset was chosen as the reference, to which the remaining moving datasets were registered in two steps - rigid initialization followed by the deformable registration, both detailed in the following subsections.

2.3.1. Initial rigid alignment

There are many approaches for finding rigid transformations. The chosen procedure is independent from the skeleton information and is the same no matter the chosen deformable registration model. In that way, later comparisons of registration results are not affected by the initialization. The whole initialization procedure relies solely upon the extracted surface meshes, but the calculated rigid transformations were also applied to the gray-scale volumes and their segmentations.

Translation: Let $p_{(i,j)}$ be the jth vertex position of dataset i. A translation was applied so that the center of mass is placed in position (0,0,0), i.e. the mean vertex position, $\bar{p}_i$, was subtracted from all vertices. This places all datasets in a coordinate system where the inner ear center of mass of each dataset is in the origin.

Rotation: Let Σ_i be the 3×3 covariance matrix of the mesh vertex positions of dataset i (after the translation). The eigenvectors, W_i, of Σ_i provides a rotation matrix, which when applied transforms the data to the principal component directions. This essentially corresponds to fitting an ellipsoid to the point cloud and aligning the axes.

Check directions: This alignment procedure is robust due to the asymmetry of the inner ear shape (Fig. 1, right). However, the sign of a principal direction in the ith dataset could potentially be opposite compared to that of the reference. To handle this we make a simple check. The bounding box of the reference and of the moving point cloud is divided into a coarse grid. We use the sum of squared grid vertex-density difference between the two as a check metric. If the axis-flip would result in a lower metric, then the flip is made to the moving dataset. While there is no guarantee for this to work in all cases, it has worked well for our data. In principle, any kind of rigid alignment could be used instead of the one suggested here.

2.3.2. Deformable registration

The non-rigid image registration follows the formulation and framework of `elastix` [14].

The registration is done between the segmentations rather than the gray-scale volumes for two reasons. First, the μCT data contain smaller artifacts and certain weakly contrasted edges, that were handled during the segmentation. Secondly, the registration should not be influenced by the anatomical differences in the surrounding bone structure.

The registration of the moving dataset, I_M, towards the reference, I_F, is formulated as a (parametric) transformation, T_μ, where the vector μ containing the p-parameters of the transformation model are found as an optimization of a cost function, $\mathcal{C}$.

$$\hat{\mu} = \arg\min_{\mu} \mathcal{C}(T_\mu, I_F, I_M) \tag{1}$$

The transformation model used in this paper is the cubic B-spline in a multi-resolution setting. We apply image smoothing with a Gaussian kernel to both the fixed and moving image. For each level of resolution the spacing between grid points and the width of the smoothing kernel follows a decreasing scheme, starting with a coarse registration that is gradually refined. The following scheme was chosen by experimentation:

Control point grid spacing (isotropic, voxels):

[144, 72, 48, 48, 36, 24, 18, 12, 6]

Width of Guassian kernel (isotropic, voxels):

[10, 10, 1, 1, 1, 1, 1, 1, 1]

The width of the kernel was deliberately kept narrow in most levels to avoid that small and sharp features would be blurred out (for instance the separation of the cochlear turns).

Please cite this article as: H.M. Kjer et al., Free-form image registration of human cochlear μCT data using skeleton similarity as anatomical prior, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.07.017

The cost function used in this 'basic' registration set-up:

$$\mathcal{C}_1 = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, I_F, I_M) + (1-\alpha) \cdot \mathcal{P}_{\text{BE}}(\mu) \quad (2)$$

where α is a weight parameter in the interval [0,1]. The similarity term, $\mathcal{S}_{\text{Sim}}$, is chosen as the sum of squared differences (SSD). The term $\mathcal{P}_{\text{BE}}$ is the energy bending regularization used to penalize strong changes and foldings in the transformation [21]. The weighting of the similarity term was chosen to 0.9 by experimentation. Increasing α would provide more freedom for deformation of the shapes, but also increase the risk of having non-plausible anatomical results.

The optimization is solved using Adaptive Stochastic Gradient Descent [13]. The maximum number iterations was set to 2500. To reduce the computational burden of the optimization only a subset voxels are sampled for the evaluation. For each iteration 2^{14} random coordinate points were sampled. These settings were fixed for all resolutions.

2.3.3. Deformable registration with guidance from skeleton

The free-form registration set-up remains largely the same when a skeleton is included in order to make comparisons fair. The cost function is modified to include a landmark similarity term [3]:

$$\mathcal{C}_2 = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, I_F, I_M) + \beta \cdot \mathcal{S}_{\text{CP}}(\mu, \mathcal{Z}_F, \mathcal{Z}_M) + (1-\alpha-\beta) \cdot \mathcal{P}_{\text{BE}}(\mu) \quad (3)$$

where α and β are weightings in the interval [0,1] and fulfilling $\alpha + \beta \leq 1$. The landmark similarity term, $\mathcal{S}_{\text{CP}}(\mu, \mathcal{Z}_F, \mathcal{Z}_M)$, uses the Euclidean distance between the set of corresponding landmarks, $\mathcal{Z}_F$ and $\mathcal{Z}_M$. In this way intensity-based image registration is guided with features extracted from the anatomical skeleton (i.e. using $\mathcal{Z}_i^{\text{LM}}$ from Section 2.2). By experimentation the weightings were set to $\alpha = 0.8$ and $\beta = 0.11$. The landmark similarity is kept small in order not to force the alignment, and the ratio between image similarity and bending energy regularization is kept similar to the previous set-up $\mathcal{C}_1$ (Eq. (2)). Settings for the transformation model and optimizer were unchanged from the previous registration model.

2.4. Evaluation

We are interested in comparing the 16 registration results of model 1 (Eq. (2)) and model 2 (Eq. (3)) using a number of different image and mesh based metrics.

Image based evaluation: Let $I_i(\mu)$ be the moving segmentation volume after application of the resulting transformation. We compare the Dice Score [8] to the segmentation of the reference dataset, I_{Ref}.

$$\text{DSC} = \frac{2 \cdot \left| I_{\text{Ref}} \bigcap I_i(\mu) \right|}{|I_{\text{Ref}}| + \left| I_i(\mu) \right|} \quad (4)$$

Mesh based evaluation: We define the surface based scores as follows. Let $\mathbb{S}_{\text{Ref}}(\mu)$ be the reference surface mesh after application of the resulting transformation. There is no direct point correspondence between the reference and the ground truth surfaces, $\mathbb{S}_i$, and they each contain a varying number of vertices. Metrics are therefore based on the closest points, i.e. the minimum Euclidean distance from a point, p, to any of the points, q, in the other surface, $\mathbb{S}$:

$$d(p, \mathbb{S}) = \min_{\forall q \in \mathbb{S}} (||p - q||_2) \quad (5)$$

The mean surface error, $d_{\bar{s}}$, of each sample is defined as the average of all the closest point distances:

$$d_{\bar{s}} = \frac{1}{N_{\text{Ref}} + N_i} \left(\sum_{\forall p \in \mathbb{S}_{\text{Ref}}(\mu)} d(p, \mathbb{S}_i) + \sum_{\forall p \in \mathbb{S}_i} d(p, \mathbb{S}_{\text{Ref}}(\mu)) \right) \quad (6)$$

where N_{Ref} and N_i are the total number of points in the reference and the moving surface respectively.

The Hausdorff distance, d_H, is the maximum of all the closest point distances:

$$d_H = \max \left\{ \max_{\forall p \in \mathbb{S}_{\text{Ref}}(\mu)} d(p, \mathbb{S}_i), \max_{\forall p \in \mathbb{S}_i} d(p, \mathbb{S}_{\text{Ref}}(\mu)) \right\} \quad (7)$$

The above mentioned metrics are very generic and will hardly be able to reflect and evaluate the change in the registration model that we intend to explore. We therefore include two additional scores, apex error and torque.

First, we calculate the Euclidean distance between apexes of the target data and of the reference.

$$d_A = \left|\left| A'_{\text{Ref}}(\mu) - A_i \right|\right|_2 \quad (8)$$

The apex is one of the few locations on the cochlea that can be placed relatively precisely. Even though an arc-length distance might be more correct to use, the Euclidean apex error should be indicative of the registration model behavior in the apical region, even though this point is also included in the registration model.

Secondly, we look at the differences in the vector deformation fields obtained by the registration models. The cochlear samples have a different number of turns, and we wish to evaluate the registration models ability to capture this rotational behavior of the anatomy. Our postulation and assumption is that this ability of the registration model should correlate with the 'torque', τ, on the central axis of the cochlear exerted by the deformation field.

Let the force vector, $\vec{F_p}$, on the vertex, p, in the reference mesh be defined simply as the vector between the vertex position before and after application of the registration transformation:

$$\vec{F_p} = p(\mu) - p$$

Further, we can calculate the perpendicular arm from the central axis to the mesh vertex, $\hat{v}_p$. This vector is normalized to unit length, so that the vertices farther from the axis will not contribute with a greater torque.

The scalar projection of the force vector, F_p, onto the unit arm that is perpendicular to both the central axis and $\hat{v}_p$ is then the acting force contributing to the torque:

$$F_p = \vec{F_p} \cdot (\vec{n} \times \hat{v}_p)$$

Using this local vertex torque force leads to our definition of the torque of the registration:

$$\tau = \frac{1}{N_{\text{Ref}}} \sum_{\forall p \in \mathbb{S}_{\text{Ref}}} F_p = \frac{1}{N_{\text{Ref}}} \sum_{\forall p \in \mathbb{S}_{\text{Ref}}} (p(\mu) - p) \cdot (\vec{n} \times \hat{v}_p) \quad (9)$$

3. Results

The registrations were done on a desktop with a quad-core 3.6 GHz processor, 64 GB RAM, running elastix v4.7. The average time per registration was approximately 0.8 h and we observed no notable

Table 1
Statistics of registration evaluation metrics, reported as the mean +/− 1 std. Model 1 is the non-rigid image registration model and model 2 the non-rigid image registration model incorporating a skeleton prior.

Metric	Dice score	Surface error [mm]	Hausdorff [mm]	Apex error [mm]	Avg. torque [mm^2]
Model 1	0.96 ± 0.01	0.040 ± 0.01	0.69 ± 0.24	1.01 ± 0.59	−0.04 ± 0.09
Model 2	0.95 ± 0.01	0.045 ± 0.01	0.73 ± 0.35	0.69 ± 0.52	−0.53 ± 0.28

Please cite this article as: H.M. Kjer et al., Free-form image registration of human cochlear μCT data using skeleton similarity as anatomical prior, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.07.017

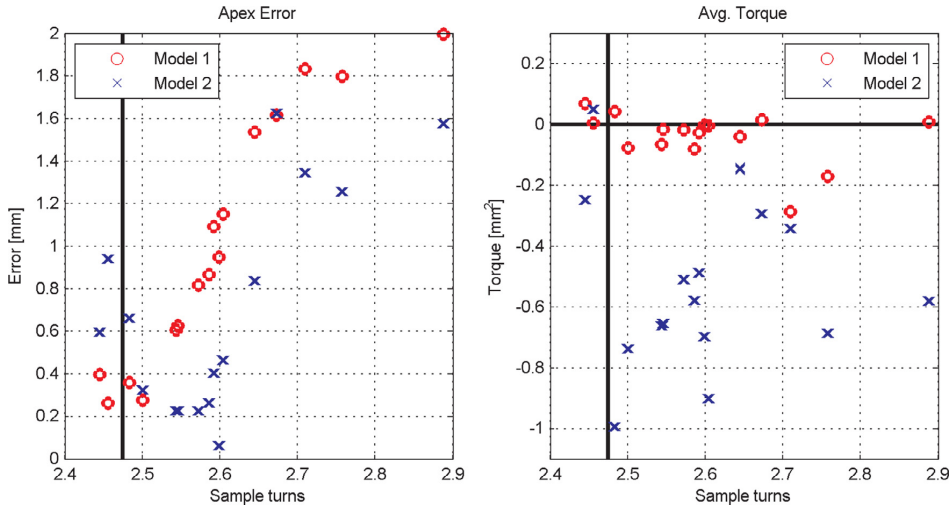


Fig. 3. Sample-wise apex error (Left) and average torque (Right) plotted against the number of cochlear turns of the target samples. Vertical black line indicate the number of turns in the reference sample.

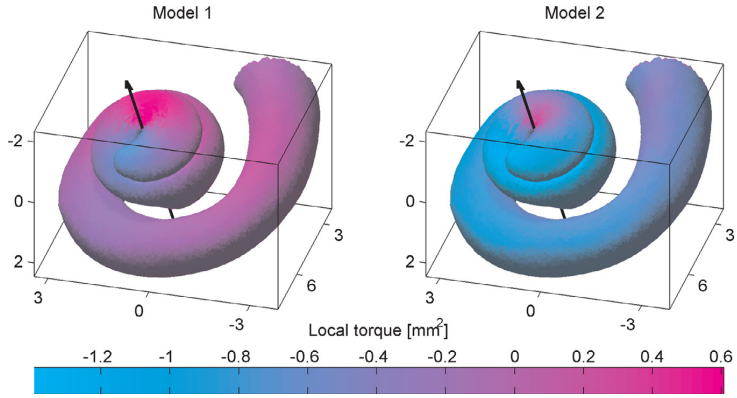


Fig. 4. Qualitative difference in the local torque acting on the cochlea central axis (black vector). The target sample has 2.60 turns, compared to the 2.46 of the reference (the shown surface). Positive direction of the central axis is defined from the cochlea base towards the apex.

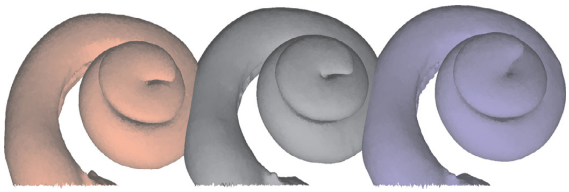


Fig. 5. The visual difference between registration models. The reference surface is deformed using either model 1 (purple, right) or model 2 (red, left) to align with the target sample (grey, middle). The surfaces have been moved apart to avoid overlap between shapes.

difference in run times or convergence speed between the two registration models.

The statistics of the different metric scores are presented in Table 1. Fig. 3 elaborates on the sample-wise apex error and torque metric, and Figs. 4 and 5 show the qualitative difference between the registration models.

The general metrics (DSC, $d_{\bar{s}}$, d_H) show a small decrease in performance accuracy for model 2.

From Fig. 3 it is observed that the apex errors of model 1 grow more or less proportionally to the discrepancy in cochlear turns. The torque is close to zero on average. These observations reflect that model 1 only adapts very locally and behaves indifferently with regards to the turning of the target shape. I.e. the resulting cochlear shapes after registration have little variation in the turns.

The apex errors are seen to be generally lower for model 2. Note, that the apex landmark used to calculate this error was a part of the optimization procedure. That the error is reduced is therefore no surprise and it is a biased metric for considering the model accuracy and precision. However, it provides a summarizing pseudo-measure of how much more turning registration model 2 on average is able to capture, which is further illustrated in Fig. 5. For very large differences in cochlear turns it would seem that both of the registration models have trouble with aligning the apexes.

The torque of model 2 is in most of the cases negative. This indicate vector fields pointing more tangentially in the direction of the spiral towards to the apical region. This would be the expectation as most of the target samples have more turns than the reference. The torque is not a measure of accuracy nor precision. The torque merely

Please cite this article as: H.M. Kjer et al., Free-form image registration of human cochlear μCT data using skeleton similarity as anatomical prior, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.07.017

provides a simple quantification of the overall rotation of the cochlear shape. Further is gives a good way of illustrating the differences between the registration models as demonstrated in Fig. 4.

4. Discussion

The Dice Score, surface error and Hausdorff distance serves as very general metrics for evaluating the local adaptability of the registration models. Further, they indicate the general accuracy and precision that we are achieving with the data. The performance with model 2 was decreased on these scores. It would seem that we are trading some local adaption for guiding the model with the landmarks. The determination of the skeleton inherently carries some uncertainties. By introducing the landmarks into the registration model extra noise is added to the procedure. It may happen that a poor skeleton estimate is drawing the spiral in the wrong direction. By providing a more robustly determined skeleton that additionally could fully reach the most apical turn, we expect that the performance of model of 2 could be increased.

The current approach by basing the skeletons on manually picked medial points is only feasible because of the low number of data samples, and because of the specific anatomy where the ridge of the *lamina spiralis* is easily identifiable and happens to define a medial line of the object. Manually placing medial points in a 3D model would generally be impossible.

Aspects of the skeletonization and its influence can be studied furthermore. For instance the number of landmarks used to represent the skeleton. By experimentation we found an amount of cross-sections that appeared to work, but the number of landmarks per cross-section could potentially be reduced. However, the primary concern is the current lack of information in the most apical cochlear turn. For this to be included it would be interesting to look into other skeletal representations. That would in turn potentially require a different way of measuring the similarity of skeletons and possibly an extension to the registration framework to accommodate this. It holds an interesting research potential as both the field of skeletonization and image registration are well-researched areas, but so far joining the two have received little focus. A reason might be the challenge in automatically obtaining consistent skeletons from volumetric data. In this work the skeletons were based on the surface models (i.e. the data segmentation), which in many cases are difficult and/or time-consuming to obtain. Ideally the skeletons should be extracted from volumetric gray-scale data similar to the work of [1,2].

Using the B-spline grid as the transformation model in the registration has limitations. Choosing a fluid- or optical flow-based model [18] could potentially be more suited for this kind of spiral anatomy. Alternatively, the performance of the B-spline approach could perhaps be improved with some data preprocessing. If the cochlea was unfolded, possibly based on the skeleton cross-sections, it would be in a space more suited for a B-spline grid transformation. Along the same line of thinking, the deformation control points could be placed in a non-cubic grid structure favoring the spiral nature of the data. However, these suggestions may be difficult to realize and involves adapting the registration method to one very specific task or anatomy. In this and potentially other cases finding a skeleton and including it into the a registration model may be an easier or more feasible approach. The results reflect that it is possible to modify and regularize the registration by using skeleton similarity as a prior, even though there is room for improvements in our methodology.

The registration parameters used in this work were manually determined. A set of parameters that works well on all data samples while running within a reasonable time frame can be difficult to find. Regarding the choice of metric weights, an interval of $\alpha = 0.7 - 0.9$, would seem to be the most appropriate for model 1. Higher α increases the flexibility of the model, which is needed for capturing the cochlear turning. However, increasing beyond 0.9 made some cases fail. In particular the behavior of the deformations in the semi-circular canals performed poorly. The same holds true for model 2. For having a fair comparison between the registration models, the same relative weight of the image similarity and bending energy metric was kept. Having $\beta < 0.15$ was found to be reasonable. Forcing more weight on the landmarks could result in too strong deformations in some cases, and going much lower counters the idea of having the landmarks. Variable metric weights throughout the resolutions were also tested for model 2. I.e. a scheme where a strong weighting was placed on the landmarks in the initial resolutions and then gradually reduced. It worked well in some cases only, so to keep the registration models comparable the fixed weightings scheme was used. Regarding the optimization only the default optimizer and automatically determined settings were used. A number of samples in the range of $2^{14} - 2^{17}$ and a maximum number of iterations between 1000 and 2500 seemed to produce stable results. Tweaking of registration parameters could result in minor changes of the performance scores, but the same tendencies of the registration models would be observed.

The local torque forces (Fig. 4) provides the most qualitative view of the differences between the registration models. There is no ground truth torque, but it illustrates that the normal registration model is very local in its adaption, whereas model 2 provides more turning in the region where the skeleton is defined. Ideally we could have shown a more convincingly stronger negative correlation (Fig. 3) between the differences in the cochlear turns and the average size of the torque. However, we have a low number of samples and the registration also has to deal with general differences in the size and orientation of the samples apart from the turning. In future work the torque could perhaps even be used as a regularization in the registration model, where it could favor a constant torque in the B-spline grid points near the spiral.

5. Conclusion

We have demonstrated the use of parameterized skeletons to act as an anatomical similarity prior for cubic B-spline non-rigid image registration of cochlea μCT data.

The B-spline transformation model is only locally defined, and the registration is challenged by the spiral nature of the cochlear data, that is locally similar throughout all turns. We have created a simple parameterized skeleton, and included corresponding parametric pseudo-landmarks into the registration cost function. The skeleton provides global similarity to the registration model, that allows cochlear turning to be captured to a larger degree.

Developments in the use of skeletons in the non-rigid alignment could lead to better registration models. This requires further work in combining research in skeletonization with the field of image registration.

Acknowledgments

HEAR-EU is a collaborative project between Alma Medical Imaging, Med-El, Scanco Medical AG, the Univ. of Bern, Univ. Pompeu Fabra and the Tech. Univ. of Denmark. We would like to thank Scanco and the University/University Hospital of Bern for acquiring and providing the data.

The research leading to HEAR-EU results has received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement n 304857.

Debora Gil is a Serra Húnter fellow and supported by Spanish Government Project TIN2012-33116.

References

[1] S.S. Abeysinghe, M. Baker, W. Chiu, T. Ju, Segmentation-free skeletonization of grayscale volumes for shape understanding, in: Proceedings of IEEE International Conference on Shape Modeling and Applications 2008, 2008, pp. 63–71.

[2] E. Antúnez, L. Guibas, Robust extraction of 1D skeletons from grayscale 3D images, in: Proceedings of the International Conference on Pattern Recognition, 2008.
[3] M. Baiker, M. Staring, C.W.G.M. Löwik, J.H.C. Reiber, B.P.F. Lelieveldt, Automated registration of whole-body follow-up microCT data of mice, Med. Image Comput. Comput. Assist. Interv. 14 (Pt 2) (2011) 516–523, doi:10.1007/978-3-642-23629-7_63.
[4] F.F. Berendsen, U.A. van der Heide, T.R. Langerak, A.N.T.J. Kotte, J.P.W. Pluim, Free-form image registration regularized by a statistical shape model: application to organ segmentation in cervical MR, Comput. Vision Image Underst. 117 (9) (2013) 1119–1127, doi:10.1016/j.cviu.2012.12.006.
[5] S. Bouix, K. Siddiqi, A. Tannenbaum, Flux driven automatic centerline extraction, Med. Image Anal. 9 (3) (2005) 209–221, doi:10.1016/j.media.2004.06.026.
[6] M. Ceresa, N. Mangado Lopez, H. Dejea Velardo, N.C. Herrezuelo, P. Mistrik, H.M. Kjer, S. Vera, R.R. Paulsen, M.Á. González Ballester, Patient-specific simulation of implant placement and function for cochlear implantation surgery planning, in: Proceedings of MICCAI, 2014, pp. 49–56, doi:10.1007/978-3-319-10470-6_7.
[7] N.D. Cornea, D. Silver, P. Min, Curve-skeleton properties, applications, and algorithms, IEEE Trans. Vis. Comput. Graphics 13 (3) (2007) 530–548, doi:10.1109/TVCG.2007.1002.
[8] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[9] E. Erixon, H. Högstorp, K. Wadin, H. Rask-Andersen, Variational anatomy of the human cochlea: implications for cochlear implantation, Otol. Neurotol. 30 (1) (2009) 14–22, doi:10.1097/MAO.0b013e31818a08e8.
[10] T. Heimann, H.-P. Meinzer, Statistical shape models for 3D medical image segmentation: a review, Med. Image Anal. 13 (4) (2009) 543–563, doi:10.1016/j.media.2009.05.004.
[11] D. Hill, P. Batchelor, M. Holden, D. Hawkes, Medical image registration, Phys. Med. Biol. 46 (3) (2001) R1–R45, doi:10.1088/0031-9155/46/3/201.
[12] T. Ju, M.L. Baker, W. Chiu, Computing a family of skeletons of volumetric models for shape description, Comput.-Aided Des. 39 (5) (2007) 352–360, doi:10.1016/j.cad.2007.02.006.
[13] S. Klein, J.P.W. Pluim, M. Staring, M.A. Viergever, Adaptive stochastic gradient descent optimisation for image registration, Int. J. Comput. Vis. 81 (3) (2009) 227–239, doi:10.1007/s11263-008-0168-y.
[14] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P. Pluim, Elastix: a toolbox for intensity-based medical image registration, IEEE Trans. Med. Imaging 29 (1) (2010) 196–205.
[15] A. Liu, E. Bullitt, S. Pizer, 3D/2D registration via skeletal near projective invariance in tubular objects, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention - MICCAI'98, 1496, 1998, pp. 952–963.
[16] W.E. Lorensen, H.E. Cline, Marching cubes: a high resolution 3D surface construction algorithm, SIGGRAPH Comput. Graph. 21 (4) (1987) 163–169, doi:10.1145/37402.37422.
[17] J.B.A. Maintz, M.A. Viergever, A survey of medical image registration, Med. Image Anal. 2 (1) (1998) 1–36, doi:10.1016/S1361-8415(98)80001-7.
[18] F.P.M. Oliveira, J.M.R.S. Tavares, Medical image registration: a review, Comput. Methods Biomech. Biomed. Eng. 17 (2) (2014) 73–93, doi:10.1080/10255842.2012.670855.
[19] R. Paulsen, J. Baerentzen, R. Larsen, Markov random field surface reconstruction, IEEE Trans. Visual. Comput. Graphics 16 (4) (2010) 636–646, doi:10.1109/TVCG.2009.208.
[20] T. Rohlfing, C. Maurer, D. Bluemke, M. Jacobs, Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint, IEEE Trans. Med. Imaging 22 (6) (2003) 730–741, doi:10.1109/TMI.2003.814791.
[21] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, D. Hawkes, Nonrigid registration using free-form deformations: application to breast mr images, IEEE Trans. Med. Imaging 18 (8) (1999) 712–721, doi:10.1109/42.796284.
[22] K. Saeed, M. Tabedzki, M. Rybnik, M. Adamski, K3M: a universal algorithm for image skeletonization and a review of thinning techniques, Int. J. Appl. Math. Comput. Sci. 20 (2) (2010) 317–335, doi:10.2478/v10006-010-0024-4.
[23] M. Staring, S. Klein, J.P.W. Pluim, A rigidity penalty term for nonrigid registration, Med. Phys. 34 (11) (2007) 4098–4108, doi:10.1118/1.2776236.
[24] H. Sundar, D. Silver, N. Gagvani, S. Dickinson, Skeleton based shape matching and retrieval, in: Proceedings of Shape Modeling International 2003 (SMI), 2003, pp. 130–139.
[25] L. Tang, G. Hamarneh, Smrfi: Shape matching via registration of vector-valued feature images, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[26] J.W. Tangelder, R.C. Veltkamp, A survey of content based 3D shape retrieval methods, in: Proceedings of Shape Modeling International SMI 2004, 2004, pp. 145–156.
[27] B. Tom, S. Efstratiadis, A. Katsaggelos, Motion estimation of skeletonized angiographic images using elastic registration, IEEE Trans. Med. Imaging, 13 (3) (1994) 450–460.
[28] S. Vera, D. Gil, A. Borras, M.G. Linguraru, M.Á. González Ballester, Geometric steerable medial maps, Mach. Vis. Appl. 24 (6) (2013) 1255–1266, doi:10.1007/s00138-013-0490-4.
[29] S. Vera, F. Perez, C. Balust, R. Trueba, J. Rubió, R. Calvo, X. Mazaira, A. Danasingh, L. Barazzetti, M. Reyes, M. Ceresa, J. Fagertun, H. Kjer, R. Paulsen, M. González Ballester, Patient Specific Simulation for Planning of Cochlear Implantation Surgery, Lecture Notes in Computer Science, Springer, pp. 101–108. doi: 10.1007/978-3-319-13909-8_13.
[30] B.S. Wilson, M.F. Dorman, Cochlear implants: a remarkable past and a brilliant future, Hear. Res. 242 (1-2) (2008) 3–21, doi:10.1016/j.heares.2008.06.005.
[31] P.A. Yushkevich, J. Piven, H.C. Hazlett, R.G. Smith, S. Ho, J.C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, Neuroimage 31 (3) (2006) 1116–1128.

Please cite this article as: H.M. Kjer et al., Free-form image registration of human cochlear μCT data using skeleton similarity as anatomical prior, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.07.017

Appendix B

Image Registration of Cochlear μCT Data Using Heat Distribution Similarity

Image Registration of Cochlear μCT Data Using Heat Distribution Similarity

Hans Martin Kjer[1( )], Sergio Vera[2,3], Jens Fagertun[1], Debora Gil[3], Miguel Ángel González-Ballester[4,5], and Rasmus Paulsen[1]

[1] Department of Applied Mathmatics and Computer Science, Technical University of Denmark, Copenhagen, Denmark
hmkj@dtu.dk
[2] Alma Medical Systems, Barcelona, Spain
[3] Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain
[4] Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
[5] ICREA, Barcelona, Spain

Abstract. Better understanding of the anatomical variability of the human cochlear is important for the design and function of Cochlear Implants. Good non-rigid alignment of high-resolution cochlear μCT data is a challenging task.

In this paper we study the use of heat distribution similarity between samples as an anatomical registration prior. We set-up and present our heat distribution model for the cochlea and utilize it in a typical cubic B-spline registration model. Evaluation and comparison is done against a corresponding normal registration of binary segmentations.

1 Introduction

Image registration and finding data correspondence is a well known challenge in biomedical image analysis. The choice and performance of the registration model depends highly on the involved imaging modalities, the anatomy of the object, the desired end-goal, etc. [9]. Larger and more complex deformations make it more challenging to set-up the optimal registration procedure. It becomes more difficult for the optimization to avoid local minima as the amount of parameters in the transformation model increases. To efficiently solve the problem it is often required to include additional data pre-processing, prior knowledge, regularization or constraints.

In this study we work with registration of μCT images of the inner ear, which is the structure controlling hearing and balance. Establishing correspondences between samples is required to understand the anatomical/shape variability, which has uses in a range of interesting clinical applications regarding Cochlear Implants [13]. It can be in a patient-specific context, by advising recipients and surgeons to choose an electrode design suited for the anatomy of the user [12], and/or optimizing the CI-programming based on patient-specific physiological simulations [4]. Description of population-based anatomical variability further

R.R. Paulsen and K.S. Pedersen (Eds.): SCIA 2015, LNCS 9127, pp. 234–245, 2015.
DOI: 10.1007/978-3-319-19665-7_20

allow manufacturers to explore the general implant design. The inner ear is terminologically divided into two parts - the spiral formed cochlea (for hearing) and the vestibular system (for balance) consisting of three semi-circular canals in a close to perpendicular configuration (see Figure 1). We will focus mostly on the cochlea part in this study, as this to our experience is the more difficult part to register. The challenge is mainly due to the relatively large observable differences compared to the compactness of the spiral. The spiral outer dimensions approximates to 10x8x4 mm, and on average the cochlea winds 2.6 turns [5] but can approach up to three full turns. This corresponds to a difference in the order of 1-2 mm following the path of the spiral. The separation between the cochlear turns in the same region is typically one order of magnitude smaller. Further, the spiral lacks distinct features to identify corresponding locations. The whole spiral is a tube-like structure (see Figure 1) with a large degree of self-similarity in cross-sections.

Our aim is to set up an image registration procedure for cochlear data. Conventional practice is to start with a global alignment (rigid, affine etc.) followed by a non-rigid step consisting of very local deformations. The common global transformations cannot align cochlear shapes very well, as they cannot take into account the spiral nature of the data. This lack of global fit should be handled by the non-rigid registration instead. The desired non-rigid model should not just expand or reduce the apical part of the spiral to cope with variability in cochlea turning, but instead it should try to readjust the entire spiral. Essentially the model should be able to handle very local deformations while still adhering to a more global structure.

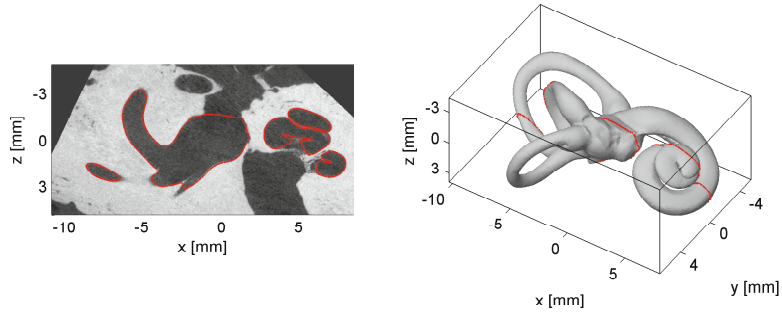


Fig. 1. Left: A slice from a μCT dataset showing the segmentation of the inner ear. Right: A surface model of the inner ear constructed from the segmentation of a μCT dataset.

There are different ways to modify the registration model to achieve this, and inclusion of prior-knowledge have been studied previously. An anatomical shape prior can be in form of a statistical shape model [2,6]. However, building shape models is in itself a labor intensive task rivaling if not surpassing the task of the registration, as the prerequisite for building the model is data that is already registered to have correspondences.

In the work of [1] an articulated skeleton model was pre-registered to intra-mouse data studies in order to recover large pose-differences between data acquisitions. Landmark correspondences of the skeleton model were then used to regularize the registration cost function. While in principle it is a useful approach to our case, corresponding landmarks are to our experience difficult to establish coherently and consistently for the cochlea.

It is also possible to do clever preprocessing of the data to obtain an image similarity measure more suited for the registration. In the work of [11] a selection of shape features were calculated from the objects and transformed into vector-valued 2D feature images, that were then registered using a classic image registration formulation. Skeletal features were used to provide the more global similarity between samples, while curvature and convexity features handles local similarity. The principle idea of processing the data to find a new similarity metric is sound and an approach that we follow. Instead of working from meshes and shape features we will work from the images and explore the use of a single image feature to improve our registrations - the heat distribution. Techniques for calculating volumetric maps of 3D heat distributions already exist, and have the potential to provide the required global context to the registration model. Since a heat map can be considered just another type of gray-scale image data, this approach would fit into already existing registration frameworks with little modification.

The purpose of this study is then to test and evaluate deformable registration using a B-spline transformation model on a series of inner ear/cochlear datasets with and without the use of heat distribution similarity in the registration model.

2 Materials and Methods

In this section we provide a more detailed description of the data and its processing, the set up and evaluation of registration models and the computation of heat distributions. In order to conceptualize and test the overall procedure we create and use some simple synthetic 2D images. This data is introduced in Section 2.1 and corresponding registration model described in Section 2.4. Details concerning the cochlear data and the initial processing are found in Section 2.2. The computation of the heat distribution in this data is described in Section 2.3. Finally, the registration models (for both data types) are detailed in Section 2.4 and their evaluation in Section 2.5.

2.1 2D Synthetic Data

Data Generation: Small synthetic 2D images are generated to demonstrate the concept of using maps of heat distributions in a registration model. Two binary 100x100 images, B_i, were created manually containing a foreground spiral region and a background. In order to create a classic gray-scale version of the images (I_i), random noise was added to each pixel. The noise models for both regions were Gaussian, $\mathcal{N}_j(\mu_j, \sigma_j)$. The corresponding volumetric heat distributions (H_i)

were generated simplistically; The apex of the spiral was manually selected as the source. Heat was propagated iteratively with a 4-neighborhood kernel. The heat spreads to previously untouched foreground voxels, and for each iteration the heat is decreased by 1. The synthetic data is presented in Figure 2.

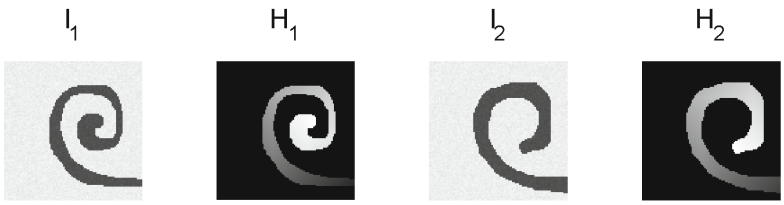


Fig. 2. Synthetic 2D data with I_1 and I_2 showing the two gray-scale images, and H_1 and H_2 the corresponding heat map versions

Image Registration: In short, we initially calculate one rigid transformation that is applied to both of the moving images, i.e. I_2 and H_2. This is followed by non-rigid registrations between respectively the pair of grayscale and heat map images. The formulation and details on the images registrations are described in Section 2.4. The result of the deformations are visually compared and evaluated.

2.2 Cochlear Data

A collection of N=17 dried temporal bones from the University of Bern were prepared and scanned with a Scanco Medical μCT100 system. The data was reconstructed and processed to obtain image volumes of 48 micron isotropic voxel-sizes containing the inner ear (Figure 1, left).

Image Segmentation: The inner ear border was delineated semi-automatically to obtain a binary segmentation, B_i, of each dataset using ITK-SNAP [14].The images contain some openings and non-anatomical artifacts that had to be manually handled to obtain comparable segmentations across datasets. A surface model was generated for each dataset using Marching Cubes followed by a surface reconstruction [10] to obtain a well-formed triangular mesh (Figure 1, right).

Initial Rigid Alignment: Before proceeding further we choose to perform an initial registration of the data to bring it into a common space and orientation. We use a rigid transform to remove translational and rotational differences between the datasets while retaining variation in size or scaling.

The initial registration is done with principal component transformations calculated from the extracted surface models. One dataset was chosen as the reference, to which the others were aligned. In short, the mean vertex position, $\bar{p_i}$, of the i-th dataset is subtracted from all vertices translating the center of mass to position (0,0,0). Finding the eigenvectors of the 3-x-3 covariance matrix of the mesh vertex positions (after the translation) provides a rotation matrix to

the principal axes. This essentially corresponds to fitting an ellipsoid to the point cloud of each dataset and aligning the ellipsoid axes, and it works consistently due to the asymmetry of the inner ear shape.

After the initial registration all images were cropped to contain only the cochlea region, since this is the structure of our focus. The first principal axis described above separates the cochlea from the vestibular region. Therefore, the cropping is accomplished easily by splitting each dataset at x = 0 (see Figure 1 and 3a).

2.3 Heat Distribution

The usage of heat distribution models for registration of complex anatomical shapes allows the generation of a global similarity between samples regardless of local differences in shape.

Calculating a heat map, H_i, for each dataset is done after the initial registration and is initiated by the placement of a source and sink. The source was manually set at the cochlear apex (see Figure 3a) and the sink was chosen as the inner ear center of mass (i.e. position (0,0,0)). The heat values at the two extremes were fixed. The heat distribution over time in a region is governed by the Partial Differential Equation (PDE)

$$\frac{\partial u}{\partial t} - \alpha \Delta u = 0 \tag{1}$$

which describes the distribution of heat, u, over time, t, in a medium with thermal conductivity α. With Δ being the Laplacian operator $\Delta u \equiv u_{xx} + u_{yy} + u_{zz}$. Assuming α to be one and $t \to \infty$ we obtain the thermal equilibrium state solution described by the Laplacian equation

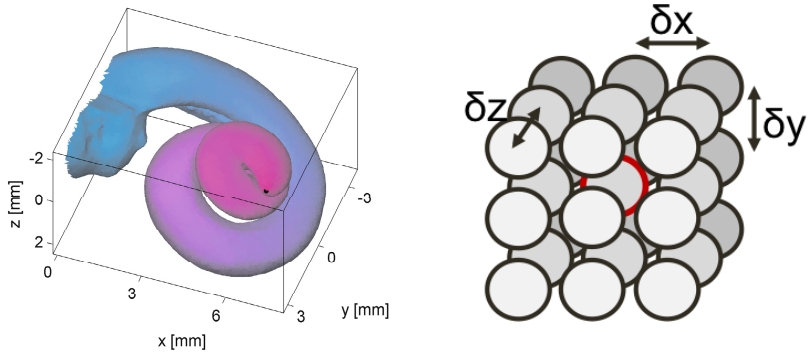
$$\Delta u = 0 \tag{2}$$

Solving this PDE (Eq. 2) implies solving a boundary value problem with boundary conditions [3]. Careful assignment of the boundary values can be positive for the registration process. Assigning Dirichlet boundary conditions f to specific anatomical sites $\mathcal{A}$ such as the apex and the center of the cochlea creates a shape metric image that can be used for steering the registration process to an improved solution.

Discrete Laplacian on a Closed Domain: We want to solve

$$\Delta u = 0 \qquad u_{|\mathcal{A}} = f \tag{3}$$

for $f = f(x, y, z)$ the boundary values defined at anatomical sites $\mathcal{A}$. All the foreground voxels of a binary image segmentation, B_i, will constitute the domain Ω where the heat diffusion will be applied. By applying the Laplace discrete operator to all image voxels, Equation 2 can be written in matrix form as $\mathbf{A}u = 0$. The matrix $\mathbf{A}$, called the graph Laplacian or adjacency matrix, encodes the



(a) Heat distribution at the surface of the cochlear. The source point in the apex is marked in black.

(b) Discrete voxel lattice. Central voxel highlighted in red. Neighbors in x, y and z axis are at distances $\delta x, \delta y, \delta z$ respectively.

Fig. 3. Heat distribution for 3D data

neighboring relations between voxels. This is because we in the voxel image lattice (see Figure 3b) can approximate the discrete Laplacian by computing the second order finite differences.

The solution to the Laplacian with Dirichlet anatomical conditions is obtained by solving the system of equations $\mathbf{A}u = \mathbf{b}$, with $\mathbf{b}$ being a row matrix encoding the boundary values at $\mathcal{A}$ given by:

$$f(x,y,z) = \begin{cases} 1, & \text{for } (x,y,z) \in \text{ cochlear apex (source)} \\ 0, & \text{for } (x,y,z) \in \text{ cochlea center of mass (sink)} \end{cases} \tag{4}$$

Although $\mathbf{A}$ and $\mathbf{b}$ are sparse by definition, solving the system of equations with standard techniques might be unfeasible. Given that $\mathbf{A}$ is symmetric and positive definite we can use the Preconditioned Conjugate Gradient method using the Incomplete Cholesky Factorization of $\mathbf{A}$ as a pre-conditioner [7]. This allows solving the system iteratively in a short time and with a low memory footprint. For the generation of the map to be useful, the cochlea segmentation must ensure that the cochlear turns does not intersect each other in the segmentation.

Two Step Heat Map Generation: When solving the entire 3D domain using just a sink and a source voxel the decay of the heat function is very fast to our experience. We obtain a very small heat gradient throughout the cochlea and the resulting heat map is too flat to be used properly. We improve the heat map by applying a two step process for the generation of the map. First, we compute the heat map with a single voxel sink and source point, but using only the boundary voxels of the cochlea segmentation as the diffusion domain. This ensures that heat spreads over a reduced number of voxels, generating a boundary heat map with better gradient and slower heat decay. Second, the resulting boundary heat

map is used as boundary condition for a second heat propagation over the entire cochlea segmentation domain.

2.4 Deformable Image Registration

The work-flow for non-rigid image registration is quite standard. The (N-1) moving datasets are registered to a chosen reference, following the formulation and framework of the `elastix` [8] toolbox.

The registration of the moving dataset, I_M, towards the fixed image, I_F, is formulated as a (parametric) transformation, T_μ, where the vector μ containing the p-parameters of the transformation model are found as an optimization of the cost function, $\mathcal{C}$.

$$\hat{\mu} = \arg\min_{\mu} \mathcal{C}(T_\mu, I_F, I_M) \tag{5}$$

The transformation model used in this paper is the cubic B-spline grid in a multi-resolution setting. The spacing between grid points follow a gradually decreasing schedule to start with a rough alignment that is slowly refined. The particular schedule used was (24, 16, 8) voxels.

Binary Set-up: We make a 'normal' registration between the segmentation volumes, to have something to compare our proposed usage of heat distributions against. The registration is done between the binary segmentation volumes, B_i, rather than the gray-scale volumes, I_i, for two reasons. First, the μCT data contain smaller artifacts and certain weakly contrasted edges, that were dealt with during the segmentation (Section 2.2). Secondly, the registration should not be influenced by the anatomical differences in the surrounding bone structure. The following cost-function was used for the registration:

$$\mathcal{C}_1 = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, B_F, B_M) + (1 - \alpha) \cdot \mathcal{P}_{\text{BE}}(\mu) \tag{6}$$

where α is a weight parameter in the interval [0,1], here chosen to 0.9 by means of experimentation. For the similarity term, $\mathcal{S}_{\text{Sim}}$, the sum of squared differences (SSD) is chosen. The term $\mathcal{P}_{\text{BE}}$ is the energy bending regularization used to penalize strong changes and foldings in the transformation.

Heat-map Set-up: In order to use the similarity of heat distributions in the registration, the cost function can essentially remain the same, simply replacing the underlying image data from the segmentations, B_i, to the heat maps, H_i. By experimentation we set $\alpha = 0.1$, while all other settings remains un-altered in order to provide a fair comparison between the two set-ups.

$$\mathcal{C}_2 = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, H_F, H_M) + (1 - \alpha) \cdot \mathcal{P}_{\text{BE}}(\mu) \tag{7}$$

Table 1. Statistics of registration evaluation metrics, reported as the mean +/- 1 std. Model 1 and 2 refers respectively to Eq. 6 (binary) and Eq. 7 (heat).

Metric	**Dice Score**	**Mean Err.** [mm]	**Max Err.** [mm]	**Apex Err.** [mm]
Model 1	0.96 ± <0.01	0.07 ± <0.01	0.54 ± 0.13	0.97 ± 0.58
Model 2	0.99 ± <0.01	0.05 ± <0.01	0.43 ± 0.12	0.66 ± 0.58

Registration Model - 2D Synthetic Data: The two registrations models described above (Eq. 6 and Eq. 7) are also applied in the case of synthetic 2D data with some minor modifications. The cost function terms remains unaltered, i.e. the SSD was used for the similarity metric with added bending energy regularization. However, the weighting parameter was set to $\alpha = 0.7$ in both cases, and only a single resolution was used with no smoothing and a B-spline grid point spacing of 3 pixels.

2.5 Evaluation

We are interested in comparing the registration results, $\hat{\mu}_j$, from using either model 1 (Eq. 6) or model 2 (Eq. 7) with different metrics. For a voxel based score we calculate the Dice Coefficient between the deformed binary volume, $B_i(T_{\mu_j})$, and the reference, B_{Ref}. The ground truth mesh, M_i, can be compared to the deformed reference mesh, $M_{\text{Ref}}(T_{\mu_j})$, from which we calculate the two-sided mean surface error and Hausdorff distance.

Since the above mentioned evaluation metrics are very generic and global we also use the apex error. A landmark is manually placed in the apex in all datasets, and we calculate the euclidean distance between the anatomical correspondence.

3 Results

2D Synthetic Data: The results with the synthetic data are shown in Figure 4. The 'normal' intensity registration (model 1), μ_I, was unable to fully capture

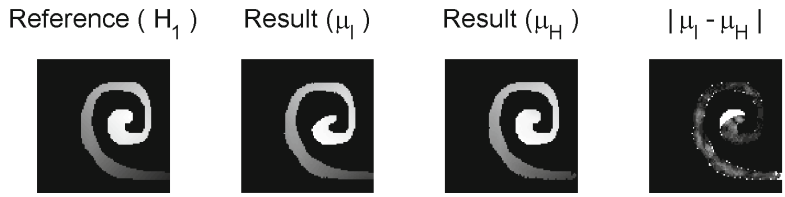
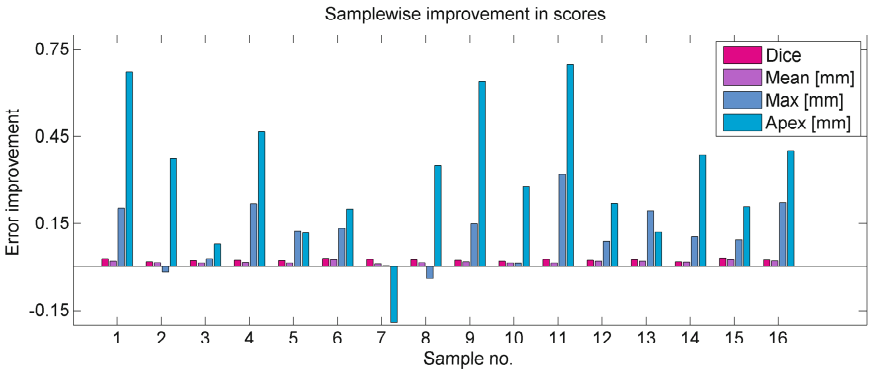


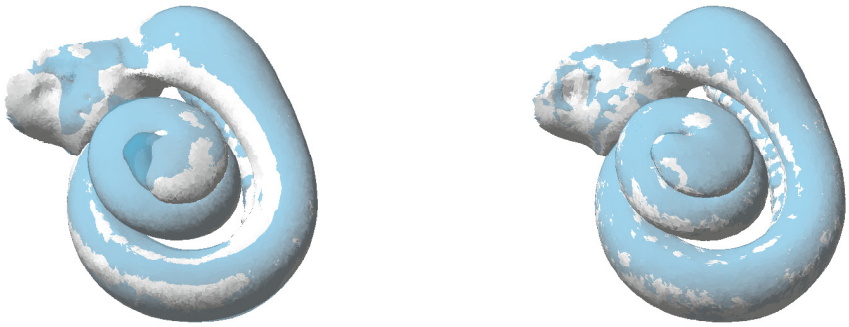
Fig. 4. Registration results on the synthetic 2D data. From left to right: The reference heat distribution H_1, The moving heat map, H_2, after transformation using the registration results, i.e. $H_2(T_{\mu_I})$ and $H_2(T_{\mu_H})$. Finally, the difference between the results, i.e. $|H_2(T_{\mu_I}) - H_2(T_{\mu_H})|$.

the full spiral with the given settings. When comparing the subtraction image of the two deformed heat maps, it should be noted that there is a difference throughout the entire spiral - suggesting that the registration based on the heat maps, μ_H, provides a more global twist of the spiral.

3D Cochlea Data: The quantitative evaluation is presented in Figure 5a and Table 1. The visual differences between the two registration models are illustrated in Figure 5b for a single case. In three cases the deformations resulted in an anatomically incorrect warping, where for instance the separation of the cochlea turns was not preserved. It is not apparent from the evaluation metrics, but very clear from a visual inspection of those samples.



(a) Graph of the sample wise improvement between model 2 (heat) and model 1 (binary) in evaluation scores.



(b) Comparison of deformations. Left: Model 1 (binary) Right: Model 2 (heat). Blue transparent surface is ground truth and gray is the fitted surface.

Fig. 5. Results of the registration of 3D data

4 Discussion and Conclusion

The contribution of this study is a presentation and evaluation of using heat distribution similarity as a way of aiding image registration of cochlear samples. For now we have demonstrated its ability to add flexibility to the transformation and to improve registration accuracy. However, there are other important aspects to consider. Introducing additional pre-processing obviously adds more computations and processing time. It is likely to involve extra parameters to tune and it may contribute with noise and uncertainty to the registration. Our approach rely heavily upon a good segmentation of the data, which required a lot of manual labor. The additional processing time for calculating the heat-distributions were not a substantial issue considering the observed benefits. There are some factors that may affect the usage of the heat distributions. First, the manual placement of the source landmark inherently carries some uncertainty, but it is not immediately considered problematic. Secondly and more importantly, is the effect of the heat gradient throughout the cochlea. If the change in heat is too small, it might not have the desired guiding effect upon the registration, and too much change might force very strong deformations. Finding the right balance should be part of our future work. It is further important to note that the transformation model remains the same with this pre-processing strategy. The registration between the original images (segmentation or gray-scale) have the potential to find the same transformations as with the heat maps included. Without the heat distributions, the convergence rate may be slower and more levels of resolution and tweaking of registration parameters is required. The data pre-processing can have the added benefit of easing the registration set-up.

Our evaluation of the method is based mostly on global metrics. Their usefulness is limited, as they are hardly able to show whether a registration is successful or not. For that we need more anatomically meaningful evaluation metrics like the apex error. This error is very local, so it does not reveal if the desired deformation throughout the cochlea is achieved. The lack of consistent landmarks or uniquely identifiable locations on the cochlear makes it challenging to perform better quantitative evaluation of the registration. It would require setting up larger elaborate methods, exceeding the scope of this study. What we conclude from the results, is that heat distribution similarity adds flexibility to the registration, thereby allowing turning of the cochlea apex to be captured in a more anatomically correct manner.

Considerations can also be made to the chosen registration model. We work with the B-spline grid as the transformation model which has some limitations. Choosing a fluid- or optical flow-based model [9] could potentially be more suited for this kind of spiral anatomy.

Working on the binary data, better registrations were achieved with little regularization ($\alpha = 0.9$) in order to allow larger deformations. But since there is nothing to guide the transform to the anatomical correct place, then the method falls short. The maps of heat distributions provide global similarity to the otherwise locally defined B-spline transformation. It adds a lot of flexibility to the procedure - in fact too much. Strong regularization ($\alpha = 0.1$) was required

to ensure reasonable deformations. Even though the results look very positive, our experience is that the procedure is not stable. It is likely to run into very large unreasonable deformations, that for instance goes across the separation of the cochlea turns. Essentially, we have a difficulty in applying enough and/or correct regularization. This is a problem to be addressed in future work. Additional regularization could for instance be forcing local rigidity in the bony structures surrounding the cochlea.

Of the 16 tested registrations, we observed three failed cases where deformations were unreasonable. These datasets are the cases least similar to chosen reference data. The issue is therefore not only lack of regularization, but part of the solution could also be by introducing coarser resolution levels in the registration. Further, this study was run on down-sampled data. The data is originally in a higher resolution, where the separation between cochlea turns is more clear. Using this data may help prevent the unreasonable deformations.

To conclude, we have presented a data pre-processing strategy for aiding non-rigid image registration using similarity of heat distributions. We have tested the idea on synthetic 2D data and on μCT of the cochlea, and demonstrated its ability to provide a global guidance and flexibility to the registration procedure.

Acknowledgments. The research leading to HEAR-EU results has received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement n°304857.

References

1. Baiker, M., Staring, M., Löwik, C.W.G.M., Reiber, J.H.C., Lelieveldt, B.P.F.: Automated registration of whole-body follow-up MicroCT data of mice. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 516–523. Springer, Heidelberg (2011)
2. Berendsen, F.F., van der Heide, U.A., Langerak, T.R., Kotte, A.N.T.J., Pluim, J.P.W.: Free-form image registration regularized by a statistical shape model: application to organ segmentation in cervical MR. Computer Vision and Image Understanding **117**(9), 1119–1127 (2013)
3. Brechbuhler, C., Gerig, G., Kubler, O.: Parametrization of closed surfaces for 3-d shape-description. Computer Vision and Image Understanding **61**(2), 154–170 (1995)
4. Ceresa, M., Mangado Lopez, N., Dejea Velardo, H., Carranza Herrezuelo, N., Mistrik, P., Kjer, H.M., Vera, S., Paulsen, R.R., González Ballester, M.A.: Patient-specific simulation of implant placement and function for cochlear implantation surgery planning. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 49–56. Springer, Heidelberg (2014)
5. Erixon, E., Högstorp, H., Wadin, K., Rask-Andersen, H.: Variational Anatomy of the Human Cochlea: Implications for Cochlear Implantation. Otology and Neurotology **30**(1), 14–22 (2009)
6. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis **13**(4), 543–563 (2009)

7. Hestenes, M.R., Stiefel, E.: Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards **49**(6), 409–436 (1952)
8. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1), 196–205 (2010)
9. Oliveira, F.P.M., Tavares, J.M.R.S.: Medical image registration: a review. Computer Methods in Biomechanics and Biomedical Engineering **17**(2), 73–93 (2014)
10. Paulsen, R., Baerentzen, J., Larsen, R.: Markov random field surface reconstruction. IEEE Transactions on Visualization and Computer Graphics **16**(4), 636–646 (2010)
11. Tang, L., Hamarneh, G.: SMRFI: shape matching via registration of vector-valued feature images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
12. Vera, S., Perez, F., Balust, C., Trueba, R., Rubió, J., Calvo, R., Mazaira, X., Danasingh, A., Barazzetti, L., Reyes, M., Ceresa, M., Fagertum, J., Kjer, H.M., Paulsen, R., Ballester, M.Á.G.: Patient specific simulation for planning of cochlear implantation surgery. In: Linguraru, M.G., Laura, C.O., Shekhar, R., Wesarg, S., Ballester, M.Á.G., Drechsler, K., Sato, Y., Erdt, M. (eds.) CLIP 2014. LNCS, vol. 8680, pp. 101–108. Springer, Heidelberg (2014)
13. Wilson, B.S., Dorman, M.F.: Cochlear implants: A remarkable past and a brilliant future. Hearing Research **242**(1–2), 3–21 (2008)
14. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. Neuroimage **31**(3), 1116–1128 (2006)

Appendix C

Shape Modelling of the Inner Ear From Micro-CT Data

015 **SHAPE MODELLING OF THE INNER EAR FROM MICRO-CT DATA**

Kjer, H.M.[1], Fagertun, J.[1], Vera, S.[2], González-Ballester, M.A.[3], Paulsen, R.R.[1],

[1]Technical University of Denmark, Copenhagen, DENMARK. [2]Alma IT Systems, Barcelona, SPAIN. [3]Universitat Pompeu Fabra, Barcelona, SPAIN.

Background: It is crucial to understand the anatomy and anatomical variability of the inner ear in order to improve the design and functionality of implantable hearing devices such as Cochlear Implants (CI). Statistical shape modelling of the inner ear provides a versatile tool that can aid numerous interesting Cochlear Implant applications, in particular implant design optimization and surgical planning.
The spiral-shaped cochlear of the inner ear presents an anatomy with small important features that can only be properly perceived in high resolution micro-CT scans of cadaveric specimens. The complex anatomy and the large data sizes make it a challenging dataset to handle and to build a shape model from.

Aims: To build a statistical shape model of the inner ear from high resolution micro-CT data.

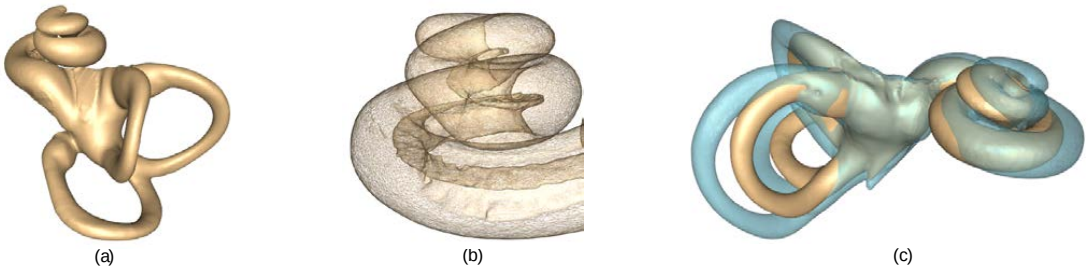
Methods: 17 temporal bones excised from human cadavers were dried and scanned with a micro-CT system (Scanco Medical, Switzerland). The region of interest in the datasets was reconstructed in 24 micron isotropic voxels (resulting in approximately 6GB of data per scan).
Segmentation: The inner ear (cochlear and vestibular system) was segmented manually using ITK-SNAP[1]. The surfaces of the segmentations were extracted using Marching Cubes and post-processed using MRF surface reconstruction[2] to provide smooth and well-formed surface meshes.
Registration: One dataset was chosen as a reference. An initial rigid transformation aligning the center of mass and the principal directions was calculated. The principal directions are consistent due to the asymmetric shape of the inner ear. This was followed by a deformable registration using elastix[3]. A multi-level B-spline grid minimizing sum of squared differences with bending energy regularization was used. The registration was between the segmentation images to reduce the influence of the noise in the micro-CT data.
Model Building: The transformations were applied to the reference surface model, to create surfaces representing the anatomy in the individual datasets with point correspondences. Using Statismo[4] a point distribution model (PDM) was built (Figure 1-3).
Evaluation: The quality of the model is constrained by the accuracy of the registration which is evaluated against the ground truth segmentation using Dice score and Hausdorff distance (the latter calculated with the 'ground truth' surface model).

Results: A PDM of the inner ear containing 466k vertices and 16 modes of variation (Figure 1-3). The average Dice score was 0.96 ± 0.01 and average Hausdorff distance 0.69 ± 0.24 mm.

Conclusions: An inner ear shape model has been built using open source libraries and tools. The model has more anatomical detail and modes of variation than what has previously been reported.

Acknowledgements: The research leading to HEAR-EU results has received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement №304857.

References: [1] P.A. Yushkevich, J. Piven, H.C. Hazlett, R.G. Smith, S. Ho, J.C. Gee, G. Gerig, "User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability", Neuroimage, 2006.
[2] R.R. Paulsen, J.A. Baerentzen, R. Larsen, "Markov Random Field Surface Reconstruction", IEEE Transactions on Visualization and Computer Graphics, 2010.
[3] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P.W. Pluim, "elastix: a toolbox for intensity-based medical image registration", IEEE Transactions on Medical Imaging, 2010.
[4] M. Lüthi, R. Blanc, T. Albrecht, T. Gass, O. Goksel, P. Büchler, M. Kistler, H. Bousleiman, M. Reyes, P. Cattin, T. Vetter, "Statismo - A framework for PCA based statistical models", 2012.



Figures: (a) Mean shape. (b) Close up on the cochlear. (c) 1st mode showing the change in overall size.

Appendix D

Predicting Detailed Inner Ear Anatomy From Pre-Operational CT For Cochlear Implant Surgery

#283 - ISCAS / Lecture or Poster

Predicting Detailed Inner Ear Anatomy From Pre-Operational CT For Cochlear Implant Surgery

Clinical Applications / Surgery and Subspecialties / ENT

Hans Martin Kjer[1], Sergio Vera[2], Jens Fagertun[1], Frederic Peréz[2], Javier Herrero[2], Miguel Angel González-Ballester[3], Rasmus Reinhold Paulsen[1]

1. Technical University of Denmark, Copenhagen, Denmark
2. Alma Medical Systems, Barcelona, Barcelona, Spain
3. Universitat Pompeu Fabra, Barcelona, Barcelona, Spain

keywords: Cochlear Implants, Image Registration, Shape Model Regularization, CT

Purpose: A Cochlear Implant is a surgically inserted prosthetic device for restoration of hearing given to persons who are profoundly deaf or severely hard of hearing.
Pre- and post-operational CT scans are routinely used in planning and assessment of Cochlear Implant surgeries. However, due to the small size of the implant and cochlea, the images contain only very gross anatomical information about the inner ear.
Providing additional patient-specific anatomical information about the inner ear is very valuable. It allows surgeons and manufacturers to make decisions about the design and programming of the inserted implant, in a manner that optimizes the restored hearing capabilities of the recipient. A promising way of achieving this is to use statistical shape models from high-resolution imaging techniques such as μCT. Previous work already [1] shows the potential and the interesting clinical implications/applications.
In this study we present an alternative image registration approach for predicting detailed inner ear anatomy in pre-operative CTs using Statistical Deformation Model (SDM) regularization. Further, we present some preliminary evaluation of the clinically predictive accuracy.

Methods: The statistical model: A Statistical Deformation Model (SDM) [2] was built from 17 μCT datasets of inner ear cadaverous specimens. One segmented dataset was used as a reference to which the remaining datasets were non-rigidly registered using a B-spline registration model [3]. The high resolution of this type of data allows us to segment finer anatomical details not visible in normal CT.
Segmentation of clinical CT: In order to project the high resolution μCT inner ear model into the clinical CT data we use a series of image registrations (see Figure 1) following the formulation of the elastix toolbox [4].
Two rigid transformations (T1 and T2 Figure 1) are used for transforming the CT data to the corresponding position and orientation of the μCT reference.
The first transform (T1) is calculated from four anatomical landmarks [5] manually placed in both datasets. This landmark alignment roughly brings the CT data to same coordinate system as the μCT reference, and in the process the CT data is also cropped to a smaller anatomical region, as all of the surrounding CT information is not needed in the subsequent steps.
The second transformation (T2) is a rigid image registration calculated between the μCT and the cropped CT data. The placement of landmarks used in TI will always have some user-based uncertainty. To reduce that potential noise, we run a simple rigid image registration procedure that provides a small correction in position and orientation.
A non-rigid image registration (Tsdm) between the μCT and CT data then follows. The procedure is regularized with the SDM using the statismo-elastix software.
Technically it would be perfectly possible to do a normal free-form image registration, but the results would be unpredictable and unreliable, due to the large differences in image resolutions and the relatively noisy CT data.
Our solution is to regularize the procedure based on our knowledge of anatomical variability. The SDM represents a probability distribution of valid free-form registrations between μCT datasets. The SDM regularized registration finds the optimal transformation between the μCT reference and the target CT image within this subset of deformations, ensuring that the output is well-behaved. A detailed μCT based surface model is then projected into the CT dataset.

Results: For this preliminary study we have tested the SDM registration on two clinical CT datasets.
The qualitative accuracy of a registration can be visually inspected (see Figure 1 and 2). Further, we have manually segmented the CT data for a more quantitative comparison with the fitted model.
We calculate the mean and largest surface error (respectively average and biggest Euclidean distance from all CT vertices to

their nearest µCT vertex). We refrain from calculating the symmetric distances (i.e. also including distances from µCT to nearest CT vertex) as there are many µCT vertices with no meaningful correspondences in the rougher CT-mesh.
The largest surface errors for the two cases were 1.15 mm and 0.89 mm, with corresponding mean surface errors (± 1 std.) of 0.22 ± 0.16 mm and 0.23 ± 0.18 mm.
The larger errors were observed to be located in the fitting of the semi-circular canals.

Conclusions: We have demonstrated a method for providing additional information about the inner ear from pre-operational CT scans used in Cochlear Implant surgeries by the use of a statistical model. We have further presented some initial steps towards a more clinical validation of the procedure.
A statistical model based on 17 samples has some limitations in its predictive capabilities, as this number of samples is unlikely enough to capture the total variance of the population. The model represents the entire inner ear. It was observed that fitting errors were larger in the regions of the semi-circular canals. A possible improvement to model could therefore be to model the vestibular system and the cochlea separately.
Some caution is advised when interpreting the numbers of the clinical validation results on CT data. The segmentation of CT is coarse and prone to uncertainties. Even so, the largest surface error might be a useful metric for conveying an upper bound error. The mean surface error approaches the typical image resolution of such data, so it is difficult to say something conclusive about the accuracy without further investigation. Other more cochlea specific error metrics should be considered in the evaluation.

The research leading to HEAR-EU results has received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement nº 304857.

References: [1] J. H. Noble, R. F. Labadie, R. Gifford, and B. Dawant, “Image-guidance Enables New Methods for Customizing Cochlear Implant Stimulation Strategies," Neural Systems and Rehabilitation Engineering, IEEE Transactions on, vol. 21, no. 5, pp. 820-829, 2013.
[2] D. Rueckert, A. F. Frangi, and J. A. Schnabel, “Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration,"IEEE Transactions on Medical Imaging, vol. 22, no. 8, pp. 1014-1025, 2003.
[3] H. M. Kjer, J. Fagertun, S. Vera, M.A. González Ballester, and R. R. Paulsen, “Shape modelling of the inner ear from micro-CT data," Proceedings of Symposium on Statistical Shape Models and Applications, Shape 2014,p. 21, 2014.
[4] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, “elastix: a toolbox for intensity-based medical image registration," IEEE Transactions on Medical Imaging, vol. 29, pp. 196-205, January 2010.
[5] W. Wimmer, F. Venail, T. Williamson, M. Akkari, N. Gerber, S. Weber, M. Caversaccio, A. Uziel, and B. Bell, “Semiautomatic cochleostomy target and insertion trajectory planning for minimally invasive cochlear implantation," Biomed Research International, 2014.

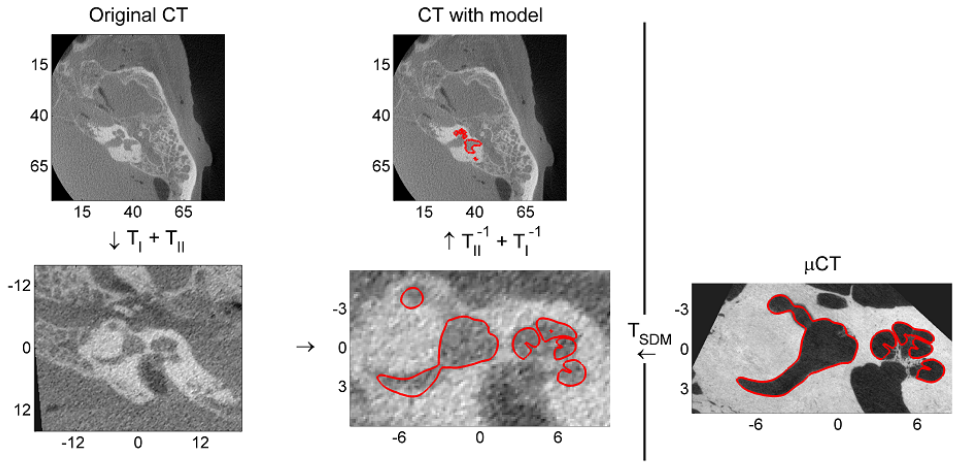


Figure 1: Illustration of the transforms in the proposed procedure. T_I and T_{II} are rigid transforms bringing data from the CT space (upper row) to μCT space (lower row) and back. The registration between μCT and CT is regularized with the SDM. Axes units are in mm.

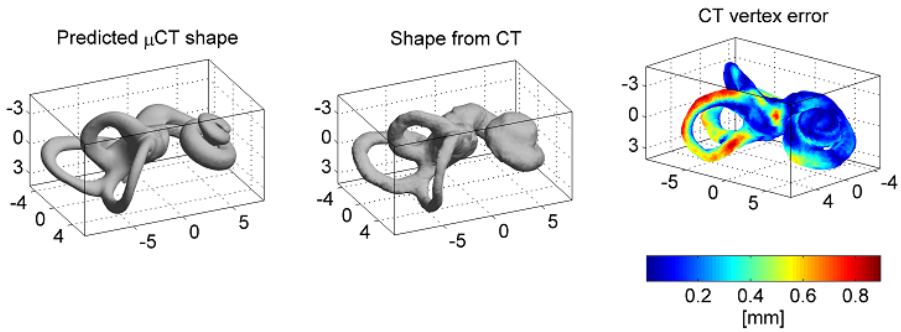


Figure 2: Comparison of the fitted model resulting from the SDM regularized registration against the shape segmented from the CT data. Axes units are in mm.

Appendix E

Patient-Specific Simulation of Implant Placement and Function for Cochlear Implantation Surgery Planning

Patient-Specific Simulation of Implant Placement and Function for Cochlear Implantation Surgery Planning

Mario Ceresa[1], Nerea Mangado Lopez[1], Hector Dejea Velardo[1], Noemi Carranza Herrezuelo[3], Pavel Mistrik[4], Hans Martin Kjer[5], Sergio Vera[6], Rasmus R. Paulsen[5], and Miguel Angel González Ballester[1,2]

[1] Simbiosys group, Universitat Pompeu Fabra, Barcelona, Spain
[2] ICREA, Barcelona, Spain
[3] Multicellular system biology group, Centre for Genomic Regulation, Barcelona, Spain
[4] Med-EL, Innsbruck, Austria
[5] Denmark Technical University, Copenhagen, Denmark
[6] Alma IT Systems, Barcelona, Spain

Abstract. We present a framework for patient specific electrical stimulation of the cochlea, that allows to perform in-silico analysis of implant placement and function before surgery. A Statistical Shape Model (SSM) is created from high-resolution human μCT data to capture important anatomical details. A Finite Element Model (FEM) is built and adapted to the patient using the results of the SSM. Electrical simulations based on Maxwell's equations for the electromagnetic field are performed on this personalized model. The model includes implanted electrodes and nerve fibers. We present the results for the bipolar stimulation protocol and predict the voltage spread and the locations of nerve excitation.

1 Introduction

Hearing impairment or loss is among the most common reasons for disability. Worldwide, 27% of men and 24% of women above the age of 45 suffer from hearing loss of 26dB or more. The cochlear implant (CI) is a surgically placed device that converts sounds to electrical signals, bypassing the hair cells and directly stimulating the auditory nerve fibers.

Even if cochlear implantation is able to restore hearing in patients with severe or complete functional loss, the level of restoration varies highly between subjects and depends on a variety of patient-specific factors [1]. Moreover, extreme care has to be taken when inserting the CI's electrode array into the cochlea to obtain the best possible positioning while not damaging residual hearing capabilities [2].

The HEAR-EU[1] project aims at reducing the inter-patient variability in the outcomes of surgical electrode implantation by improving CI designs and surgical protocols. In this context, we propose that the availability of an accurate

[1] `http://www.hear-eu.eu/`

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 49–56, 2014.

and personalized stimulation model of the cochlea can improve implant design, insertion planning and selection of the best treatment strategy for each patient.

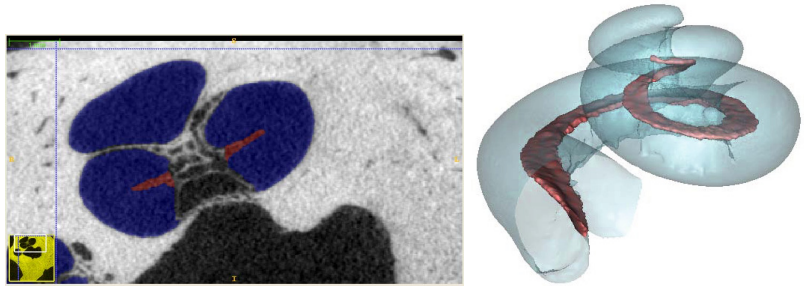
In this work a model is built from high-resolution μCT data to create a detailed simulation of the electrical properties of the cochlea. To the best of our knowledge, this is the first study combining human high-resolution imaging techniques, finite element (FE) methods and a nerve fiber model to predict implantation results in humans. Previous works used synthetic geometrical models [3,4,5] or were restricted to animals [6]. In this work, we advance the state of the art by using a realistic and anatomically detailed model of the cochlea based on a statistical shape model (SSM) created using human μCT images.

Using our model, we can predict potential spread in the cochlea and location of nerve activation after surgery. As higher spreads correlate with inter-electrode interference and distorted pitch perception, we can use the results of our model to detect which configurations of electrode placement are to be avoided during surgery.

The rest of the paper is structured as follows: Section 2 describes the methodology for image acquisition and processing, the creation of the SSM, of the FEM mesh and its adaptation for specific patients. Section 3 describes the electrical conduction model and the different experiments performed. Section 4 presents the neural nerve fiber model and its initialization with the results of the FEM simulations. Results are reported in Section 5 and discussion and directions for future work are provided in Section 6.

2 FEM Construction and SSM-Based Patient-Specific Adaptation

Our model is based on 17 temporal bones excised from human cadavers. The samples were dried and scanned with a high-resolution Scanco μCT 100 system (Scanco Medical AG, Switzerland). Each dataset has a nominal isotropic resolution of 24.5 μm.



(a) (b)

Fig. 1. a) Segmented transversal slice of the μCT, showing the cochlea (blue) and the spiral bone (red). b) Surface reconstruction of the cochlea, built from the μCT image.

After acquisition, data was segmented using a semi-automatic level-set method [7] and manual corrections. Fig. 1a shows a transversal slice of the μCT image, where the cochlea is segmented in blue and the bony spiral lamina in red. Once segmented, the surface of the cochleas were extracted from the segmentation using Marching Cubes [8]. A Markov Random Field surface reconstruction was applied [9] to obtain a surface more suited for generating a FEM mesh. Fig. 1b shows a 3D surface reconstruction based on the μCT image.

In order to create the SSM one dataset was chosen as a reference and an initial rigid transformation aligning the center of mass and the principal directions was calculated. Then the transformations were applied to the reference surface model to create surfaces representing the anatomy in the individual datasets and a point distribution model (PDM) was built. Further information are available in [10]. The SSM can be instantiated to generate deformation fields corresponding to valid deformations of cochlear shapes, and in particular, it can be used to find the deformation that best fits the patient's image data.

The FEM mesh is built from the surface model with the exception of the basilar membrane and the nerves that are created manually (Figure 2). A sensitivity analysis to prove the accuracy versus computational cost of the model was carried out, leading to a finite element mesh of 8.764.7273 tetrahedral elements.

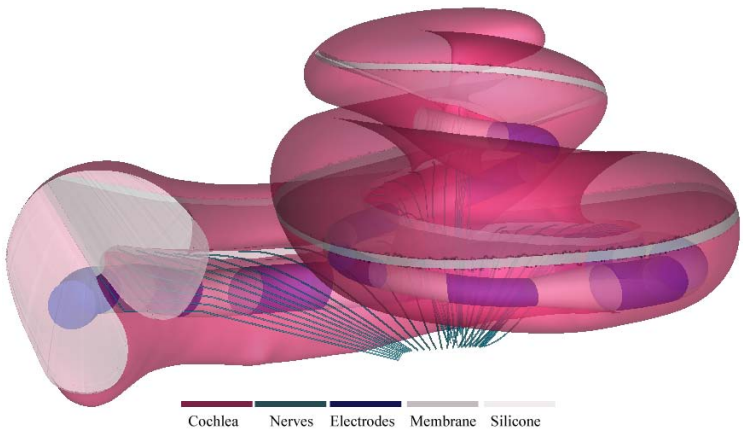


Fig. 2. Geometry of the FEM of the cochlea. Different colors represent the cochlea (purple), the nerves (aquamarine), the electrodes (violet), the basilar membrane (gray) and the silicon insulator (white). The maximum size of element selected for each part of the model and their electrical conductivity associated are as follows: 0.4, 0.005, 0.01, 0.008 , 0.8 (mm) and 1.43, 0.3, 1e7, 0.09375 and 1e-3 (S/m) for the cochlea, nerves, electrodes, membrane and silicone, respectively.

In our full pipeline the FE mesh is created on the mean shape of the SSM. This mesh is registered to the patient data, usually a lower resolution CT scan, in order to transfer the results to the clinical setting. The registration is regularized by the anatomical variability learnt in the shape model. This will produce a

geometrical fit but also capture the change in cochlear turning and basilar membrane length in an anatomical appropriate manner, thus making the simulation really patient specific.

The aspect ratio of FEM elements was checked in order to ensure a good mesh quality and the effect of the deformation on the quality of the elements is studied in [11]. All bio-mechanical properties and boundary conditions are also propagated, so new simulations can be run directly on the patient-specific mesh.

3 Electrical Conduction Model and Implant Stimulation Protocol

In order to simulate the electrical potentials we used the electrostatic solver of the open source tool Elmer [12]. We choose to use the quasi-static approximation and solve in this regime the Poisson equation:

$$\nabla \cdot \sigma \nabla \phi = \frac{\partial \rho}{\partial t} \tag{1}$$

where σ is the electric conductivity, ϕ the electric scalar potential and ρ the total charge density. For electric potential either Dirichlet or Neumann boundary condition can be used. The former prescribes the value of the potential on specified boundaries, the latter the current J_b on specified boundaries. Additional information on the electric model is available in [13]. A typical electrode array with 12 contacts (based on Med-EL Flex$^{\text{soft}}$design) was modeled and inserted in the FE model (Figure 2). The electrode array is placed through the round window into scala tympani along the lateral wall under the cochlea partition. In the full pipeline the post-operative electrode-array position would be estimated from a post-op CT scan. This is interesting as a surgical planning tool, as the surgeon can evaluate optimal electrode array design and position.

The boundary conditions for the electrode activation in the FE model are given by a stimulation pattern generation (SPG) and modeled after the manufacturer's indications. In this work we present the results relative to the bipolar (BP) stimulation protocol, where one electrode emits the current and the other is set to ground. In the following text, we refer to the bipolar stimulation as BPij where i is the source electrode and j the one set to ground. Typical stimulation currents for electrodes are in range of 0.3-1 mA and we set 1mA for all experiments. The conductivity parameters for the electrical simulation are taken from reference [6] from closely related animals.

4 Nerve Fiber Model

In order to describe the electrical properties of the nerve fibers, we use the Generalized Schwarz-Eikhof-Frijns (GSEF) model [14]. Each fiber is composed by a peripheral axon, a soma and a central axon. It has 16 compartments (Figure 3) where the voltages sampled from the FE model are applied to initialize the model, for a total of 64 coupled non-linear first-order differential equations per

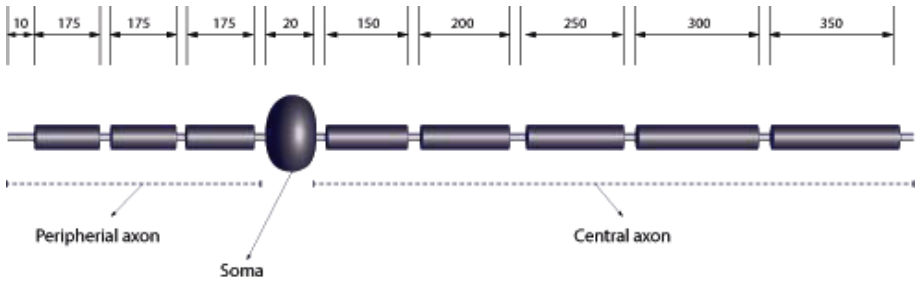


Fig. 3. Illustration of the first nodes of the nerve model used in our simulation, adapted from [14]. All dimensions are in μm, internodal gaps length is 1 μm, diameter of each compartment is 3 μm, except for the soma which is 10 μm. Internodal gaps represent the Ranvier nodes where we apply the voltage obtained by the FE model.

fiber. A total of 49 nervous fibers are modeled (Figure 2). In order to obtain the potential values to initialize our nerve model, we parameterized each of the 49 nerve bodies of the FEM using the arrival time from a heat diffusion equation. The parameterization was constructed in such a way that we could sample the potential field in 16 points per bodies, giving a total of 784 potential readings. Those points represent the un-myelinated parts of our fiber model, where the current flows in (Ranvier nodes). The readings in those points are shown in our potential spread plots (Figure 5). We implemented and solved the model in python using numpy, scipy and matplotlib open source tools [15].

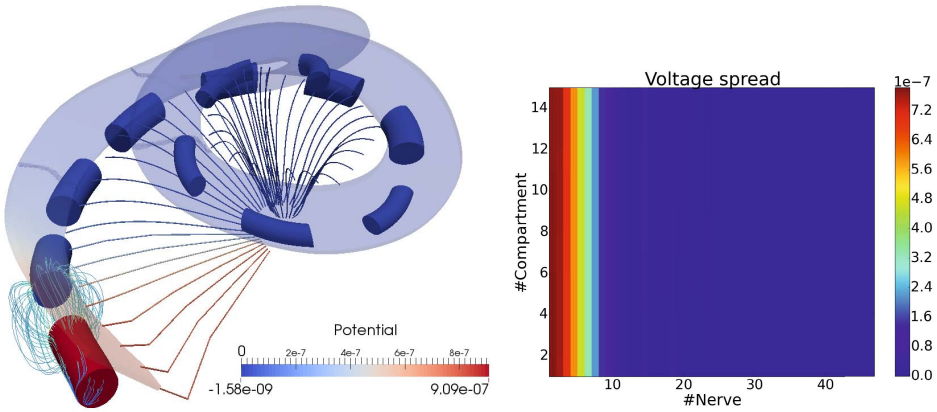
5 Results

A total of 11 FE simulations were run in steady state formulation. Each simulation run until convergence on the cluster at our Institution, that consists of 11 compute nodes, with four 16-core processors per node, for a total of 704 cores and a peak of 7876 Gflops.

In Figure 4a we show an example simulation for the BP12 protocol, with the complete 3D model and electrode 1 setup as source and electrode 2 as ground. Solid lines represent the current flowing between the electrodes.

In Figure 4b we sampled the potentials from the BP12 stimulation in 784 points of the nerve mesh in order to feed them to the spiking model described in Section 4. The colors represent the intensity of the electric field. We see that the stimulation is stronger in the closest nerve, yet several other nerves are also affected.

In Figure 5 we present the readings of the electric field for each nerves under different stimulation protocols. On the horizontal axis we have the stimulation protocols and on the vertical axis the indexes of the nerves. Differences in electric reading depend on the size of the electrode and its distance from the recording nerve.



(a) (b)

Fig. 4. a) 3D view of BP12 stimulation protocol. Electrode 1 is active and emitting 1mA. Electrode 2 is set to ground. We can see how several nerves are affected by the stimulation. b) Voltage spread curve for BP12. Potential generated by the electrode is sampled in 49 nerves and 16 points per nerve for a total of 784 points. We see how the stimulation is not limited to the nerve closest to the electrode (nerve 1)

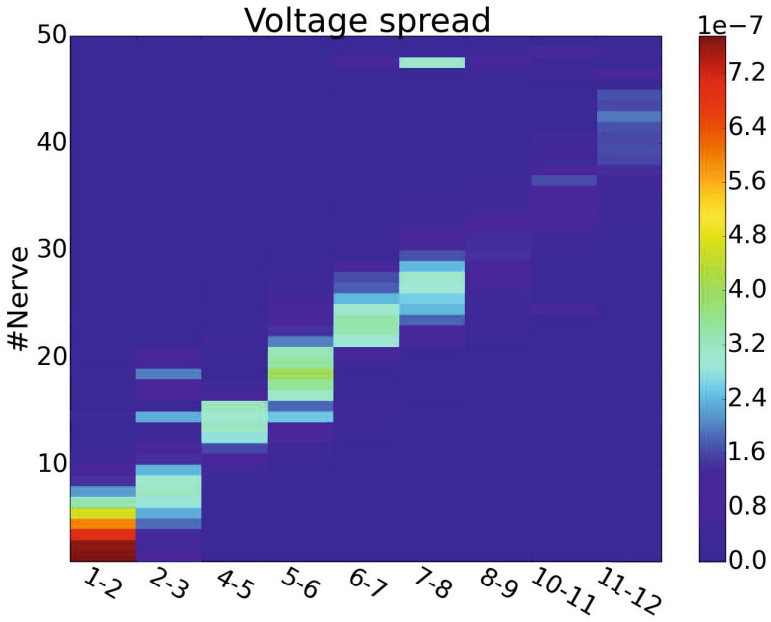


Fig. 5. Potential in each nerve for each stimulation protocol. We see how each electrode stimulates several nerves. Differences in potentials depend on the size of the electrode and its distance from the recording nerve.

In the presented results, we simulated electrode activation protocols BP1-2 to BP11-12 to show that we are able to deduce nerve activation and catch electrode cross-talk. Qualitatively, the modeled spread of excitation is in a reasonable agreement with clinical data published recently [16]. The predicted cross-talk between some of the most apical electrodes is in agreement with observations in some implanted patients [17].

The fact that, as can be seen from Figure 4b and 5, each electrodes activates more than one nerve, is at the basis of discrepancy between electrical and acoustical hearing perceptions. We thus plan to use the results of our model to predict which stimulation parameters and configurations of electrode placement are the best for each patient and thus improve the functional output of the CI surgery.

6 Discussion and Future Work

The main contribution of our work is the creation of a finite element (FE) model based on high-resolution human data and its use to predict nerve activation, which in turn allows the selection of the best electrode array for each patient from available electrode portfolios of established CI manufactures.

The model is tailored to humans and we believe that, once validated and refined, this model could be of great use for the optimization of the intracochlear position of an electrode array of an cochlear implant. We also believe it can be used for prediction of mapping in patients who cannot reliably provide auditory feedback such as, on one side, infants and young children, and on the other side, psychologically challenged patients.

This work is a step forward towards a complete personalization of CI surgery where array insertion strategy and expected response could be planned well ahead of the surgery. Virtual testing of new implants will in the future help surgeons to select the most suitable electrode array for an implant accordingly to the anatomy of the patient's cochlea. Further, being able to study the whole range of cochlear shapes and frequency distribution of the target population will lead to better fitting of implants, as well as a considerable cost reduction in the design process.

In order to assess the appropriateness of implant's electrode array design, further development should be done to define the different scenarios of the electrode array insertion, in terms of positions where it is likely to be placed and the percentage of residual hearing preserved.

Channel interaction predicted by the model will be evaluated in a follow up clinical study together with the correlation between cochlear anatomy, determined by pre-clinical CT scanning, and the spread of excitation. This will improve the selection of the best electrode array for each patient, with optimally distanced electrodes to minimize the cross-talk, from electrode portfolios available for CI surgery.

Acknowledgement. The research leading to these results received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement 304857.

References

1. World Health Organization: Deafness and hearing impairment (2012)
2. World Health Organization: The global burden of disease: 2004 update (2008)
3. Edom, E., Obrist, D., Kleiser, L.: Simulation of fluid flow and basilar-membrane motion in a two-dimensional box model of the cochlea. In: AIP Conference Proceedings, vol. 1403, p. 608 (2011)
4. Nogueira, W.: Finite element study on chochlear implant electrical activity. In: ICBT Proceeding (2013)
5. Nogueira, W., Penninger, R., Buchner, A.: A model of the electrically stimulated cochlea. In: DAGA Proceeding (2014)
6. Malherbe, T.K., Hanekom, T., Hanekom, J.J.: Can subject-specific single-fibre electrically evoked auditory brainstem response data be predicted from a model? Medical Engineering & Physics 35(7), 926–936 (2013)
7. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3), 1116–1128 (2006)
8. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: ACM Siggraph Computer Graphics, vol. 21, pp. 163–169 (1987)
9. Paulsen, R.R., Baerentzen, J.A., Larsen, R.: Markov random field surface reconstruction. IEEE Transactions on Visualization and Computer Graphics 16(4), 636–646 (2010)
10. Kjer, H., Vera, S., Perez, F., Gonzalez-Ballester, M., Paulsen, R.: Shape modeling of the inner ear from micro-ct data. In: Proceedings of Shape Symposium on Statistical Shape Models and Applications (2014)
11. Kjer, H.M., Ceresa, M., Carranza, N., Vera, S., Gonzalez-Ballester, M.A., Paulsen, R.R.: Cochlear finite element modelling, mesh quality under ssm-driven deformations. In: MeshMED Workshop - Medical Image Computing and Computer-Assisted Intervention, MICCAI 2013, pp. 70–78. Springer, Heidelberg (2013)
12. Ruokolainen, J., Lyly, M.: ELMER, a computational tool for PDEs–Application to vibroacoustics. CSC News 12(4), 30–32 (2000)
13. Råback, P., Malinen, M., Ruokolainen, J., Pursula, A., Zwinger, T.: Elmer models manual. CSC–IT Center for Science, Helsinki, Finland (2013)
14. Frijns, J.H.M., De Snoo, S., Schoonhoven, R.: Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea. Hearing Research 87(1), 170–186 (1995)
15. Oliphant, T.E.: Python for scientific computing. Computing in Science & Engineering 9(3), 10–20 (2007)
16. Vanpoucke, F.J., Boermans, P., Frijns, J.H.: Assessing the placement of a cochlear electrode array by multidimensional scaling. IEEE Transactions on Biomedical Engineering 59(2), 307–310 (2012)
17. Gani, M., Valentini, G., Sigrist, A., Kós, M.I., Boëx, C.: Implications of deep electrode insertion on cochlear implant fitting. Journal of the Association for Research in Otolaryngology 8(1), 69–83 (2007)

Appendix F

Patient Specific Simulation for Planning of Cochlear Implantation Surgery

Patient Specific Simulation for Planning of Cochlear Implantation Surgery

Sergio Vera[1,2( )], Frederic Perez[2], Clara Balust[2], Ramon Trueba[2], Jordi Rubió[2], Raul Calvo[2], Xavier Mazaira[2], Anandhan Danasingh[3], Livia Barazzetti[5], Mauricio Reyes[5], Mario Ceresa[4], Jens Fagertum[6], Hans Martin Kjer[6], Rasmus Paulsen[6], and Miguel Ángel González Ballester[4,7]

[1] Computer Vision Center, UAB, Bellaterra, Spain
sergio.vera@cvc.uab.es
[2] Alma IT Systems, Barcelona, Spain
[3] MED-EL GMBH, Wien, Austria
[4] Universitat Pompeu Fabra, Barcelona, Spain
[5] Universitat Bern, Bern, Switzerland
[6] Denmark Technische Universitet, Lyngby, Denmark
[7] ICREA - Catalan Institution for Research and Advanced Studies, Barcelona, Spain

Abstract. Cochlear implantation is a surgical procedure that can restore the hearing capabilities to patients with severe or complete functional loss. However, the level of restoration varies highly between subjects and depends on patient-specific factors. This paper presents a software application for planning cochlear implantation procedures that includes patient-specific anatomy estimation using high resolution models, implant optimization for patient-specific implant selection, simulation of mechanical and electrical properties of the implant as well as clinical reporting.

Keywords: Cochlear implant · Patient specific · Simulation · Planning

1 Introduction

A Cochlear Implant (CI) is a sound-to-electrical transducer device that can restore hearing to patients suffering hearing impairment, a condition affecting over 24 % of the population worldwide [12]. Cochlear Implants consist of a speech processor which performs filtering of the audio signal to improve the hearing of specific frequencies, and a sub-cutaneous transductor and an Electrode Array (EA) that is inserted into the cochlea and can stimulate the auditory nerve fibers, bypassing the damaged hair cells (Fig. 1).

Cochlear implantation surgery requires to gain access to the inner ear, to make the cochlea accessible, by drilling the temporal bone behind the ear. The target structure is small and the access through the middle ear is close to delicate

Miguel A. Gonzalez Ballester: This research has been funded by the European Union FP7 grant agreement no. HEAR-EU 304857.

M.G. Linguraru et al. (Eds.): CLIP 2014, LNCS 8680, pp. 101–108, 2014.
DOI: 10.1007/978-3-319-13909-8_13

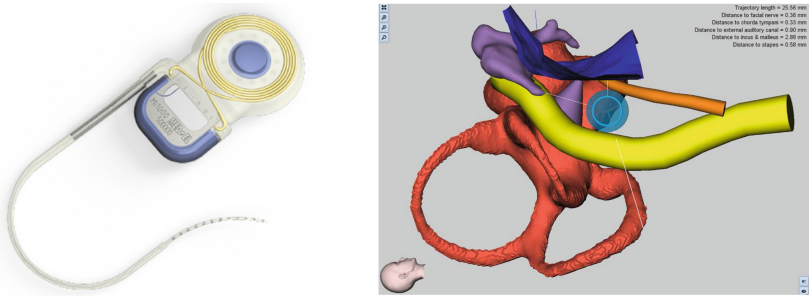


Fig. 1. (Left: Sub-cutaneal part of the cochlear implant with the transductor and electrode array. Right: Segmentation of structures of the middle and inner ear: Cochlea and semicircular canals (red) ossicles (purple), external auditory canal (blue), facial nerve (yellow) and chorda timpany (orange) (Color figure online).

structures such as the ossicles, chorda tympani and facial nerve. Careful planning of the access path considering the risk areas, is the element that decides if the electrode insertion will be performed through the membrane that covers the round window of the cochlea or through a hole drilled into the cochlea (cochleostomy). In this complex scenario, a planning software can help the surgeon to estimate the risks of the intervention and choose the best approach. Extreme care has to be taken during the insertion of the electrode array inside the cochlea. The depth and angle of insertion has to be the adequate to provide improved hearing without jeopardizing residual hearing capabilities. This is because the cochlear inner structures are delicate, and can be damaged easily by an incorrect insertion procedure. It follows that the specific anatomical variability of the cochlea of the patient plays an important role in the optimal insertion angle and depth. But the traditional Computerized Tomography (CT) or Cone Beam CT (CBCT) acquired prior to the surgery procedure cannot provide the surgeon with sufficient shape information given that the resolution of the current devices is not high enough to capture the small structures of the cochlea.

In this paper we present a software for planning electrode array insertion, that enriches conventional imaging based planning with data coming from high resolution models adapted to the patient specific anatomy. The rest of the paper is organized as follows: Sect. 2 describes the overall infrastructure of the software. Section 3 describes the modules and methods used by the application. Section 4 includes final remarks and future work.

2 Software Description

The outcome of the surgical procedure depends among other factors on the correct position of the CI's electrode array inside the cochlea and the depth of the insertion. However, conventional preoperative CT does not provide enough resolution to perform detailed analysis or simulations. High resolution models are needed to better evaluate the outcome of the procedure. The application

presented herein is designed to provide surgeons with insight of what happens inside the cochlea when the electrode is inserted.

By combining high resolution models with patient-specific information, we can use several analysis tools that would be difficult to use with the low resolution pre-clinical data. Out application closes the gap between the clinical planning stage and advanced high resolution tools applied to the electrode insertion stage. This is achieved following a workflow (Fig. 2) of tasks that starts with the patient's pre-clinical images and ends with cochlea response simulations after the implantation procedure.

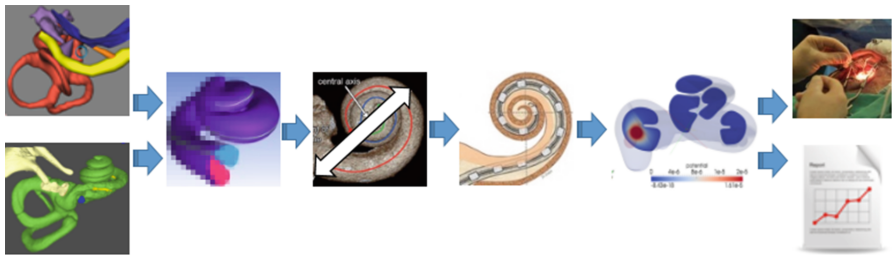


Fig. 2. Workflow of the software. From left to right: segmented structures and high resolution Statistical Shape Model as input. Patient-specific high resolution fitting. Cochlea Characterization, virtual insertion, electrical simulation, and finally, surgery and reporting.

The software runs on top of solid proven open source technologies as shown in Fig. 3. It is designed to be agnostic of operating system so it is compatible with the most popular operating systems. The Visualization Toolkit (VTK) is used as main graphical library. Qt and the Common Toolkit (CTK) are the basis of the User Interface. The communication with the clinical planning software [5] is performed using XML files defining the CT/CBCT and the segmented structures, as well as the planned path, safety volumes and any other patient relevant data.

3 Modular Structure

The software is comprised of different modules (Fig. 3) that provide individual information: patient specific high resolution anatomy model, cochlear characterization, virtual electrode insertion, electrical simulation and reporting.

3.1 Patient Specific Anatomy Model

To improve visualization and allow a more detailed modeling, a Statistical Shape Model (SSM) has been built using 17 microCT (μCT) samples of cadaveric temporal bone [7] obtained with Scanco Medical AG microCT-100 at 24 micron resolution. The inner ear structures were segmented semi-automatically using

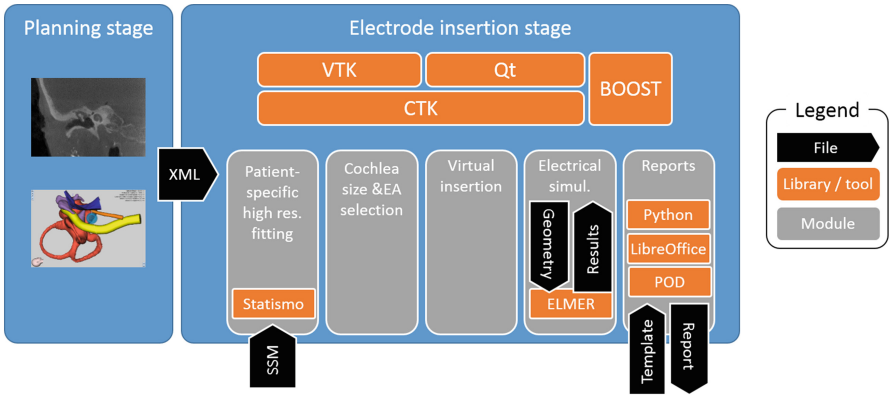


Fig. 3. Overview of the application structure, showing its modular structure as well as its software elements.

ITK-SNAP [13] and Seg3D2 [2]. The mesh resulting of the segmentations were post-processed using Markov Random Field Surface Reconstruction [10]. The datasets were registered (using Elastix [8]) to a image chosen as a reference. The transformation was applied to the reference segmentation so obtain the individual datasets with point correspondence. The SSM was built using the Statismo [9] software package. An Active Shape Model (ASM) is used to fit the high resolution model to the pre-clinical CT. The software allows inspection and generation of the SSM space through generation of specific samples (Fig. 4).

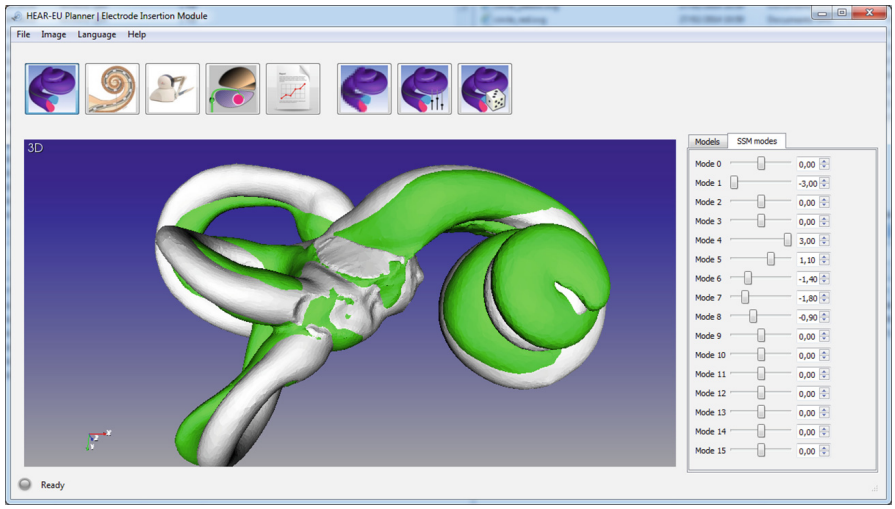


Fig. 4. Cochlear SSM loaded in the software. The mean shape of the SSM is displayed in white. Patient specific models can be generated according to the low resolution anatomy.

3.2 Cochlear Characterization

Measuring the cochlear size and shape is the first step to a correct electrode implant. The length of the cochlear duct, and the patient specific hearing impairment are key information to select the best fitting EA. The length of the unrolled cochlea has been extensively studied, and literature reports a 40 % variability with cochlear length ranging from 25 to 36 mm [6]. The final maximum insertion depth of the cochlear implant EA correlates with the diameter of the cochlea in the basal turn plane measured from the round window to the distal lateral wall [3]. This, in turn, enables the selection of the ideal electrode array from the portfolio of electrode array types that are integrated in the application (Fig. 5).

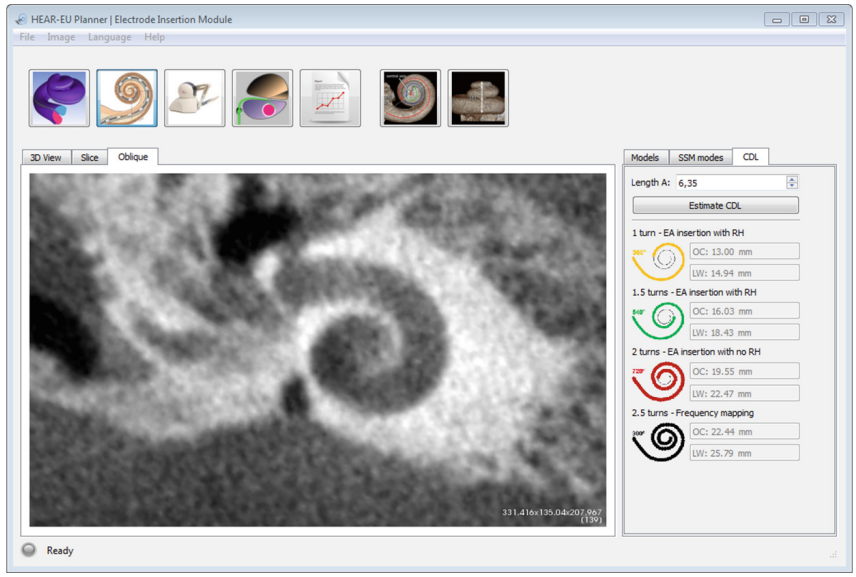


Fig. 5. Using the measurement from the diameter of the cochlea at the basal turn, the application estimates the unrolled length of the cochlea, and the different insertion depths of the electrode array.

3.3 Virtual Insertion

Once we have the patient's specific shape and a suitable electrode array has been selected, we can simulate the expected activation patterns of the implant. The last element needed for the simulation is to set the (virtual) position of the electrode array inside of the scala tympani, the chamber of the cochlea where the electrode is placed. An iterative method is used to compute the trajectory of a free-fitting electrode array, given the insertion point and direction. At each iteration the position and direction of the electrode tip with respect to the scala tympani is evaluated, ensuring that the tip proceeds tangentially and its distance from the wall is at least equal to the array radius. At each step the angle of impact

to the wall and the margin between the cochlear implant array and the cochlear walls are evaluated too, providing an indirect measure of pressure against the wall. The iteration can stop prematurely if the electrode does not fit in the scala tympani dimensions or if it is subjected to excessive bending (Fig. 6).

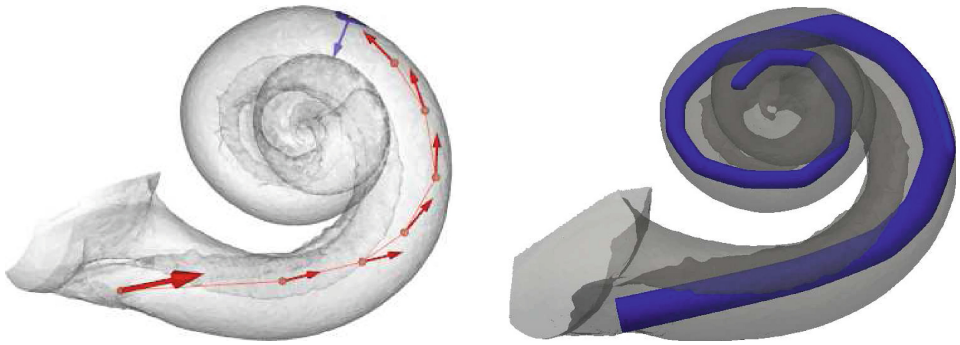


Fig. 6. Virtual insertion. Left: At each step of the insertion simulation, the tip position respect to the wall is evaluated and the direction is adjusted in order to lie tangentially to the wall. Right: Simulated electrode insertion. The final trajectory of the electrode is tangential to the scala tympani wall.

3.4 Electrical Simulations

The placed electrode is the last required step to perform the electrode simulations [1]. The simulation is performed using the multiphysics Finite Element Method (FEM) open source solver software ELMER [11]. In its current stage, the software can simulate bipolar simulation protocols (Fig. 7, left), where one electrode emits electrical current and the other is set to ground. Simulations also include modelizations of the electrical properties of nerve the fibers that start at the organ of Corti in the basilar membrane and form the auditory nerve, using the Generalized Schwarz-Eikhof-Frijns (GSEF) model [4] (Fig. 7, right).

3.5 Reporting

During the planning process, the operator has the option to save screenshots, possibly annotated with relevant information. After the process, the commented screenshots, along with the patient's clinical data, a Portable Document Format (PDF) report is generated for clinicians to review. The generation of the report employs the open source LibreOffice engine and POD (Python Open Document)[1] library to generate the report. For the generation of the reports with these technologies, a series of document templates are created that include embedded

[1] http://appyframework.org/pod.html.

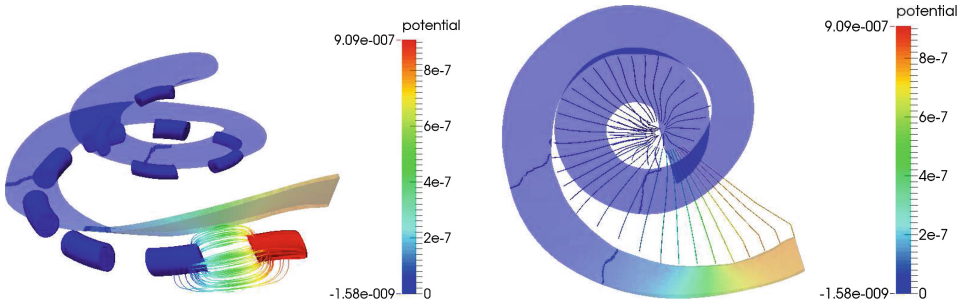


Fig. 7. Visualization of the simulation results. The basilar membrane has been rendered semitransparent for ease of visualization. Left: bipolar stimulation protocol of first two electrodes. Right: Nerve fiber stimulation after electrode activation pattern

Python code inserted into the document structure. The templates are post-processed using a Python script that can execute the embedded Python code and perform the adequate substitution of the variables. These variables include patient information and user generated screenshots and captions (Fig. 8).

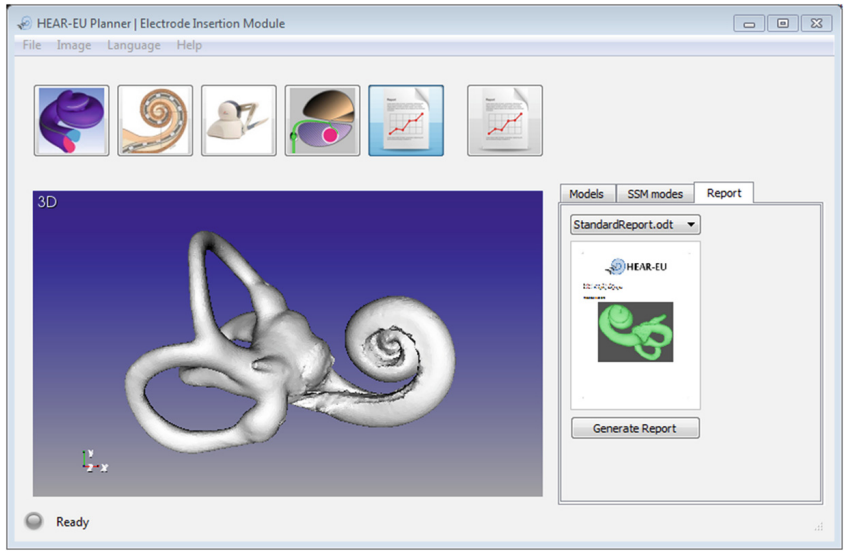


Fig. 8. Report generation interface.

4 Conclusions and Future Work

We have presented a software for the estimation of the patient specific inner ear and intra-cochlear anatomy, the planning and simulation of both the electrode

insertion procedure, and the outcome of the surgery to the hearing capabilities for the patient. The software represents also a tool for the selection of the best electrode array for the patient and the reporting of the surgical procedure, making it a helping tool in the clinical practice. While the software is still evolving, it represents a collaborative effort in integrating many medical imaging tools, bringing the pre-surgery planning to a new level of information analysis.

Future work includes additional integration with more electrode models and tools from the electrode manufacturer, improvements on the virtual insertion phase using real-time simulation, and validation of the electrical simulations using audiometric tests are some of the future tasks planned for the software.

References

1. Ceresa, M., Mangado, N., Dejea, H., Carranza, N., Mistrik, P., Kjer, H.M., Vera, S., Paulsen, R., Ballester, M.A.: Patient-specific simulation of implant placement and function for cochlear implantation surgery planning (2014)
2. CIBC, Seg3D: Volumetric Image Segmentation and Visualization. Scientific Computing and Imaging Institute (SCI) (2014). http://www.seg3d.org
3. Escudé, B., James, C., Deguine, O., Cochard, N., Eter, E., Fraysse, B.: The size of the cochlea and predictions of insertion depth angles for cochlear implant electrodes. Audiol. Neurotol. **11**(1), 27–33 (2006)
4. Frijns, J., De Snoo, S., Schoonhoven, R.: Potential distributions and neural excitation patterns in a rotationally symmetric model of the electrically stimulated cochlea. Hear. Res. **87**(1), 170–186 (1995)
5. Gerber, N., Bell, B., Gavaghan, K., Weisstanner, C., Caversaccio, M., Weber, S.: Surgical planning tool for robotically assisted hearing aid implantation. Int. J. Comput. Assist. Radiol. Surg. **9**, 11–20 (2014)
6. Hardy, M.: The length of the organ of Corti in man. Am. J. Anat. **62**(2), 291–311 (1938)
7. Kjer, H.M., Fagertun, J., Vera, S., Pérez, F., González-Ballester, M.A., Paulsen, R.R.: Shape modelling of the inner ear from micro-CT data. In: SHAPE 2014 (2014)
8. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: elastix: A toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging **29**(1), 196–205 (2010)
9. Lüthi, M., Blanc, R., Albrecht, T., Gass, T., Goksel, O., Büchler, P., Kistler, M., Bousleiman, H., Reyes, M., Cattin, P., Vetter, T.: Statismo - a framework for pca based statistical models, July 2012
10. Paulsen, R., Bærentzen, J., Larsen, R.: Markov random field surface reconstruction. IEEE Trans. Vis. Comput. Graph. **16**(4), 636–646 (2010)
11. Ruokolainen, J., Lyly, M.: ELMER, a computational tool for PDEs-Application to vibroacoustics. CSC News **12**(4), 30–32 (2000)
12. World Health Organization. Deafness and hearing impairment (2012)
13. Yushkevich, P., Piven, J., Cody, H., Ho, S., Gee, J.C., Gerig, G.: User-guided level set segmentation of anatomical structures with ITK-SNAP. NeuroImage **31**, 1116–1128 (2005)

Appendix G

Monopolar Stimulation of the Implanted Cochlea: a Synthetic Population-Based Study

Monopolar stimulation of the implanted cochlea: a synthetic population-based study

Nerea Mangado[1], Mario Ceresa[1], Hector Dejea Velardo[1], Hans Martin Kjer[2], Sergio Vera[3], Rasmus R. Paulsen[2], Jens Fagertun[2], Pavel Mistrik[4], Gemma Piella[1] and Miguel Angel Gonzalez Ballester[1,5]

[1] Simbiosys group, Universitat Pompeu Fabra, Barcelona, Spain
[2] Denmark Technical University, Copenhagen, Denmark
[3] Alma Medical Imaging, Barcelona, Spain
[4] Med-EL, Innsbruck, Austria
[5] ICREA, Barcelona, Spain

Abstract. Cochlear implantation is carried out to recover the sense of hearing. However, its functional outcome varies highly between patients. In the current work, we present a study to assess the functional outcomes of cochlear implants considering the inter-variability found among a population of patients. In order to capture the cochlear anatomical details, a statistical shape model is created from high-resolution human μCT data. A population of virtual patients is automatically generated by sampling new anatomical instances from the statistical shape model. For each virtual patient, an implant insertion is simulated and a finite element model is generated to estimate the electrical field created into the cochlea. These simulations are defined according to the monopolar stimulation protocol of a cochlear implant and a prediction of the voltage spread over the population of virtual patients is evaluated.

1 Introduction

Over 5 % of the worldwide population over the age of 45 years suffer from severe hearing impairment, thus they would be considered eligible for cochlear implantation (CI) surgery [19]. However, there is a high variability in the outcomes of CI since patient-specific factors influence directly the level of hearing restoration that the implant can achieve. Consequently, an accurate prediction of the surgery outcome of the patient is needed to estimate the performance of the cochlear implant. Although computational models have not been applied as a common technique into the clinical practice of CI, some authors have reported promising results predicting its outcomes [16,11,2]. Specifically, we have previously presented in-silico studies with promising results for patient-specific cases, where the outcomes of a personalized CI model were assessed [2,3,9]. However, the developed automatic framework has the potential to predict CI outcomes not only for patient-specific cases, but also for a more complete virtual study of the population. This is especially useful to carry out evaluations on the implant performance among a group of patients in order to be able to optimize CI electrode array design to the widest range of the population possible.

In this work, a statistical shape model (SSM) has been created from high-resolution μCT data to capture inter-patient variability and to provide a computational tool for virtual patient sampling. Special attention has been given to the insertion depth of the electrode array of the cochlear implant since it highly contributes to the variability in CI outcomes [13]. We presented a virtual insertion algorithm which physically deform the electrode array according to the geometry of the cochlear anatomy of the patient and at the same time, it allows controlling surgical insertion parameters, such as the depth of insertion of the electrode array [4,10]. This virtual insertion approach is included within the automatic framework proposed which allows obtaining a full finite element model of the CI and therefore, carrying out the computational electrical simulations [9]. We have improved the computational method by using a more detailed model of the cochlea with respect to our previous work [2]. In addition, we obtain an accurate insertion by using a surgical simulator software to compute the final position of the electrode. Thus, we believe that a more realistic virtual insertion is achieved and consequently, more accurate results of the electrode stimulation can be obtained. Using our framework, we can get realistic CI models and predict the voltage spread in the cochlea created by the stimulation protocol under study. Applying this framework to a group of virtual patients, information about the nerve activation zones after CI can be assessed. Therefore, valuable information for electrode design and stimulation optimization is obtained.

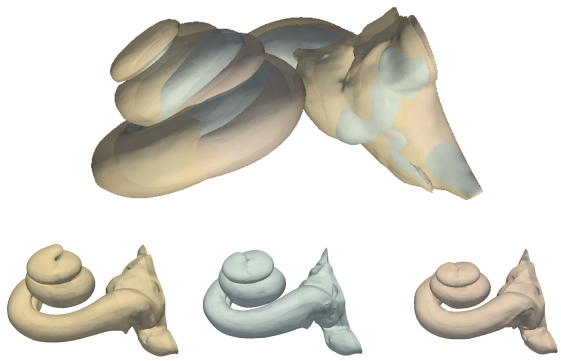


Fig. 1. Three virtual patients from the statistical shape model are overlapped to show the inter-patient variability on the cochlear shape.

2 Generation of computational models

The framework includes a cochlear Statistical Shape Model (SSM) for generation of a virtual patient anatomy (Fig. 1), a virtual insertion of the electrode array, and a procedure for generating the volumetric finite element (FE) mesh of the cochlea with realistic nerve fibers. Additional background information is found in [9]. In comparison, some changes have been made to obtain an even more realistic and detailed computational model.

Firstly, the SSM is generated from a more suited anatomical reference and with an improved registration procedure [6], allowing the SSM to capture the cochlear population variability in a more satisfactory manner. Most notably, the semi-circular canals are no longer included. The model extends far enough into the vestibule to include the oval and round window. Even though the ending in vestibule is rough and abrupt, the change to the new reference model is motivated and justified by the addition of the well-defined cochlear partition (i.e. a basilar membrane approximation) present in this particular dataset [1]. This provides additional realism to the anatomical model, and facilitates a change to the procedure for virtual placement of the electrode array.

The electrode position of the real cochlear implantation procedure has been computed by means of a planning simulator software. It consists on real-time simulations based on a deformation model which includes the mechanical properties of both electrode and cochlea and a collision model [18]. Afterwards, the virtual insertion algorithm is applied over the original electrode geometry. This algorithm allows obtaining a deformation for the electrode array according to this surgical insertion position [4,10]. Thus a final electrode mesh is obtained with a realistic placement of the implant for the given patient. This electrode array mesh consisted in a Med-EL Flex28 design, with 12 stimulating channels (electrodes) and a length of 28 mm.

Within the automatic framework, 100 nerve fiber bundles were generated according to the patient's anatomy and an outer box was created to model as the surrounding bone of the cochlea. Finally all elements were merged and transformed into a single volumetric mesh (Fig.2). This procedure was repeated in an automatic way for each of the virtual patients sampled from the SSM.

3 Finite element simulation: electrical model and stimulation protocol

For the FE electrical simulation, the static current conduction solver of the open source multiphysics Elmer software has been used [12]. Maxwell's equations are defined in the quasistatic approximation and the electrical porential is obtained by solving the Poisson equation. Both Dirichlet or Neumann boundary conditions can be used to describe the electric potential, describing the potential and the current values on the boundary, respectively [14].

Three stimulation protocols can be set up in a cochlear implant according to the electrode configuration. In this work, we use the monopolar stimulation

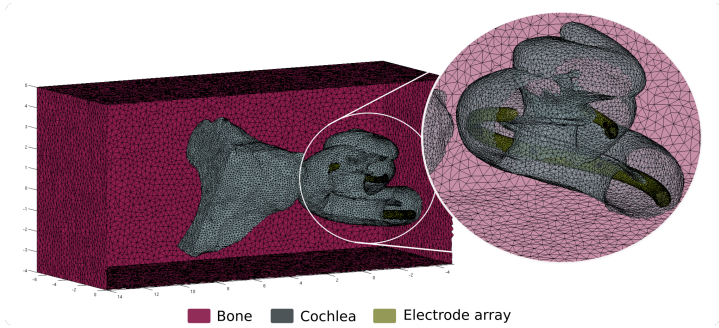


Fig. 2. Finite element mesh obtained for a single patient. A cut of the element faces is displayed for visualization purpose.

(Fig. 3). In this configuration, one electrode is activated emitting current while the bone surrounding the cochlea has been set to ground. For all models, the value of the current stimulation is 1mA [2]. The conductivity parameters of the cochlea structures defined for the electrical simulation were chosen according to [8,15]. Each simulation was run in steady state formulation and comprised one activated electrode, thus resulting in 12 simulations per virtual patient.

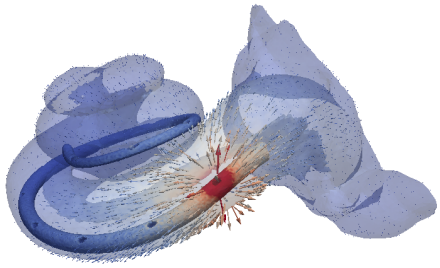


Fig. 3. Illustration of the monopolar stimulation. The first electrode has been activated and the volume current direction is shown.

4 Results

A total of 30 virtual patients were sampled randomly from the SSM and studied under a monopolar stimulation. The electrical simulation framework was

run automatically, thus 30 electrical simulations were finally obtained. Since the cochlear shape varies between patients, different lengths of virtual insertion depth were obtained. The length obtained was 25.2 ± 1.2 mm with a number of turns of 1.56 ± 0.04, corresponding to 563 ± 15^o. The virtual insertion algorithm was successfully run in all cases. However, changes in the element area of the electrode array mesh were observed, which prompted us to further quantify these local geometry changes (see Fig.4). The average changes of element area for all virtual patients evaluated were -4.6 ± 3.9%.

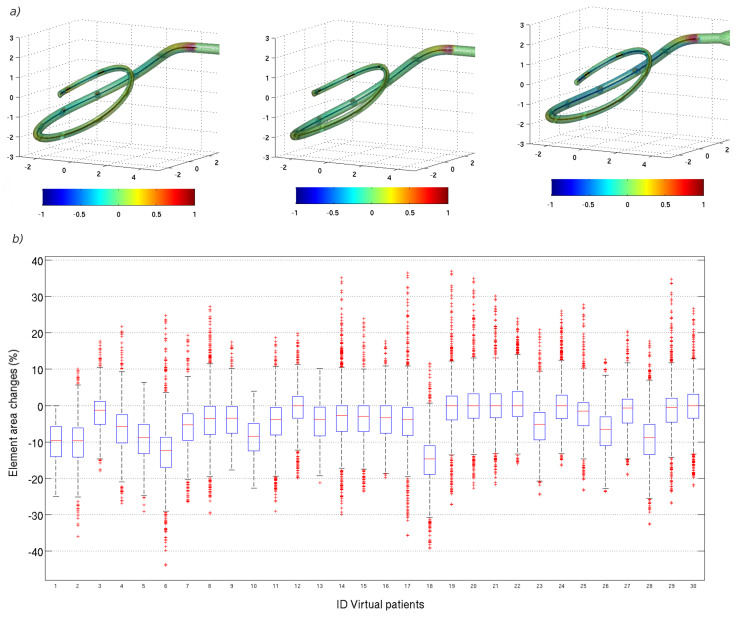


Fig. 4. Local changes on the area of each electrode mesh element after the deformation by the virtual insertion. (a) Changes are represented over the surface of the electrode in a scale of -1 to 1, being the maximum decrease and increase, respectively, compared to the area before the deformation. (b) Central mark of the box is the median and its edges the 25th and 75th percentiles of the element area changes of each of the 30 virtual patients.

The generation of the computational CI model took 228 ± 18 seconds, obtaining a volumetric mesh of 1.2×10^6 ± 7×10^4 of tetrahedral elements with a mesh quality of 0.785 ± 0.001. The mesh quality of each model was assessed by computing the aspect ratio of each element, expressed in a range from 0 to 1,

corresponding to nearly degenerated mesh element and perfect one, respectively [7].

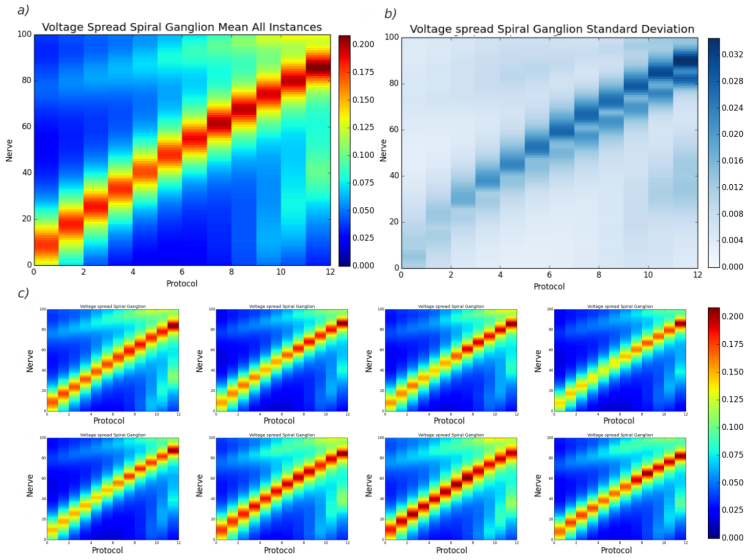


Fig. 5. Potential (V) generated for each stimulation protocol (horizontal axis) in each nerve fibers (vertical axis). (a) Mean and (b) standard deviation of the voltage spread for all virtual patients evaluated. (c) Examples of the voltage spread for a single patient, where differences can be appreciated due to changes in cochlear anatomy. Patient ID shown in (c) are respectively 1,2,9,11,15,17,23 and 30.

12 electrical simulations were run for each model, for a total of 360 runs. Each of them achieved convergence on a workstation of our institution, an Intel Core i5 of 16GB. Figure 5 shows the electric field for each nerve fiber under the stimulation of the 12 monopolar stimulation protocols. The magnitude of the potential depends on its distance from the nerve. It can be observed that some stimulated zones are located far from the perfect diagonal of stimulation. This implies that each electrode does not exclusively activate the most nearby nerve fiber.

This effect is called cross-talk and it is a reason of discrepancy between electrical hearing perceptions in patients with a cochlear implant and normal, acoustical, hearing. All these virtual patients have in common the cross-talk presented in the apical part of the cochlea. Therein, the nerves located in the

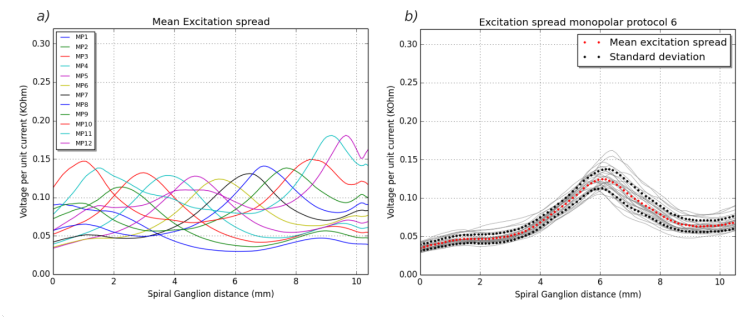


Fig. 6. (a) Mean excitation spread measured along the spiral ganglion. (b) Excitation spread of monopolar stimulation protocol 6 for all virtual patients.

basal part are nonspecifically activated by electrodes 10 to 12 (see Fig. 5). This corresponds to cross-turn stimulation.

We show in Figure 6(a) the mean excitation spread along the spiral ganglion (anatomical structure composed of soma for all neural fibers). Each curve corresponds to one of the 12 monopolar stimulation protocol. Figure 6(b) shows in detail the excitation spread of the monopolar stimulation 6 for all 30 patients, where the sixth electrode has been activated.

5 Discussion and Future Work

The main contribution of this work is the CI assessment on a population of virtual patients sampled from a SSM. As far as we know, this is the first population-based study to evaluate the results of a CI electrical simulation. Additionally, we have improved the CI model with respect to our previous work [2], providing a more realistic finite element model based on high-resolution data, real electrode array design and virtual surgical placement. The virtual insertion has proved to be consistent in all cases tested, so a realistic mesh deformation after the virtual insertion is obtained. Simulations have been run successfully in all cases, obtaining results in agreement with previous reported clinical results [17], including the cross-talk zones [5].

Nonetheless, our work has some limitations. The mean excitation spread evaluated along the spiral ganglion has some discrepancies compared to literature [15]. Even though the behaviour is similar and shows a general tendency, we believe that some work needs to be done regarding the geometrical nerve generation since their position could modify the results obtained from the electrode stimulation. Despite this, we do believe that this work is a step closer to the accurate prediction of the nerve activation.

The results obtained help to better explain the behaviour of the excitation spread within a group of patients and to observe the variations obtained accounting for the inter-patient anatomy variability. This framework has promising potential to optimize stimulation parameters and electrode placement that better fit the anatomy and level of impairment of each patient to provide the best functional outcome possible. In future work, other sources of variability can be taken into account. For example the implant placement or the electrode array configuration which would provide additional valuable information in the process of optimizing the CI.

6 Acknowledgement

References

1. Braun, K., Böhnke, F., Stark, T.: Three-dimensional representation of the human cochlea using micro-computed tomography data: presenting an anatomical model for further numerical calculations. Acta Otolaryngol 6(132), 603–13 (2012)
2. Ceresa, M., Mangado, N., Dejea, H., Carranza, N., Mistrik, P., Kjer, H., Vera, S., Paulsen, R., González Ballester, M.: Patient-specific simulation of implant placement and function for cochlear implantation surgery planning. In: MICCAI, LNCS. vol. 8674, pp. 49–56 (2014)
3. Ceresa, M., Mangado, N., J.Andrews, R., Ballester, M.A.G.: Computational models for predicting outcomes of neuroprosthesis implantation: the case of cochlear implants. Journal of Molecular Biology (Accepted) (2015)
4. Duchateau, N., Mangado, N., Ceresa, M., Mistrik, P., Vera, S., González Ballester, M.: Virtual cochlear electrode insertion via parallel transport frame. In: Proceedings of International Symposium on Biomedical Imaging (2015)
5. Gani, M., Valentini, G., Sigrist, A., K'os, M., Bo"ex, C.: Implications of deep electrode insertion on cochlear implant fitting. J Assoc Res Otolaryngol 8, 69–83 (2007)
6. Kjer, H., Vera, S., Fagertun, J., Gil, D., González-Ballester, M., Paulsen, R.: Image registration of cochlear microct data using heat distribution similarity. In: Image Analysis, Lecture Notes in Computer Science, vol. 9127, pp. 234–245. Springer International Publishing (2015)
7. Liu, A., Joe, B.: Relationship between tetrahedron shape measures. BIT Numerical Mathematics 34(2), 268–287 (1994)
8. Malherbe, T.K., Hanekom, T., Hanekom, J.J.: Can subject-specific single-fibre electrically evoked auditory brainstem response data be predicted from a model? Medical engineering & physics 35(7), 926–936 (2013)
9. Mangado, N., Ceresa, M., Duchateau, N., Dejea Velardo, H., Kjer, H., R.Paulsen, R., Vera, S., Mistrik, P., Herrero, J., González Ballester, M.: Automatic generation of a computational model for monopolar stimulation of cochlear implants. In: Proceedings of Computer Assisted Radiology and Surgery (2015)
10. Mangado, N., Duchateau, N., Ceresa, M., Kjer, H., Vera, S., Mistrik, P., Herrero, J., González Ballester, M.: Patient-specific virtual insertion of electrode array for electrical simulations of cochlear implants. In: Proceedings of Computer Assisted Radiology and Surgery (2015)

11. Nogueira, W.: Finite element study on chochlear implant electrical activity. In: ICBT Proceeding (2013)
12. Råback, P., Malinen, M., Ruokolainen, J., Pursula, A., Zwinger, T.: Elmer models manual. CSC–IT Center for Science, Helsinki, Finland (2013)
13. Roland Jr, J.T.: Cochlear implant electrode insertion. Operative Techniques in Otolaryngology-Head and Neck Surgery 16(2), 86 – 92 (2005)
14. Ruokolainen, J., Lyly, M.: ELMER, a computational tool for PDEs–Application to vibroacoustics. CSC News 12(4), 30–32 (2000)
15. Saba, R., Elliott, S.J., Wang, S.: Modelling the effects of cochlear implant current focusing. Cochlear Implants International 15(6), 318–326 (Nov 2014)
16. Smit, J.E., Hanekom, T., Hanekom, J.J.: Estimation of stimulus attenuation in cochlear implants. Journal of Neuroscience Methods 180(2), 363–373 (2009)
17. Vanpoucke, F., Boermans, P., Frijns, J.: Assessing the placement of a cochlear electrode array by multidimensional scaling. IEEE Trans Biomed Eng 59, 307–10 (2012)
18. Vera, S., Caro, R., Perez, F., Bordone, M., Herrero, J., Kjer, H., Fagertun, J., Paulsen, R., Dhanasingh, A., Barazzetti, L., Reyes.M, González Ballester, M.: Cochlear implant planning, selection and simulation with patient specific data. In: Proceedings of Computer Assisted Radiology and Surgery (2015)
19. World Health Organization: Deafness and hearing impairment (2012)

Appendix H

Multi-Region Statistical Shape Model for Cochlear Implantation

Multi-Region Statistical Shape Model for Cochlear Implantation

Jordi Romera Romero*[a], H. Martin Kjer[c], Mario Ceresa[a], Miguel A. González Ballester[a,b]
[a]Simbiosys Group, Universitat Pompeu Fabra, Barcelona, Spain**
[b]ICREA, Barcelona, Spain
[c]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Copenhagen, Denmark

ABSTRACT

Statistical shape models are commonly used to analyze the variability between similar anatomical structures and their use is successfully established as a tool for the analysis or segmentation of medical images. However, just using a simple model to capture the global variability of complex structures is not enough to achieve the best results. Typically, the anatomical variability between structures is associated to the variability of their physiological regions. In this paper, a complete pipeline is proposed for building a statistical shape model to study the local variability of physiological regions of the inner ear. The proposed model, which is based on an extension of the Point Distribution Model (PDM), is built for a training set of 17 high-resolution images (24.5 μm) of the inner ear. The model is evaluated according to its generalization ability. The results are compared with the results of a global model (without regions) built directly using a PDM. The evaluation results suggest that better accuracy can be achieved using a regional modeling of the inner ear.

Keywords: Statistical Shape Model, Anatomical regions, Patient specific modeling, Inner Ear, Cochlear Implantation.

1. DESCRIPTION OF PURPOSE

Hearing loss is among the most common reasons for disability. Worldwide, 27% of men and 24% of women above the age of 45 suffer from severe hearing impairment[1]. The situation of these people could be improved or restored with a medical treatment such as Cochlear Implantation (CI) surgery. However, the level of restoration and implantation design is highly dependent on patient-specific factors.

Consequently, an accurate prediction of the surgery outcome of the patient is needed to estimate the performance of the cochlear implant. Statistical Shape Models (SSM) from high-resolution micro-CT images are great for generating anatomical and computational models of the inner ear, which have formidable applications for the optimization of hearing restoration with CI[2]. They have the potential for capturing the global inter-patient variability, to provide a computational tool for virtual patient instantiation and to predict patient-specific CI outcomes.

However, global modeling of complex anatomical structures as the inner ear is a difficult task just with a standard PDM[3], since the amount of micro-CT data is limited. Typically, the global anatomical variability between complex structures of the inner ear is associated to the local variability of their corresponding physiological regions, i.e. the global variability can be explained in a local form. In this work, we hypothesize that if we construct a Multi-Region PDM of the inner ear, by independently modeling each of its anatomical regions, we will be able to break the global constraints in the variability and subsequently achieve better estimations in a given virtual patient instantiation problem.

The goal of this work is then to construct a Multi-Region SSM of the inner ear from high-resolution μCT data to capture the regional inter-patient variability and to evaluate and compare it with a classic SSM.

2. METHODS

In the following sections we give a brief overview of the original data and the methodology that we follow to create a multi-region SSM and to validate it.

2.1. Dataset, Imaging and Image Processing

A total of 17 temporal bones were scanned with a micro-CT system (Scanco Medical, Switzerland), processed and downsampled to contain the inner ear region in 24.5 micron voxels. The inner ear was segmented as a single object and a

* jordi.romera01@estudiant.upf.edu
** The research leading to these results received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement 304857.

surface model was extracted for each dataset. To build SSMs we need point-to-point correspondence between the surface models, which we obtain adopting a similar process to the one presented by Frangi et al.[4]. Aligning each individual segmentation dataset to a chosen reference using affine plus intensity-based deformable registration (Diffeomorphic Demons algorithm[5]) and applying the resulting transformations to the reference surface model.

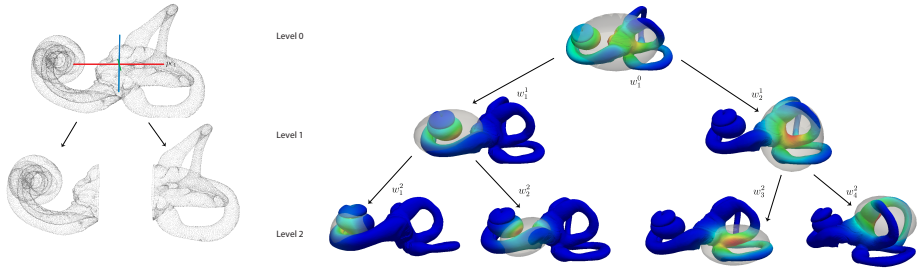


Figure 1. Region separation approach. Left: first iteration of the algorithm over the contour vertices of a reference surface. Right: hierarchical binary tree of inner ear regions example. Regions boundaries are calculated using Gaussian weights[6].

2.2. Construction of the Point Distribution Model

To construct a PDM[3], first we use the Generalized Procrustes Analysis (GPA) approach to align the shapes of the training set to a common reference coordinate system with respect to rotations and translations and to obtain the representation of the mean shape, $\bar{x}$. The next step is to obtain the principal modes of variation in shape by using the Principal Component Analysis (PCA) approach. The modes of variation, P, and some model coefficients, b, can be used to sample new shapes in the allowable shape domain:

$$x_i = \bar{x} + Pb. \tag{1}$$

In our work, we have built a global PDM of the inner ear using the entire surface models (obtained as seen in Section 2.1) to be evaluated and compared with the proposed multi-region PDM approach detailed in succeeding sections.

2.3. Region separation

The multi-region SSM that we present in this paper is based on the concept of PDM, but it uses multiple PDMs for modeling local regions with an anatomical meaning. The first step to construct a multi-region model is the identification and splitting of regions in the training surface models with point correspondences. To identify anatomical regions we adopt the approach presented by Seiler et al.[6] that automatically identifies anatomical regions at different scales in the form of a hierarchical binary tree. The approach uses PCA to obtain the principal axis from the vertex distribution of a reference dataset and classifies each of its vertices into two new groups (regions) along the first principal axis, as shown in Fig. 1-left. Applying this approach in a recursive manner we can get the hierarchical tree representation of regions. To provide greater dynamics in the delineation of regions so as to ensure an overlapping area between them, each region is modeled independently using a Gaussian distribution, w_k, (as shown in Fig. 1-right) with mean and covariance computed with the PCA approach.

To split each surface mesh of the training set we just need the Gaussian weight distributions of the desired regions and to heuristically define some thresholds that determine which vertices lay inside and outside of each region. Using the same weight distributions and the same thresholds we can assure that the splitting is the same in each surface model and that the vertex correspondences between surface regions is maintained. After the splitting procedure we get new regional training sets of surfaces that can be modeled independently using the PDM approach.

2.4. Region merging

After the construction of each local or regional PDM (rPDM) we can use model coefficients to instantiate new valid surface regions that remain similar to those in each regional training set. The purpose is now to have a procedure for regaining the global representation of the original object by combining and merging the regional instances.

The merging process employs an interpolation of the overlapping vertices to produce smooth transitions between the sampled regions. For this to work properly, the overlapping vertices must be close or roughly aligned. This can be done in numerous ways, for instance by rigidly aligning these corresponding and shared vertices.

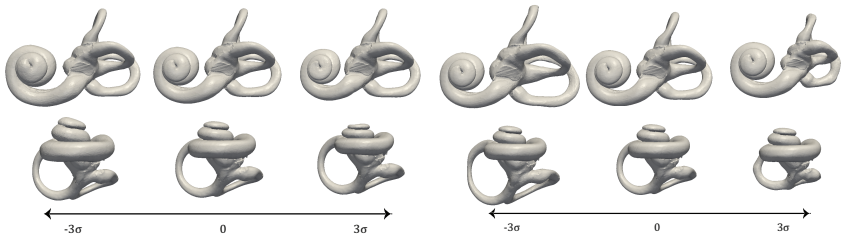


Figure 2. Samples of the two-regional PDM model of the inner ear. Left: first variation mode of the cochlear rPDM merged with the mean shape of the semicircular canals rPDM. Right: first variation mode of the semicircular canals rPDM merged with the mean shape of the cochlear rPDM.

To perform the interpolation we use a weighted sum of the corresponding vertex coordinates in the overlapping area:

$$x_j = \sum_{k=1}^{n_r} \overline{w}_{kj} x_{kj}\,,\ \text{for } j \in (\Omega_1 \cap \Omega_2 \dots\ \Omega_{n_r}), \tag{2}$$

where $\overline{w}_{kj}$ are the normalized weights of w_{kj} for the jth vertices in the overlapping area of the kth region:

$$\overline{w}_{kj} = \frac{w_{kj} - min(w_{kj})}{max(w_{kj}) - min(w_{kj})},\ \text{for } j \in \left(\Omega_1 \cap \Omega_2 \dots\ \Omega_{n_r}\right) \tag{3}$$

and

$$\sum_{k=1}^{n_r} \overline{w}_{kj} = 1\,,\ \text{for } j \in \left(\Omega_1 \cap \Omega_2 \dots\ \Omega_{n_r}\right). \tag{4}$$

2.5. Evaluation

The evaluation of the model is carried out using the property of generalization ability of a statistical shape model. The generalization ability measures the capability of a model to estimate unknown instances of surfaces of the same class. This property is critical and tells us the precision in which the model has captured the real inter-patient variability from a limited set of training shapes.

To measure the generalization ability of the proposed multi-region model we use a leave-one-out cross-validation evaluation. We should measure the accuracy achieved when using a multi-region model to estimate global unknown surfaces of the inner ear. For a given target surface, x_i, we use each rPDM to sample each corresponding target region of the surface using a determined number of variation modes, M. For this purpose, first we rigidly align each rPDM mean, $\bar{x}_i^k$, to the corresponding target region, x_i^k, and we compute the M least squares model coefficients:

$$b_i^k(M) = P_i^k(M)^T \left(x_i^k - \bar{x}_i^k\right),\ \text{for } k \in [1, n_r]. \tag{5}$$

After the instantiation, the overlapping vertices are assumed to be close enough so we can directly merge the estimated regions using the weighted sum of the vertices in that area (Eq. 2). Finally, to measure the accuracy of the global estimation we compute the sum of squared differences (SSD) between the target surface vertices, x_i, and the estimated ones, x_i'. The metric is given as a function of the number of variation modes, M, used during the estimation[7]:

$$G(M) = \frac{1}{N} \sum_{i=1}^{N} |x_i - x_i'(M)|^2. \tag{6}$$

It is noteworthy to mention that this evaluation process is also valid for the evaluation of a standard PDM. In that case the estimation is done using a simple PDM that models the entire shape of the inner ear and the evaluation of the generalization capacity is then identical to the one presented by Davies et al.[7].

3. RESULTS

Once we have obtained the surface meshes (as seen in Section 2.1) with 53.6k vertices each, we have constructed a two-regional model build from the regions identified in the first level of the hierarchical tree (cochlea and semicircular canals). We have selected this level since these regions maintains a more intuitive anatomical meaning. The first principal mode of variation for both the cochlea and the semicircular canals is shown in Fig. 2. We can note that a region based model produces modes of variation with isolated or local effects.

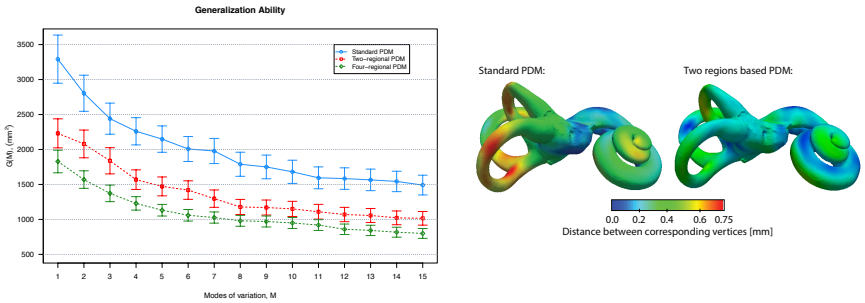


Figure 3. Comparison between the Multi-Region PDM and the standard PDM. Left: generalization ability measured for the standard PDM (blue), the two-regional PDM (red) and a four-regional PDM (green). Right: vertex distances between a randomly used target shape and the estimations using the standard PDM (left) and the two-regional PDM (right).

The generalization ability evaluation is shown in Fig. 3-left. We have measured it for the standard PDM, the two-regional PDM and an additional four-regional PDM. We can note that a multi-regional model can estimate unknown shapes of the inner ear with more precision than the standard PDM. In advance, it is also important to note that using more regions can further reduce the estimation error, but further attention and special care should be given then to the region merging process to avoid surface or transition artifacts. Fig. 3-right also shows the refinement in the estimation of a given surface in terms of vertex distances when using the two-regional based model instead of the standard model.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

The use of a classical SSM for complex anatomical structures presents the problem of how to separate the variability contribution for each of their anatomical regions. In this paper we present a method, which extends from the theory of Point Distribution Models, devised to statistically model complex objects by capturing their global variability in a local manner. The pipeline comprises three main functionalities. First, the identification and splitting of anatomical regions in training structures with point correspondences. Second, the application of local PDMs to capture and sample the desired regional deformations. And finally, the merging of sampled regions to recover the entire structure. Additionally, we present an approach to validate the generalization capability of the new model with respect to the standard PDM.

5. CONCLUSIONS

In this paper we proposed and built a multi-region statistical shape model of the inner ear based on anatomical regions, primarily the cochlea and the semicircular canals. This has been validated and compared with the standard PDM model by assessing the generalization capacity. The results show that the multi-region model provides greater control of deformations and therefore it is more accurate in the reconstruction of new samples. This suggests that a significant improvement in the estimation or segmentation accuracy could be achieved when compared to the standard model.

REFERENCES

[1] World Health Organization: Deafness and hearing impairment (2012)

[2] Noble, J. H., Labadie, R. F., Gifford, R. H., & Dawant, B. M. (2013). Image-guidance enables new methods for customizing cochlear implant stimulation strategies. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, *21*(5), 820-829.

[3] Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, *61*(1), 38-59.

[4] Frangi, A. F., Niessen, W. J., Rueckert, D., & Schnabel, J. A. (2001, January). Automatic 3D ASM construction via atlas-based landmarking and volumetric elastic registration. In *Information Processing in Medical Imaging* (pp. 78-91). Springer Berlin Heidelberg.

[5] Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, *45*(1), S61-S72.

[6] Seiler, C., Pennec, X., & Reyes, M. (2012). Capturing the multiscale anatomical shape variability with polyaffine transformation trees. *Medical Image Analysis*,*16*(7), 1371-1384.

[7] Davies, R. H. (2002). Learning shape: optimal models for analysing natural variability. University of Manchester.

Appendix I

Patient Specific Estimation of Detailed Cochlea Shape From CT Images

Patient Specific Estimation of Detailed Cochlea Shape From CT Images

IN PREP. - TBD

Abstract: The anatomical information about the inner ear gained from pre- and post-operative CT scan for Cochlear Implant (CI) procedures is limited as the image resolution is coarse compared to the small size of the anatomical structure. Image analysis and processing combined with statistical modeling techniques can be applied to infer additional patient-specific anatomical knowledge, which has the potential to aid decisions regarding the surgical CI-insertion procedure, the choice of implant type suited for the recipient and the programming of the CI. This paper presents and evaluates a framework for building a statistical deformation model of the cochlea from high-resolution μCT datasets and the subsequent fitting of the model to patient pre-operative CT data.

1 INTRODUCTION

Cochlear Implants (CI) have become an established and successful way of treating severe hearing impairment and deafness. In the top performing cases recipients achieve near normal hearing, but a large variability in the auditory restoration outcomes remains a challenge to be dealt with [1]. Essentially there is a large group of patients who have the potential to obtain a better hearing. This outcome variability of the procedure affects the cost-benefit analysis of the device negatively, and it can be perceived as a factor of risk and uncertainty for the potential CI-recipients considering undergoing the surgery.

One of several ways of dealing with this problem is to achieve a better understanding of the cochlear shape of the individual CI-recipient, as this allows to customize and optimize the surgical planning and subsequent CI programming according to the specific patient [2].

The main source of information about patient anatomy comes from CT scans taken pre- and/or post-operatively. Current modern scanners can typically provide data with a voxel side length in the order of 0.15-0.40 mm. Compared to the relatively small size of the cochlea, which on average will fit inside a region of approximately $10 \times 8 \times 4$ mm [3, 4], the level of obtainable anatomical detail and information is very limited. The structures describing the intra-cochlear anatomy (for instances the *basilar membrane* and the *spiral lamina ossea*) are completely invisible, and even the gross complete geometry of the cochlear spiral can be difficult to observe fully (Figure 1).

There are different approaches for making a prediction of the true patient

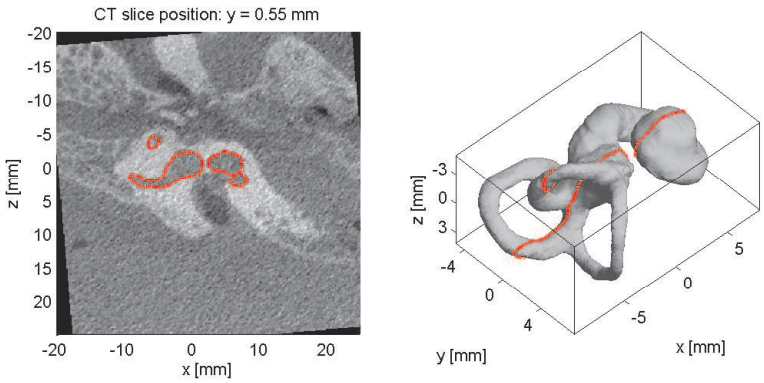


Figure 1: The information quality of the inner ear anatomy from CT pre-op data (0.15 mm isotropic voxels). (Left) An example CT slice cropped to an inner ear region. (Right) The extracted inner ear surface model from a manual segmentation.

anatomy based on a CT scan. Some studies try to estimate 'global' morphometrics of the cochlea by fitting various spiral models to the data [5, 6, 7]. The studies are characterized by manual measurements of a few observable parameters such as the cochlear height and the basal turn diameter, from which additional morphometrics are inferred using idealized mathematical spiral formulations with a low number of parameters. Depending on the intended application, these approaches can be considered too simple and generalized, and they are not able to take into account the true shape variability seen across the population.

A related approach is seen in Baker et al. [8]. A model-image based optimization procedure is used to estimate the 9 model parameters of a 3D spiral shell, that fits the CT image data in a best possible sense. The approach quickly suffers from the same limitations as the above mentioned strategies. It uses a very idealized model with too few parameters to account for the individual variability in shape. With such a model for instance all cross-sections of the cochlea are perfectly circular, contrary to what can be observed in real data, which puts a limit to the obtainable accuracy with such a strategy.

Some studies use more 'advanced' anatomical modeling from high-resolution datasets. Here, models can be based on a much larger number of parameters to describe not only the overall shape but also intra-cochlear anatomy to some extent. Most notable is the work of Noble et al. [9]. Using a selection of ex-vivo μCT scans they build a Statistical Shape Model (SSM) of the cochlea and follow the Active Shape Model (ASM) [10] approach with some modifications for estimating intra-cochlear anatomy from conventional CT. The statistical model provides a shape prior, which can guide a registration procedure between a patient CT scan and a CT-atlas enhanced with a model of the intra-cochlear anatomy. The shape prior acts as a regularization to improve and ensure that

the resulting output has an anatomical plausible cochlear shape.

The study presented in this paper follows a strategy along the same principles. We have an accurate high-resolution computational model of the cochlea extracted from μCT, which we wish to non-rigidly adapt to a patient CT scan, to make it specific for this patient. From a selection of similar μCT datasets we can extract a Statistical Deformation Model (SDM), describing statistically in which ways a μCT scan can deform to match with another cochlear shape. This SDM can act as a regularizing shape prior when adapting the computational model to a clinical CT scan to ensure anatomically plausible outputs.

The scope of this paper is to present and evaluate the overall framework for building a Statistical Deformation Model of the cochlea from high-resolution μCT data and the subsequent SDM-regularized fitting procedure to CT data.

Building, using and testing such models of anatomy and anatomical variability requires a lot of data and combines different image processing techniques (segmentation and registration) with statistical models. Several factors will impact the final results and the obtainable accuracy, precision and usability, which we will cover in this paper. The available datasets and their segmentations for this study are described in Section 2.1, detailing the involved imaging modalities, image resolutions and their usage. Further image processing of the data and the registration models used to build the SDM is detailed in Section 2.2. The concept of the SDM and the fitting procedure describing how we estimate detailed cochlear shape from clinical CT is covered in Section 2.3. The methodology is finalized with some details about the evaluation procedures in Section 2.4. Test results are presented in Section 3, followed by a discussion in Section 4.

2 MATERIALS AND METHODS

2.1 Data and segmentation

Anatomical reference model: The μCT scan that serves as our reference dataset and forms the basis for our 3D anatomical computational model is detailed in the work of Braun et al. [11]. In summary, it is a temporal bone specimen preserved by freezing with no additional staining or fixation. It was cut to contain only the region of the cochlear and most of the vestibule. Using a μCT system the specimen was scanned and the volume reconstructed in isotropic 5.9 μm voxels.

The segmentation of the sample contains the cochlear *scalae* as a combined structure and a cochlear partition, which serves as a good approximation of the basilar membrane (See Figure 2). Slight manual modifications were made to the segmentation using ITK-SNAP [12] compared to the version presented in [11]. Some spurious details were edited and smoothed away mostly in the region of the vestibule. These minor cosmetic changes were done to provide a smoother and less noisy surface model.

A triangulated surface mesh of the cochlear and the partition could be extracted

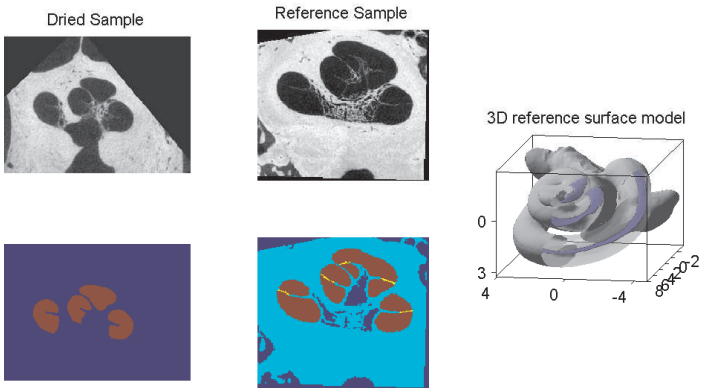


Figure 2: Illustration of the training and reference data and the extracted 3D surface model. Segmentation show the following labels - scalae (red), background (dark blue), bone (light blue) and cochlear partition (yellow). The latter two are only segmented for the reference dataset.

from the segmentations using Marching Cubes [13]. The meshes were refined with a surface reconstruction procedure [14] to obtain our 3D reference anatomical model (See Figure 2).

Training data: The μCT datasets that we use to build our Statistical Deformation Model are a series of dried temporal bone specimens. In total we have 18 bones without major deviations from the normal cochlear and vestibule anatomy. The samples were scanned with a μCT system and reconstructed with isotropic voxel sizes in the range of 16.3-24.5 μm. Since the bones are dried the images contain no soft-tissue structures.

The cochlear scalae was segmented as a single object semi-automatically using ITK-SNAP [12] and Seg3D [15](See Figure 2). Manual corrections were needed to correct for obvious errors and for having a smooth closing of the openings at the oval and round window.

Test data: For the purpose of validating our SDM-fitting procedure we have a collection of test datasets. It consists of 14 samples (originating from 8 different cadaverous head specimens), which were preserved with a thiel-solution. The samples were imaged both with a CT and μCT scanner, and reconstructed in respectively 150 and 7.6 μm isotropic voxels. However, the μCT data was downsampled to 24.5 μm to make it more manageable.

The intra-cochlear structures visible in the samples varies. Some samples retained fluid within the scalae, thereby eliminating contrast to other soft-tissues.

2.2 Image processing and registration

The registration procedure follows a common work-flow. The N training data samples are registered to the reference dataset in two steps - rigid initialization followed by deformable registration. The procedure is detailed in [16], and briefly recapped in the following to state the differences and introduce the notation.

In short, the datasets were downsampled to 24 μm voxel-sizes and then rigidly aligned to the reference to take out variability in translation and rotation between the samples. In order to ensure a good quality of the following non-rigid registration, especially for the cochlear turns and apical region, a heat distribution map through each cochlear was calculated. Instead of having a binary volume of the cochlear, each voxel of the foreground is assigned a heat value with the hottest point in the apex and coldest in vestibule. These heat distribution maps, H_i, provide a more suited way of the describing the similarity between two cochlea samples [16]. The non-rigid registration follows the framework and formulation of the `elastix` software library [17].

The registration of the moving dataset, I_M, towards the fixed image, I_F, is generally formulated as a (parametric) transformation, T_μ, where the vector μ containing the p-parameters of the transformation model are found as an optimization of the cost function, $\mathcal{C}$.

$$\hat{\mu} = \arg\min_{\mu} \mathcal{C}(T_\mu, I_F, I_M) \tag{1}$$

The transformation model used here is the cubic B-spline grid in a multi-level setting, where both the fixed and moving image is further smoothed with a Gaussian kernel. For each level of resolution the spacing between grid points and the width of the smoothing kernel follows a decreasing scheme, thus starting with a coarse registration that is gradually refined. The following scheme was used and chosen by experimentation:

Control point grid spacing (isotropic, voxels): $[60, 18, 6]$

Width of Guassian kernel (isotropic, voxels): $[6, 2, 1]$

The following cost-function was used for the registration:

$$\mathcal{C} = \alpha \cdot \mathcal{S}_{\text{Sim}}(\mu, H_F, H_M) + (1 - \alpha) \cdot \mathcal{P}_{\text{BE}}(\mu) \tag{2}$$

where α is a weight parameter in the interval [0,1], here chosen to 0.9 by means of experimentation. Sum of squared differences (SSD) is chosen as the metric for the similarity term, $\mathcal{S}_{\text{Sim}}$. H_F is the heat map of the reference dataset, and H_M in turn is the heat map of each of the 18 training data samples. The term $\mathcal{P}_{\text{BE}}$ is the energy bending regularization used to penalize strong changes and foldings in the transformation.

The optimization was solved using Adaptive Stochastic Gradient Descent [18]. The maximum number iterations was set to 1500. To reduce the computational burden of the optimization only a subset of voxels (2^{18} random coordinates) are sampled for the evaluation. The samples were only drawn close to the cochlear,

in order to focus the optimization further. This was achieved by applying a binary mask of the reference dataset, which was generated by dilating the scalae segmentation with a spherical kernel (radius = 20 voxels). The settings were fixed for all resolutions.

2.3 Statistical Deformation Model

Having obtained N=18 corresponding B-spline deformation fields, we generate the Statistical Deformation Model (SDM) by making a Principal Component Analysis (PCA) over the B-spline parameters. The process was done using the Statismo [19] software package. Additional details about the concept can be found in the work of Rueckert et al. [20].

2.3.1 Fitting procedure to clinical CT

Initialization: The SDM defines deformations of the reference volume, so the first step of the fitting is to move the target CT scan to that position and orientation. In practice, it is difficult to assume anything about the initial location of the cochlea in the CT images. An automatic initialization procedure could be developed, but for this study the initialization was done using a landmark-based rigid transformation. The following four landmarks were placed manually in each dataset; the center of the Round Window (RW) at the bony overhang, the center of the modiolus respectively in the basal and apical turn, and finally the basal turn inner wall across from the RW [21]. The landmark transform is followed by a rigid image registration between the reference dataset and the CT scan. This step can provide a slight correction of the orientation and position in order to prevent a bias or uncertainty from the user-based landmarking.

In practice both of the rigid transformations can be stored and the inverse applied in order to transform the fitted model (after next step) back to the position and orientation of the CT dataset.

Non-rigid image registration: Once the target CT and reference μCT are rigidly aligned the deformable fitting process can be started.

Using the Statismo-`elastix` integration allow us to follow a similar non-rigid image registration procedure (Equation 1). The major difference is that the transformation model changes to a SDM-regularized cubic B-spline model. In principle, it is a B-spline transform with the same grid definition as the final level of the previous described registration model. However, each grid point cannot vary freely anymore. Instead they deform in accordance with the learned covariance structure, which is described with a maximum of N-1 modes of variation of the PCA. The cost function can now be stated as:

$$\mathcal{C}_{\text{SDM}} = \mathcal{S}_{\text{Sim}}(\mu, I_F, I_{CT}) \tag{3}$$

where I_F is the reference μCT, and I_{CT} is a volume in CT resolution - for instance one of the test datasets. Mutual Information (MI) is used for the similarity term. Using Normalized Correlation Coefficient (NNC) could in many

cases be sufficient, but as some datasets contain different visible structures we use the more general applicable MI.

Other relevant settings: Smoothing applied to reference image (kernel size: 1.5 voxels), optimization running adaptive stochastic gradient descent with default settings. 2^{15} random samples per iteration for a maximum of 500 iterations.

2.4 Evaluation

Crossvalidation on training data: The training data contain a ground truth segmentation and are perfectly registered compared to the placement of the grid of the SDM. Downsampling the training data to CT resolutions ($\approx$0.2 mm isotropic voxels) provides a way of making a quantitative evaluation of the fitting procedure in a leave-one-out crossvalidation set-up. The SDM is built without the sample in question and the fitting procedure is done with all available modes of variations (N-2).

With the ground truth segmentation and the surface model it is possible to evaluate the accuracy and precision of the fit with following metrics: Dice Score of the binary segmentations, mean and maximum surface errors of the mesh models. As these scores only present a generic and overall accuracy score, we supplement the evaluation with scores that highlight the error at specific locations. We manually place a landmark at the apex and calculate the Euclidean distance between a ground truth and fitted apex. We further manually measure cochlear length, width and height according to the consensus definition of [22], allowing us to calculate some additional morphometric errors.

Test data: These datasets have no ground truth segmentations, and we therefore only evaluate using the morphometric errors stated above. The landmarks are placed in the μCT data, but the SDM is fitted to the CT data with 12 modes of variations available to optimize for. Note that μCT data versions are also registered to the reference dataset using the same initialization procedure as described in Section 2.3, in order to have a co-registration between the CT and μCT data.

3 RESULTS

The statistics of the resulting metrics are given in Table 1. A qualitative evaluation of the fitting accuracy are further illustrated in Figure 3.

Qualitatively the cochlear partition does not seem to fit very accurately on the test data. Note, however that the variability of this structure is not modelled in the SSM. It is simply a passive structure following the deformations of the fitting. This is merely a reflection of how the partition was seen in the reference dataset, but we have no guarantee for how precise and representative that actually it.

On average the morphometric (length, width and height) accuracy seems pretty good, but the observed standard deviation is quite high. However, as these measurements are made manually, there is a potentially significant uncertainty not

yet accounted for in those numbers. A more unbiased measurement procedure would be valuable.

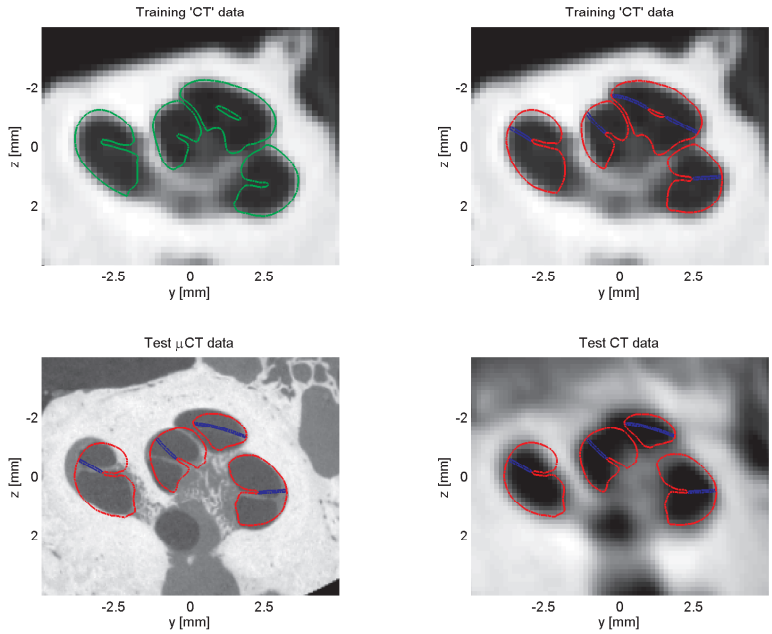


Figure 3: Ground truth (green) and fitted model (red and blue) shown on top CT/μCT data. The fitting to a CT training data sample (top row) is done in leave-one-out crossvalidation set-up. The fitting to a test sample (bottom row) is done only on the CT data, and further shown on top the μCT version for comparison.

4 DISCUSSION

There are several aspects to note and discuss regarding the results and the evaluation presented here. The true test of usability would of course be a measure of improvement in hearing restoration outcomes in CI-users. However, such a measure would be influenced by a lot of other confounding factors, than just the accuracy of cochlear shape prediction, which is the scope of this study. We have provided some quantitative scores and metrics for the prediction of the cochlear shape. However, their interpretation remains to be somewhat qualitative of nature as the absolute values are difficult to evaluate on their own. The Dice Score is in the high end, but does not immediately allow us to conclude anything about the accuracy of the predicted shape. The same is true for mean and maximum surface error, where the density of mesh vertices further influences the absolute values of the scores. The metrics are very common to report,

Table 1: Statistics of performance validation, reported as the mean +/- 1 std.

Type	**Crossvalidation**	**Test Data**
Dice Score	0.88 ± 0.02	-
Mean Err. [mm]	0.11 ± 0.01	-
Max Err. [mm]	0.58 ± 0.11	-
Apex Err. [mm]	0.8 ± 0.5	1.2 ± 0.4
Cochlear Length Err. [mm]	0.02 ± 0.2	-0.05 ± 0.4
Cochlear Width Err. [mm]	0.08 ± 0.2	-0.03 ± 0.3
Cochlear Height Err. [mm]	0.05 ± 0.2	0.2 ± 0.2
Spiral Length Err. [mm]	XX	-

but very limited and generic in their description.

In this regard the anatomically specific scores are more valuable. The apex error is shown to be larger than the maximum surface error, which demonstrates that the SDM and fitting procedure still has some issues and limitations with the apical region. A plausible explanation should likely be found in the amount of training data. The SDM it-self is built from a relatively low number of samples, which means that the modes of variation may not be able to sufficiently represent regional independence in the variability. However, μCT data is a scarce source to work with and the processing and segmentation a time consuming task. The limitations of the model should be seen in this light, and its' application can still be valuable. Providing a fair prediction of the cochlear anatomy can still be a lot more useful than assuming nothing about the anatomy, which is still the general practice today.

This is not the first approach for handling the challenge of cochlear shape prediction from conventional CT. One advantage of the SDM over the Active Shape Model approach [9, 10] is that it can handle intra-anatomical structures nicely. Since the intra-cochlear anatomy is not visible in CT there is no gradient information for guiding the fitting procedure. Further there is not enough data support to allow us to build a statistical model of the cochlear partition variability for instance. Describing everything as deformation fields with the SDM strategy allow us to quite simply just let the intra-cochlear structures follow the same deformations passively.

5 CONCLUSION

We have acquired a strong anatomical reference model describing the overall cochlea shape as well as intra-cochlear structures. From a series of μCT samples we have built a statistical shape model describing how the anatomical model can be deformed to fit into a pre-operational clinical CT scan. We have shown that we can estimate the patient specific anatomy with good accuracy especially

in the basal turn of the cochlea. The accuracy in the apical part is a bit lacking, suggesting that there is room for improvements. More training data would be valuable in order to improve on the low precision of the estimated anatomy.

Acknowledgments

The research leading to HEAR-EU results has received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement n°304857.

References

[1] B. S. Wilson and M. F. Dorman, "Cochlear implants: A remarkable past and a brilliant future," *HEARING RESEARCH*, vol. 242, no. 1-2, pp. 3–21, 2008.

[2] J. H. Noble, R. F. Labadie, R. Gifford, and B. Dawant, "Image-guidance Enables New Methods for Customizing Cochlear Implant Stimulation Strategies," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 21, no. 5, pp. 820–829, 2013.

[3] E. Avci, T. Nauwelaers, T. Lenarz, V. Hamacher, and A. Kral, "Variations in microanatomy of the human cochlea," *Journal of Comparative Neurology*, 2014.

[4] K.-J. Shin, J.-Y. Lee, J.-N. Kim, J.-Y. Yoo, C. Shin, W.-C. Song, and K.-S. Koh, "Quantitative analysis of the cochlea using three-dimensional reconstruction based on microcomputed tomographic images," *The Anatomical Record*, vol. 296, no. 7, pp. 1083–1088, 2013.

[5] D. R. Ketten, M. W. Skinner, G. Wang, M. W. Vannier, G. A. Gates, and J. G. Neely, "In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays," *The Annals of otology, rhinology & laryngology. Supplement*, vol. 175, p. 116, November 1998.

[6] B. Escudé, C. James, O. Deguine, N. Cochard, E. Eter, and B. Fraysse, "The size of the cochlea and predictions of insertion depth angles for cochlear implant electrodes," *Audiology and Neuro-Otology*, vol. 11 Suppl 1, p. 2733, 2006.

[7] E. Erixon and H. Rask-Andersen, "How to predict cochlear length before cochlear implantation surgery," *ACTA OTO-LARYNGOLOGICA*, vol. 133, no. 12, pp. 1258–1265, 2013.

[8] G. Baker and N. Barnes, "Model-image registration of parametric shape models: fitting a shell to the cochlea," *Insight Journal*, 2005.

[9] J. H. Noble, R. F. Labadie, O. Majdani, and B. M. Dawant, "Automatic Segmentation of Intracochlear Anatomy in Conventional CT," *Biomedical Engineering, IEEE Transactions on*, vol. 58, pp. 2625 –2632, sept. 2011.

[10] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[11] K. Braun, F. Boehnke, and T. Stark, "Three-dimensional representation of the human cochlea using micro-computed tomography data: Presenting an anatomical model for further numerical calculations," *ACTA OTO-LARYNGOLOGICA*, vol. 132, no. 6, pp. 603–613, 2012.

[12] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[13] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, pp. 163–169, Aug. 1987.

[14] R. Paulsen, J. Baerentzen, and R. Larsen, "Markov random field surface reconstruction," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, pp. 636–646, July 2010.

[15] CIBC, 2014. Seg3D: Volumetric Image Segmentation and Visualization. Scientific Computing and Imaging Institute (SCI), Download from: http://www.seg3d.org.

[16] H. M. Kjer, S. Vera, J. Fagertun, D. Gil, M. Á. González-Ballester, and R. Paulsen, "Image registration of cochlear μct data using heat distribution similarity," in *Image Analysis*, pp. 234–245, Springer, 2015.

[17] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 196 – 205, January 2010.

[18] S. Klein, J. P. W. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *INTERNATIONAL JOURNAL OF COMPUTER VISION*, vol. 81, no. 3, pp. 227–239, 2009.

[19] M. Lüthi, R. Blanc, T. Albrecht, T. Gass, O. Goksel, P. Büchler, M. Kistler, H. Bousleiman, M. Reyes, P. Cattin, and T. Vetter, "Statismo - a framework for pca based statistical models,"

[20] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 22, no. 8, pp. 1014–1025, 2003.

[21] W. Wimmer, F. Venail, T. Williamson, M. Akkari, N. Gerber, S. Weber, M. Caversaccio, A. Uziel, and B. Bell, "Semiautomatic cochleostomy target and insertion trajectory planning for minimally invasive cochlear implantation," *BIOMED RESEARCH INTERNATIONAL*, vol. 2014, pp. –, 2014.

[22] B. M. Verbist, M. W. Skinner, L. T. Cohen, P. A. Leake, C. James, C. Boëx, T. A. Holden, C. C. Finley, P. S. Roland, J. T. Roland, M. Haller, J. F. Patrick, C. N. Jolly, M. A. Faltys, J. J. Briaire, and J. H. Frijns, "Consensus panel on a cochlear coordinate system applicable in histologic, physiologic, and radiologic studies of the human cochlea," *Otology and Neurotology*, vol. 31, no. 5, pp. 722–730, 2010.

Appendix J

Cochlear Implant Electrode Localization in Post-Operative CT Using a Spherical Measure

Cochlear Implant Electrode Localization in Post-Operative CT Using a Spherical Measure

Benjamin Braithwaite, Hans Martin Kjer, Jens Fagertun, and Rasmus Paulsen

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Kgs. Lyngby, Denmark

Abstract. We present a method for automatically locating and determining the ordering of electrode contacts on implanted electrode arrays from post-operative CT images. Our method applies a specialized filter to the images based on a threshold and spherical measure, and selects contact positions at local maxima in the filtered image. Two datasets of 13 temporal bone specimens scanned in CBCT are used to validate the method, which successfully locates and orders each electrode contact in every image.

Keywords: Cochlear Implant, Electrode array, Spherical measure.

1 Introduction

Cochlear Implants (CIs) are used to restore hearing in patients who are profoundly deaf or severely hard of hearing. They work by directly stimulating the hearing nerve with an electrode array inserted into the cochlea. The positions of the intra-cochlear electrode contacts determine the intensity and frequency of sound sensed by the patient. Locating the electrode contacts in post-operative CT has thus proven useful for e.g. studying patient hearing outcomes [6] and automatically determining patient-specific configurations of electrodes [10]. Due to the time-consuming and error-prone nature of manually annotating the CT images, it is useful to automate the process.

Figure 1 shows an example of a temporal bone specimen with an implanted electrode array scanned in CBCT and μCT, where the metallic part has been isolated by a threshold. Both the electrode contacts and spiral wires in between are visible, but the latter diminishes or disappears further along the array. Note that scanning metal in CBCT creates artifacts, which causes the contacts appear larger, rounder, and more uneven than they are in reality. Artifacts appear to a much lesser extent in μCT,

(a) CBCT, front (b) μCT, front (c) CBCT, side (d) μCT, side

Fig. 1: Appearance of an electrode array in CBCT (0.15 mm voxel width) and μCT (0.018 mm voxel width). The metal is isolated by a threshold, see Section 3.

but cannot be used for scanning live patients. A MED-EL Standard electrode array has been used in this example, which has a 2.4 mm spacing between contacts. Other arrays have a much smaller contact spacing which changes the nature of the problem. We distinguish between distantly-spaced contacts, where there is sufficient contrast to identify individual contacts, and closely-spaced contacts, where neighbouring contacts are indistinguishable from each other in the image. The category of spacing depends both on the electrode array and the scanning process.

Noble et al. [3] from Vanderbilt University have developed a method for locating closely-spaced electrode arrays based on Gradient Vector Flow snakes. They validate the method on 7 temporal bone specimens with inserted Cochlear Contour Advance arrays, scanned in CT with a 0.3 mm isotropic voxel width. By comparing their results to a manual annotated ground truth they achieved a 0.195 mm mean curve error. Through a further refinement of the method [9] validated on 15 live patients scanned in CT with 0.4 mm isotropic voxel width, they achieved a 0.09 mm mean curve error and 0.21 mm mean electrode localization error. The snake-fitting approach used in these methods are well-suited for closely-spaced arrays, but may break down when no metal is detected between neighbouring electrodes.

In this work we instead consider the problem of locating distantly-spaced electrode arrays. We propose a method where contacts are individually located and subsequently ordered, using the knowledge of the type of electrode array present when needed. In the case of contact pairs at the same distance along the array, as found in many MED-EL electrode arrays, we only detect a single point between the contacts. We assume a $\sim$1 cm^3 region around the cochlea is given. The region can be located manually or found through registration with a known volume. [9] To validate our method we use a dataset of temporal bone specimens scanned in CBCT and μCT. Since the CBCT images are much more prone to metallic artifacts than μCT, we use the latter as ground truth for a more accurate estimation of error.

Dataset	Specimen ID	Electrode type	Intra-cochlear contacts*
ARTORG [8]	1L	Standard	12 of 12
	2L	Standard	12 of 12
	2R	Standard	12 of 12
	3L	FLEX^{28}	12 of 12
	3R	Standard	10 of 12
	4L	FLEX^{28}	12 of 12
	4R	FLEX^{28}	12 of 12
MED-EL [11]	1	Standard	9 of 12
	2	36 channel	23 of 36
	3	FLEX^{28}	11 of 12
	4	FLEX^{28}	11 of 12
	5	Standard	9 of 12
	6	Standard	10 of 12
*Contact pairs are counted as single contacts.			

Table 1: The datasets used for evaluation. The ARTORG dataset is scanned in CBCT (0.15 mm isotropic voxel width) and μCT (0.018 mm isotropic voxel width), while the MED-EL dataset is scanned in CBCT (0.10 mm isotropic voxel width).

2 Data

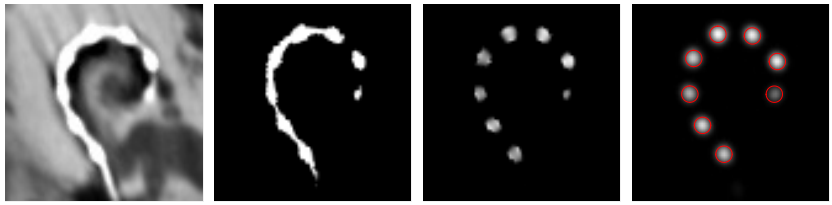
In this study we use two datasets containing CT images of temporal bone specimens with surgically inserted MED-EL electrode arrays. The ARTORG dataset [8] consists of 7 specimens scanned in CBCT (0.15 mm isotropic voxel width) and μCT (0.018 mm isotropic voxel width). The MED-EL dataset [11] consists of 6 specimens scanned in CBCT (0.10 mm isotropic voxel width). In both datasets the electrode arrays were inserted through the round window. The electrode types and insertion depths for each image are shown in Table 1.

3 Methods

Our proposed method uses a specialized filter that, when applied to the CT image, produces the greatest response at centres of electrode contacts. We note three distinctive properties of these locations:

1. The intensity of metal in CT is much higher than the surrounding tissue.
2. The contacts resemble a sphere.
3. The contacts have a certain size.

We take advantage of these properties by successively applying three filters to the volumetric image $I(\mathbf{x})$[1]: a thresholding filter $T(\mathbf{x})$, a spherical filter $S(\mathbf{x})$, and a Gaussian filter $G(\mathbf{x})$. Given an electrode array model with N contacts, we then search the filtered image for the N greatest local maxima, and use another method to determine the contact ordering. An overview of the process is shown in Fig. 2, and the individual steps are explained in the following.



(a) (b) (c) (d)

Fig. 2: Overview of the successive filtering scheme shown for a single slice of the 3D volume. Starting with a $\sim$1 cm^3 region around the cochlea (a), three filters are applied successively: thresholding (b), spherical measure (c), and Gaussian (d). Local maxima in the filtered image are selected as contact positions.

Thresholding: Metal has a much higher intensity (HU) than the surrounding tissue, and can thus be reliably isolated by a threshold. We use a binary thresholding filter:

$$T(\mathbf{x};t) = \begin{cases} 1, & \text{if } I(\mathbf{x}) \geq t \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

[1] $\mathbf{x} = [x\ y\ z]^\top$ is the coordinate vector for a voxel

The value of t was chosen by visual inspection, and we were able to use the same threshold across all images of $t = 3064$. This was possible since our images had the same range of intensities. If this were not the case, there are methods for selecting a dynamic threshold to take this into account [9].

Spherical measure: Having reduced the image to its metallic voxels, we need to separate the contact positions from wiring. The contact positions appear more spherical or blob-like than the tubular wiring, as is apparent in Fig. 1. We found Westin's spherical measure [7] to be effective at differentiating the two. Our spherical filter is defined as:

$$S(\mathbf{x}) = \begin{cases} \frac{3\lambda_3}{\lambda_1+\lambda_2+\lambda_3}, & \text{if } \lambda_{1,2,3} < 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$ are eigenvalues of the Hessian matrix of $G(\mathbf{x};\sigma) * I(\mathbf{x})$. $G(\mathbf{x};\sigma)$ is a 3D Gaussian kernel with standard deviation σ. The eigenvalues $\lambda_{1,2,3}$ represent the change in intensity along three perpendicular directions at scale σ. A point amidst a bright sphere will cause all eigenvalues to be negative and similar to each other; this is when the filter returns the highest values.

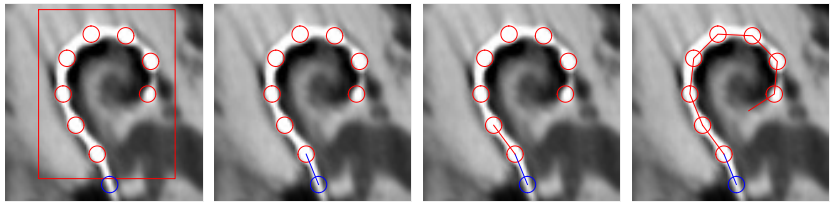
The value of σ should roughly correspond to the size of the contact positions in CT. Note that this is not the same as the physical size of the contacts due to contributions from wiring, the case of contact pairs, and the fact that metal appears larger in CT because of artifacts. For the Standard and FLEX^{28} electrode types we used $\sigma = 0.5$ mm, and for the thinner 36 channel model we used $\sigma = 0.2$ mm.

Gaussian: The response from the spherical filter appears noisy and cannot be used directly to find local maxima. To find the centres of the contact regions, we apply the Gaussian filter $G(\mathbf{x};\sigma/2)$, which smooths the response image based on the previously used scale parameter. This causes only the centres of each region to be local maxima.

Contact detection and ordering: Maxima are found in the filtered image by comparing each voxel to its 26-voxel neighbourhood, and the N maxima with the greatest filter value are selected as contact positions. They are then corrected to sub-voxel positions by computing the maxima of the voxel's Taylor expansion up to quadratic terms.

The resulting points need to be ordered according to their position on the implant. This process is illustrated in Fig. 3. First we find a point on the lead connecting the electrode array to the internal implant. This is done by defining a bounding box around the electrode contacts, and searching the area just outside this in the thresholded image. The mean position of the metal voxels is located as a point on the lead. From here, a breadth-first flood fill approach is used to follow the wiring until a contact position is reached; this point is p_1.

To order the remaining contacts we refrain from using the thresholded image, as gaps occur between contacts with little wiring in between; see Fig. 1. Instead we iteratively predict the next contact position based on a linear model. Given p_{k-1} and p_k (p_0 is the point on the lead), the predicted point is $\hat{p} = p_k + (p_k - p_{k-1})s$, where s is the contact spacing of the given electrode type. The nearest remaining contact to $\hat{p}$ is selected as the next contact p_{k+1}. Repeating this process for each point results in an ordered electrode array $p_{1..N}$.



(a) (b) (c) (d)

Fig. 3: Overview of the ordering scheme shown for the same slice as in Fig. 2. First a point on the connecting lead is found outside a bounding box containing the electrode contacts (a), then the first contact is found by a breadth-first flood fill approach (b), and finally the remaining contacts are predicted based on a linear model (c) until all the contacts are ordered (d).

Cochlea model: We use a Statistical Deformation Model (SDM) [1] to estimate the shape of the cochlea. The model was built from 17 segmented μCT images of inner ear specimens without implanted arrays, and is fitted to an image by manual landmark alignment followed by mutual information based registration. Further details can be found in the referenced paper. The model is not directly applicable to our post-op CT images due to the implants, but we found that adding a simple pre-processing step was sufficient to get a reasonable result: The metallic intensities above the threshold t are changed to a common intra-cochlear value $k = 0$.

4 Results

We present three types of results: an evaluation of the method's accuracy, a visualization using the cochlea model, and a stress test of the method with regards to the voxel size.

As noted in the introduction, artifacts in CBCT cause the appearance of the electrode contacts to be misleading whereas μCT gives a much clearer picture. We therefore base our ground truth on the μCT images from the ARTORG dataset. The ground truth was constructed by having two individuals manually annotate the contact positions, and computing the mean position for each electrode contact. Note that the annotation process is relatively straightforward due to the high amount of detail offered by μCT, recall figure Fig. 1.

The method is used to estimate the electrode contact positions in the corresponding CBCT images, which are similar to what is available in a clinical setting. In order to compare the estimated positions with the ground truth, we compute a rigid registration between the two image types. Additionally we run the method on the μCT images downsampled to the size of the CBCT images to understand the difference between the two. The results of the comparison are shown visually in Fig. 4, and the mean errors are reported in Table 2. For CBCT images the method has a mean error of 0.148 ± 0.075 mm corresponding to 0.98 ± 0.50 voxel widths. For μCT images the mean error is 0.049 ± 0.023 mm corresponding to 2.74 ± 1.28 voxel widths.

Two examples of the estimated electrode positions shown together with the cochlea surface model are shown in Fig. 5.

Since our approach is only feasible when the spacing between contacts is sufficiently large compared to the image resolution, we perform an additional experiment to

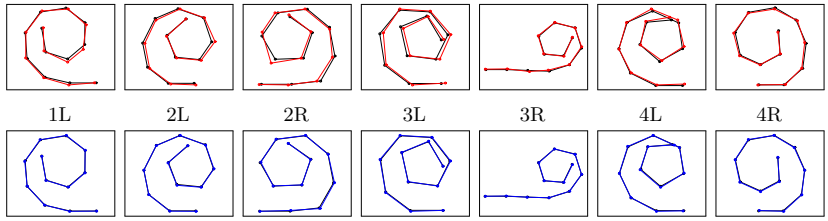
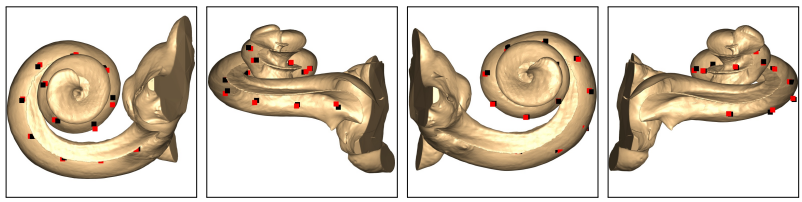


Fig. 4: Electrode localization results on the ARTORG dataset. The manually annotated positions from μCT (black) are shown with the automatically estimated contact positions from CBCT (red, above) and from μCT (blue, below). The positions have been projected onto a 2D plane.

Specimen ID	Mean error in CBCT (mm)	Mean error in downsampled μCT (mm)
1L	0.212	0.044
2L	0.126	0.042
2R	0.201	0.053
3L	0.170	0.055
3R	0.080	0.047
4L	0.158	0.051
4R	0.087	0.052
Overall	0.148 $\pm$ 0.075	0.049 $\pm$ 0.023

Table 2: Quantitative electrode localization results on the ARTORG dataset.

determine the point at which the method breaks down. Each image is downsampled to an increasingly smaller size (by increasing the voxel width in steps of 0.01 mm) until the method fails to locate the electrodes correctly. For each electrode type we report the largest voxel width in which every image of that electrode type could be successfully estimated. The results are shown in Table 3.



(a) 1L, front (b) 1L, side (c) 4R, front (d) 4R, side

Fig. 5: Manually annotated electrode positions (black) and automatically estimated positions (red) are shown together with the estimated cochlea shape.

Electrode type	Contact spacing (mm)	Min. voxel width (mm)
Standard	2.4	0.36
FLEX^{28}	2.1	0.36
36 channel	0.8	0.17

Table 3: Stress test results of the method with regards to voxel size. For each electrode type we report the largest voxel width in which every image of that electrode type could be successfully estimated.

5 Discussion and Conclusion

We have presented a method for automatically locating electrode arrays in post-operative CT images. The method takes as input a $\sim$1 cm^3 region around the cochlea, and requires three parameters: a threshold t, which should be set based on the range of intensities in the images, and a scale σ and spacing s, which should both be set based on the type of implanted electrode array.

Two datasets of 13 temporal bone specimens in total were used to validate the method. The method successfully located and determined the ordering of every electrode contact in every image. The mean error in terms of distance was evaluated based on a ground truth defined in μCT images. For CBCT images the overall mean error was 0.148 ± 0.075 mm, and when run on the μCT images of the same size the error is reduced to 0.049 ± 0.023 mm. In reality we expect the error to be somewhere in between, as the imperfect registration between CBCT and μCT causes the first error to be overstated, while the second error uses images with less artifacts. The results cannot be directly compared to the work of Zhao et al. [9] which achieved a 0.21 mm mean electrode localization error, as they used lower quality images with closely-spaced electrode arrays, significantly increasing the difficulty of the task. However we can use their error figure to justify that our method is accurate enough to use in practice.

Acknowledgements. This research is part of the HEAR-EU project, which has received funding from the European Union Seventh Frame Programme (FP7/2007-2013) under grant agreement n°304857.

References

[1] Kjer, H.M., Vera, S., Fagertun, J., Gonzlez-Ballester, M.A., Paulsen, R.R.: Predicting Detailed Inner Ear Anatomy from Clinical Pre-Op CT. In: Proceedings of CARS. vol. 10, pp. 98–99. Springer Berlin Heidelberg (2015)

[2] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality Image Registration by Maximization of Mutual Information. Medical Imaging, IEEE Transactions on 16(2), 187–198 (1997)

[3] Noble, J.H., Schuman, T.A., Wright, C.G., Labadie, R.F., Dawant, B.M.: Automatic Identification of Cochlear Implant Electrode Arrays for Post-Operative Assessment. Proc. SPIE 7962, 796217 (2011)

[4] Reda, F.A., Dawant, B.M., McRackan, T.R., Labadie, R.F., Noble, J.H.: Automatic segmentation of intra-cochlear anatomy in post-implantation CT. Proc. SPIE 8671, 86710I (2013)

[5] Skinner, M.W., Holden, T.A., Whiting, B.R., Voie, A.H., Brunsden, B., Neely, J.G., Saxon, E.A., Hullar, T.E., Finley, C.C.: In Vivo Estimates of the Position of Advanced Bionics Electrode Arrays in the Human Cochlea. Annals of Otology, Rhinology & Laryngology 116(4), 2–24 (2007)

[6] Wanna, G.B., Noble, J.H., Carlson, M.L., Gifford, R.H., Dietrich, M.S., Haynes, D.S., Dawant, B.M., Labadie, R.F.: Impact of Electrode Design and Surgical Approach on Scalar Location and Cochlear Implant Outcomes. The Laryngoscope 124(S6), S1–S7 (2014)

[7] Westin, C.F., Peled, S., Gudbjartsson, H., Kikinis, R., Jolesz, F.A.: Geometrical Diffusion Measures for MRI from Tensor Basis Analysis. In: Proceedings of ISMRM. vol. 97, p. 1742 (1997)

[8] Wimmer, W., Bell, B., Huth, M.E., Weisstanner, C., Gerber, N., Kompis, M., Weber, S., Caversaccio, M.: Cone Beam and Micro-Computed Tomography Validation of Manual Array Insertion for Minimally Invasive Cochlear Implantation. Audiology and Neurotology 19(1), 22–30 (2014)

[9] Zhao, Y., Dawant, B.M., Labadie, R.F., Noble, J.H.: Automatic Localization of Cochlear Implant Electrodes in CT. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) Medical Image Computing and Computer-Assisted Intervention MICCAI 2014, Lecture Notes in Computer Science, vol. 8673, pp. 331–338. Springer International Publishing (2014)

[10] Zhao, Y., Dawant, B.M., Noble, J.H.: Automatic electrode configuration selection for image-guided cochlear implant programming . vol. 9415, pp. 94150K–94150K–8 (2015)

[11] Zou, J., Lähelmä, J., Koivisto, J., Dhanasingh, A., Jolly, C., Aarnisalo, A., Wolff, J., Pyykkö, I.: Imaging cochlear implantation with round window insertion in human temporal bones and cochlear morphological variation using high-resolution cone beam CT. Acta Oto-laryngologica 135(5), 466–472 (2015), pMID: 25675836

Appendix K

Current Electrode Designs

Cochlear™ Implant Electrode Comparison

Current Electrodes Offered by All Manufacturers

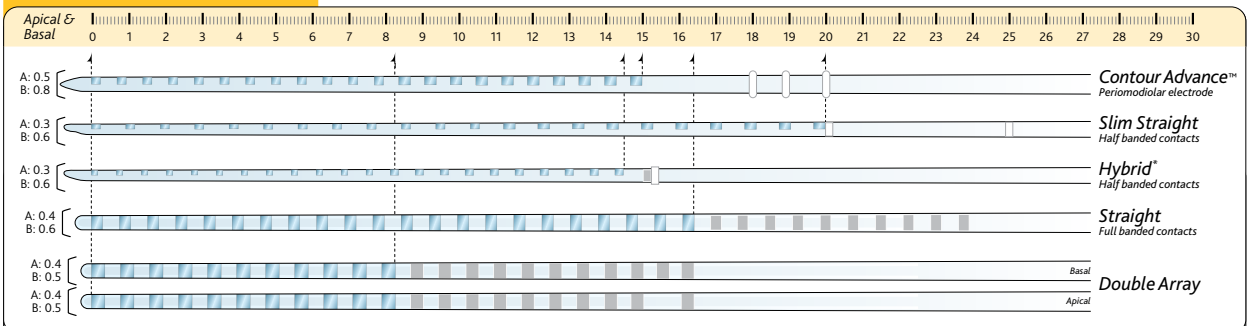
Reference Guide

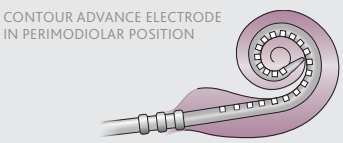
Uniquely Designed for Preservation, Performance and Preference

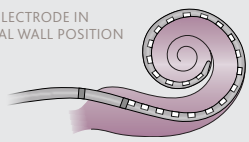
Cochlear provides a choice of electrodes with 22 active contacts for maximum stimulation and performance.

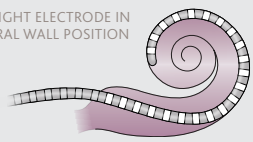
Cochlear™ Ltd. — Electrodes

Linear Electrode Depth (mm) — *Active Contacts* — *Stiffening Rings*



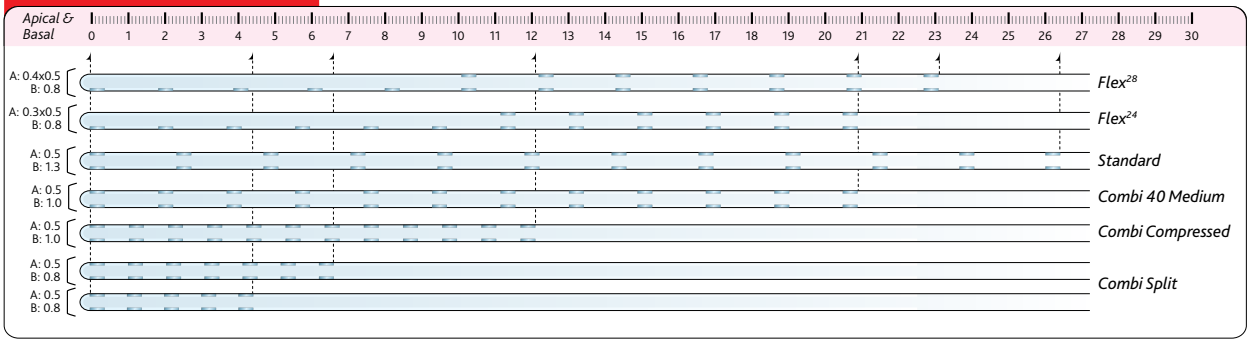






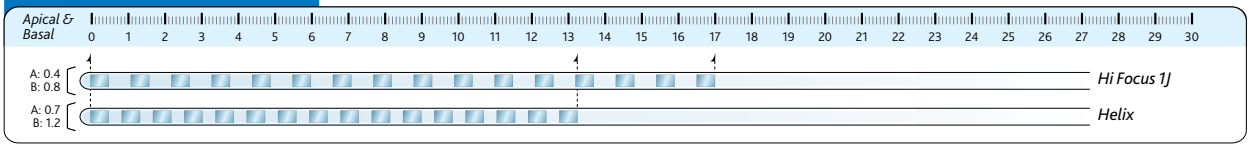
Med-El — Electrodes

Linear Electrode Depth (mm) – Note: 12 stimulation channels on all electrodes



Advanced Bionics® — Electrodes

Linear Electrode Depth (mm)



Measurements according to manufactures specifications – MedEl and Advanced Bionics specification collateral.
*CAUTION - Investigational device. Limited by United States law to investigational use.

www.CochlearAmericas.com

Cochlear Americas
13059 East Peakview Avenue
Centennial, CO 80111 USA

Tel: 1 303 790 9010
Fax: 1 303 792 9025
Toll Free: 1 800 523 5798

Please Recycle FUN1142 ISS4 JUL12

Hear now. And always

Cochlear™

Appendix L

Overview of Author Contributions

As this thesis is part of a larger research project I certainly cannot take full credit for all of the work presented. I include here a table to clarify what my contributions have been.

Table L.1: Author Contribution Table

Chapter	Section	Contribution Description
2	-	General background knowledge - no contributions
3	-	General background knowledge - no contributions
4	4.2 - 4.3	Data acquisition, sample preparation and scanning was done by University of Bern and Scanco Medical
	4.4	Data segmentation - approximately an equal effort between Sergio Vera (Alma), Livia Barazzetti (UB) and me
	4.5	Extraction of surface models using open-source software. No contributions, practical work and parameter tuning - me
	4.6.1	Initial registration procedure. Developed and implemented - me
	4.6.2	Morphometrics and torus fitting. Measured, developed and implemented - me
	4.6.3	Methodology for extraction of skeleton landmarks - mainly by Jens Fagertun (DTU)
	4.6.4	Cochlear heat distributions - Sergio Vera (Alma)
5	-	Image registration using `elastix`. Running experiments (set-up, parameter tuning), validation and data processing - me
6	-	SSM using Statismo and Statismo-`elastix`. Set-up and model building - joint effort between Jens Fagertun and me. Fitting procedure (implementation and experimentation) - me.
6	-	Multi-region PDM - Jordi Romero (UPF)
7	7.1	Patient specific anatomy estimation - me
	7.2	Procedure for electrode localization - Ben Braithwaite (DTU)
	7.3	Eletrode Insertion Simulation - Sergio Vera (Alma)
	7.4	Computational Models - Nerea Mangado, Mario Ceresa and others (UPF)

Bibliography

[11-13] NIH Publication No. 11-4798. Cochlear implants, 2013. [Online; accessed 20-July-2015].

[AML+84] David J. Apple, Nick Mamalis, Katherine Loftfield, Joseph M. Googe, Linda C. Novak, Dolores Kavka-Van Norman, Steven E. Brady, and Randall J. Olson. Complications of intraocular lenses. a historical and histopathological review. *Survey of Ophthalmology*, 29(1):1 – 54, 1984.

[ANL+14] Ersin Avci, Tim Nauwelaers, Thomas Lenarz, Volkmar Hamacher, and Andrej Kral. Variations in microanatomy of the human cochlea. *Journal of Comparative Neurology*, 2014.

[Azu97] Ronald T. Azuma. A survey of augmented reality. *PRESENCE-TELEOPERATORS AND VIRTUAL ENVIRONMENTS*, 6(4):355–385, 1997.

[BBS12] Katharina Braun, Frank Boehnke, and Thomas Stark. Three-dimensional representation of the human cochlea using micro-computed tomography data: Presenting an anatomical model for further numerical calculations. *ACTA OTO-LARYNGOLOGICA*, 132(6):603–613, 2012.

[BSL+11] Martin Baiker, Marius Staring, Clemens W. G. M. Löwik, Johan H. C. Reiber, and Boudewijn P. F. Lelieveldt. Automated registration of whole-body follow-up microCT data of mice. *Medical Image Computing and Computer Assisted Intervention*, 14(Pt 2):516–523, 2011.

[BvdHL+13] Floris F. Berendsen, Uulke A. van der Heide, Thomas R. Langerak, Alexis N. T. J. Kotte, and Josien P. W. Pluim. Free-form image registration regularized by a statistical shape model: application to organ segmentation in cervical MR. *Computer Vision and Image Understanding*, 117(9):1119–1127, 2013.

[CIB14] CIBC, 2014. Seg3D: Volumetric Image Segmentation and Visualization. Scientific Computing and Imaging Institute (SCI), Download from: http://www.seg3d.org.

[CTCG95] Timothy F. Cootes, Christopher J. Taylor, David H. Cooper, and Jim Graham. Active Shape Models - Their Training and Application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[CV15] Francis S. Collins and Harold Varmus. A new initiative on precision medicine. *New England Journal of Medicine*, 372(9):793–795, 2015. PMID: 25635347.

[Dav02] Rhodri H. Davies. *Learning shape: optimal models for analysing natural variability*. PhD thesis, University of Manchester, 2002.

[Dic45] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):pp. 297–302, 1945.

[EHWRA09] Elsa Erixon, Herman Högstorp, Karin Wadin, and Helge Rask-Andersen. Variational Anatomy of the Human Cochlea: Implications for Cochlear Implantation. *Otology and Neurotology*, 30(1):14–22, 2009.

[EJD+06] Bernard Escudé, Chris James, Olivier Deguine, Nadine Cochard, Elias Eter, and Bernard Fraysse. The size of the cochlea and predictions of insertion depth angles for cochlear implant electrodes. *Audiology and Neuro-Otology*, 11 Suppl 1:27—33, 2006.

[ERA13] Elsa Erixon and Helge Rask-Andersen. How to predict cochlear length before cochlear implantation surgery. *ACTA OTO-LARYNGOLOGICA*, 133(12):1258–1265, 2013.

[FGT14] Ana Ferreira, Fernanda Gentil, and João Manuel R.S. Tavares. Segmentation algorithms for ear image data towards biomechanical studies. *COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING*, 17(8):888–904, 2014.

[FRSN02] Alejandro F. Frangi, Daniel Rueckert, Julia A. Schnabel, and Wiro J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac

modeling. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 21(9):1151–1166, 2002.

[Gol12] David Goldman. Google unveils 'project glass' virtual-reality glasses, 2012. [Online; accessed 20-July-2015].

[Gre90] D. D. Greenwood. A cochlear frequency-position function for several species–29 years later. *Journal of the Acoustical Society of America*, 87(6):2592–2605, 1990.

[HBHH01] DLG Hill, PG Batchelor, M. Holden, and DJ Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):R1–R45, 2001.

[HM09] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *MEDICAL IMAGE ANALYSIS*, 13(4):543–563, 2009.

[HTF01] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 1st edition, 2001.

[JS04] Nathan Jeffery and Fred Spoor. Prenatal growth and development of the modern human labyrinth. *Journal of Anatomy*, 204(2):71–92, 2004.

[KPSV09] Stefan Klein, Josien P. W. Pluim, Marius Staring, and Max A. Viergever. Adaptive stochastic gradient descent optimisation for image registration. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 81(3):227–239, 2009.

[KSJ+12] Eric R. Kandel, James H. Schwartz, Thomas M. Jessell, Steven A. Siegelbaum, and A. J. Hudspeth. *Principles of Neural Science*. McGraw-Hill Medical, 5th edition, 2012.

[KSM+10] Stefan Klein, Marius Staring, Keelin Murphy, Max A. Viergever, and Josien P.W. Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196 – 205, January 2010.

[KSP07] Stefan Klein, Marius Staring, and Josien PW Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *Image Processing, IEEE Transactions on*, 16(12):2879–2890, 2007.

[KSW+98] Darlene R. Ketten, Margaret W. Skinner, Ge Wang, Michael W. Vannier, George A. Gates, and J. Gail Neely. In vivo measures of cochlear length and insertion depth of nucleus cochlear implant

electrode arrays. *The Annals of Otology, Rhinology & Laryngology. Supplement*, 175:1—16, November 1998.

[LA99] Hava Lester and Simon R Arridge. A survey of hierarchical non-linear medical image registration. *Pattern recognition*, 32(1):129–149, 1999.

[LBA+12] Marcel Lüthi, Rémi Blanc, Thomas Albrecht, Tobias Gass, Orcun Goksel, Philippe Büchler, Michael Kistler, Habib Bousleiman, Mauricio Reyes, Philippe C. Cattin, and Thomas Vetter. Statismo - a framework for pca based statistical models. *The Insight Journal*, 1:1–18, 07 2012.

[LC87] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, August 1987.

[LMM98] Gabor Lukács, Ralph Martin, and Dave Marshall. Faithful least-squares fitting of spheres, cylinders, cones and tori for reliable segmentation. In *Computer Vision—ECCV'98*, pages 671–686. Springer, 1998.

[Loi99] Philipos C. Loizou. Introduction to Cochlear Implants. *Engineering in Medicine and Biology Magazine, IEEE*, 18(1):32–42, 1999.

[LSK+13] Ju-Young Lee, Kang-Jae Shin, Jeong-Nam Kim, Ja-Young Yoo, Wu-Chul Song, and Ki-Seok Koh. A morphometric study of the semicircular canals using micro-ct images in three-dimensional reconstruction. *The Anatomical Record*, 296(5):834–839, 2013.

[LW10] John I. Lane and Robert J. Witte. *Temporal Bone: An Imaging Atlas*. Springer Science & Business Media, 2010.

[MCV+97] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, 16(2):187–198, 1997.

[Mil11] Mara Mills. Hearing aids and the history of electronics miniaturization. *Annals of the History of Computing, IEEE*, 33(2):24–45, Feb 2011.

[ML12] Daniel M. Merfeld and Richard F. Lewis. Replacing semicircular canal function with a vestibular implant. *CURRENT OPINION IN OTOLARYNGOLOGY and HEAD AND NECK SURGERY*, 20(5):386–392, 2012.

[MV98] J. B. Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[NEAT14] Guangjian Ni, Stephen J Elliott, Mohammad Ayat, and Paul D Teal. Modelling cochlear mechanics. *BioMed research international*, 2014, 2014.

[New15] BBC News. Google glass sales halted but firm says kit is not dead, 2015. [Online; accessed 20-July-2015].

[NLGD13] Jack H. Noble, Robert F. Labadie, René Gifford, and Benoit Dawant. Image-guidance Enables New Methods for Customizing Cochlear Implant Stimulation Strategies. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 21(5):820–829, 2013.

[NLMD11] Jack H. Noble, Robert F. Labadie, Omid Majdani, and Benoit M. Dawant. Automatic Segmentation of Intracochlear Anatomy in Conventional CT. *Biomedical Engineering, IEEE Transactions on*, 58(9):2625 –2632, sept. 2011.

[OdC12] Jong Min Ong and Lyndon da Cruz. The bionic eye: a review. *Clinical & Experimental Ophthalmology*, 40(1):6–17, 2012.

[OS88] Stanley Osher and James A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

[PBL10] R.R. Paulsen, J.A. Baerentzen, and R. Larsen. Markov random field surface reconstruction. *Visualization and Computer Graphics, IEEE Transactions on*, 16(4):636–646, July 2010.

[PPM10] Nathaniel R. Peterson, David B. Pisoni, and Richard T. Miyamoto. Cochlear Implants and Spoken Language Processing Abilities: Review and Assessment of the Literature. *Restorative Neurology and Neuroscience*, 28(2):237–250, 2010.

[PZK+11] Anton A. Poznyakovskiy, Thomas Zahnert, Yannis Kalaidzidis, Nikoloz Lazurashvili, Rolf Schmidt, Hans-Juergen Hardtke, Björn Fischer, and Yury M. Yarin. A segmentation method to obtain a complete geometry model of the hearing organ. *HEARING RESEARCH*, 282(1-2):25–34, 2011.

[RAD03] Gustavo K. Rohde, Akram Aldroubi, and Benoit M. Dawant. The adaptive bases algorithm for intensity-based nonrigid image registration. *Medical Imaging, IEEE Transactions on*, 22(11):1470–1479, 2003.

[RALE+12] Helge Rask-Andersen, Wei Liu, Elsa Erixon, Anders Kinnefors, Kristian Pfaller, Annelies Schrott-Fischer, and Rudolf Glueckert. Human Cochlea: Anatomical Characteristics and their Relevance for Cochlear Implantation. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, 295(11):1791–1811, 2012.

[RFS01] Daniel Rueckert, Alejandro F. Frangi, and Julia A. Schnabel. Automatic construction of 3d statistical deformation models using non-rigid registration. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, MICCAI '01, pages 77–84. Springer-Verlag, 2001.

[RMBJ03] T. Rohlfing, CR Maurer, DA Bluemke, and MA Jacobs. Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*, 22(6):730–741, 2003.

[RML+14] Fitsum A. Reda, Theodore R. McRackan, Robert F. Labadie, Benoit M. Dawant, and Jack H. Noble. Automatic segmentation of intra-cochlear anatomy in post-implantation ct of unilateral cochlear implant recipients. *MEDICAL IMAGE ANALYSIS*, 18(3):605–615, 2014.

[RSH+99] D. Rueckert, LI Sonoda, C. Hayes, DLG Hill, MO Leach, and DJ Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 18(8):712–721, 1999.

[Sha98] Craig M. Shakarji. Least-squares fitting algorithms of the nist algorithm testing system. *Journal of Research-National Institute of Standards and Technology*, 103:633–641, 1998.

[SLK+13] Kang-Jae Shin, Ju-Young Lee, Jeong-Nam Kim, Ja-Young Yoo, Chuog Shin, Wu-Chul Song, and Ki-Seok Koh. Quantitative analysis of the cochlea using three-dimensional reconstruction based on microcomputed tomographic images. *The Anatomical Record*, 296(7):1083–1088, 2013.

[SRN+03] Martin Styner, Kumar T. Rajamani, Lutz-Peter Nolte, Gabriel Zsemlye, Gábor Székely, Christopher J. Taylor, and Rhodri H. Davies. Evaluation of 3d correspondence methods for model building. volume 2732 of *Lecture Notes in Computer Science*, pages 63–75. Springer, 2003.

[SSP08] Rod R. Seeley, Trent D. Stephens, and Tate Philip. *Anatomy and Physiology, 8th Edition.* McGraw-Hill, 2008.

[Swa14] Frank Swain. I can hear wi-fi, 2014. [Online; accessed 01-July-2015].

[VKP10] D.W.F. Van Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010.

[VSC+10] Berit M. Verbist, Margaret W. Skinner, Lawrence T. Cohen, Patricia A. Leake, Chris James, Colette Boëx, Timothy A. Holden, Charles C. Finley, Peter S. Roland, J. Thomas Roland, Matt Haller, Jim F. Patrick, Claude N. Jolly, Mike A. Faltys, Jeroen J. Briaire, and Johan HM. Frijns. Consensus panel on a cochlear coordinate system applicable in histologic, physiologic, and radiologic studies of the human cochlea. *Otology and Neurotology*, 31(5):722–730, 2010.

[Wan13] Thanakul Wannaprasert. *Comparative Anatomy of the Mammalian Bony Cochlea and its Ontogenetic Development in Humans*. PhD thesis, University of Liverpool, 2013.

[WD08] Blake S. Wilson and Michael F. Dorman. Cochlear implants: A remarkable past and a brilliant future. *HEARING RESEARCH*, 242(1-2):3–21, 2008.

[WVW+14] Wilhelm Wimmer, Frederic Venail, Tom Williamson, Mohamed Akkari, Nicolas Gerber, Stefan Weber, Marco Caversaccio, Alain Uziel, and Brett Bell. Semiautomatic cochleostomy target and insertion trajectory planning for minimally invasive cochlear implantation. *BIOMED RESEARCH INTERNATIONAL*, 2014:–, 2014.

[YPH+06] Paul A. Yushkevich, Joseph Piven, Heather C. Hazlett, Rachel G. Smith, Sean Ho, James C. Gee, and Guido Gerig. User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[ZRH+08] Fan-Gang Zeng, Stephen Rebscher, William Harrison, Xiaoan Sun, and Haihong Feng. Cochlear Implants: System Design, Integration, and Evaluation. *Biomedical Engineering, IEEE Reviews in*, 1:115–142, 2008.