Automatic Segmentation of Structures and Registration of CT Images for Image-Guided Otologic Surgery and Implant Programming

By

Fitsum Aklilu Reda

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May, 2014

Nashville, Tennessee

Approved by:

Professor Benoit M. Dawant
Professor J. Michael Fitzpatrick
Professor Jack H. Noble
Professor Robert F. Labadie
Professor Robert J. Webster III

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Benoit Dawant for giving me the opportunity to be part of his group and for all his consistent support and guidance along the way all these years. His guidance has helped me progress academically and make the most out of what I was doing. He has made some of the challenging times more bearable and has contributed a great deal to my personal and professional growth. Under his tutelage, I have managed to acquire a solid background in advanced image processing techniques and how to go about researching in the field.

I would also like to express my deepest gratitude to Dr. Jack Noble for sharing his image processing expertise with me, for giving his time, for his comments and suggestions regarding my work, especially in my early days at Vanderbilt. His tool MeshEditor was also very useful for manual delineation and visualization of medical image data.

I would like to thank Dr. Robert Labadie for his support and for making the time to review and provide comments and suggestions on my work.

I would like to thank Dr. Mike Fitzpatrick and Dr. Robert Webster for their feedback on my work, and I would also like to thank the many medical specialists for the anatomical knowledge they provided me with. Validation of the techniques developed herein would not have been possible without their assistance.

I would also like to thank my family for all their support throughout my life. Finally, I wish to thank members of my laboratory Medical Image Processing Laboratory for their stimulating discussions and for their friendship. You have made my experience enjoyable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

Chapter I

Introduction

Hearing loss is the inability to sense sound waves. Approximately 1-3% of babies born in the United States have sever-to-profound hearing loss [1]. In the past, the standard of assistance for people with severe to profound sensorial hearing loss was wearable hearing aids, which amplify sound waves picked up by a microphone, or sign language. However, for patients with damaged sensory cells, amplification of sound waves alone is not adequate as it does not provide important speech cues that help discriminate or understand words.

The introduction of cochlear implants has drastically altered the rehabilitative assistance available to deaf people [2]. A cochlear implant (CI) is a neural prosthetic device that restores hearing by directly stimulating the auditory nerve using an electrode array surgically placed in the cochlea. The cochlea is a snail shaped inner ear anatomy that contains the sensory organ for hearing. During CI implantation, which has become the common treatment for deaf people, the surgeon inserts the electrode array into the cochlea and embeds an external device under the skin behind the patient's ear. The external device includes a microphone, a sound processor, and a signal transmitter component. The microphone detects sound waves. The sound processor decomposes and converts the detected sounds into electrical signals. The signal transmitter relays the electrical signals to an internally implanted receiver, which delivers the electrical signals to the implanted electrode array. The electrode array restores hearing sensation by sending the electrical stimuli to the different regions of the auditory nerve.

Conventional procedures for performing cochlear implantation involve excavation of the temporal bone region of the skull. This time-consuming and invasive excavation is necessary to safely access the cochlea and avoid damage to sensitive ear structures located between the skull and the cochlea. Typically, this cochlear implantation approach takes at least 2 hours. Several techniques have been proposed to minimize the invasiveness of conventional approaches [3]-[7]. One of the proposed techniques uses image guidance to safely access the cochlea.

Once the CI device is implanted, either via conventional techniques or image-guided techniques, the CI sound processor is programmed by an audiologist who determines a number of programming parameters that specify the electrical signals sent to implanted electrodes to attempt to optimize hearing outcome. Specifically, stimulation levels are specified for each electrode based on patient's perceived loudness, and a frequency allocation table is selected to define which electrode should be activated when specific frequencies are in the detected sound.

CI electrode arrays are designed such that when optimally placed in the lower cavity of the cochlea (the scala tympani), each electrode stimulates regions of the auditory nerve corresponding to a pre-defined frequency bandwidth [8]. However, in CI surgery, once the surgeon achieves access to the cochlea, he/she must make a small opening on the cochlea and thread the electrode array blind to the internal structures of the cochlea. Because of this, the final position of the implanted electrodes relative to the internal structures of the cochlea is unknown. Thus, the only option for CI programming has been to assume the electrode array is optimally placed in the cochlea and to use a default frequency allocation table. But CI programming technique's performance has been shown to be sensitive to how well the electrodes are positioned [9]-[10], in the cochlea. Electrodes positioned non-optimally could limit CI recipients' spectral resolution [11]-[12]. Recently, image-guidance has been proposed to aid the

determination of the position of implanted electrodes with a high degree of accuracy for individual CI recipients. This has enabled the introduction of new CI programming strategies that can be customized to individual CI recipients based on the spatial relationship between implanted electrodes and internal structures of the cochlea [22]-[27].

In the following sections, we will first present background information on the use of image guidance for minimally-invasive cochlear implantation. Then we will present background information on the use of image guidance for cochlear implant programming strategies. Within each of these sections, we will present the image processing challenges that need to be addressed to automate or extend the use of image guidance to a larger population of patients. Finally, in the last section, we will present the goal and the contribution of the dissertation.

1.1. Background on Minimally-invasive Image-guided Cochlear Implant Surgery

A recently introduced minimally-invasive image-guided cochlear implantation technique, termed percutaneous cochlear implantation (PCI), achieves safe access to the cochlea without having to drill away the mastoid portion of the temporal bone and expose the cochlea [13]. The technique is to drill a 1.5 mm diameter single linear channel from the outer part of the skull into the cochlea by passing through the facial recess (see Figure 1.1). The facial recess is a region ~2.5 mm wide bounded posteriorly by the facial nerve and anteriorly by the chorda tympani. The facial nerve is a thin tubular structure that controls ipsilateral facial memetic motion, and damage to this nerve could lead to facial paralysis. The chorda tympani is another thin tubular structure that controls ipsilateral taste of the tip of the tongue, and damage to this nerve could lead to loss of taste sensation. The PCI drilling trajectory is computed by algorithms developed to find a linear path that targets the cochlea and optimally preserves the safety of the facial nerve, chorda

tympani and other critical ear structures, such as the ossicles, ear canal, and tympanic membrane [14]. Drilling is constrained to follow the computed drilling trajectory by a patient customized micro-stereotactic drill guide, termed Microtable [15], mounted on the patient's head. Figure 1.1. shows surfaces of the ear structures and an optimally safe drilling trajectory computed based on the surfaces.

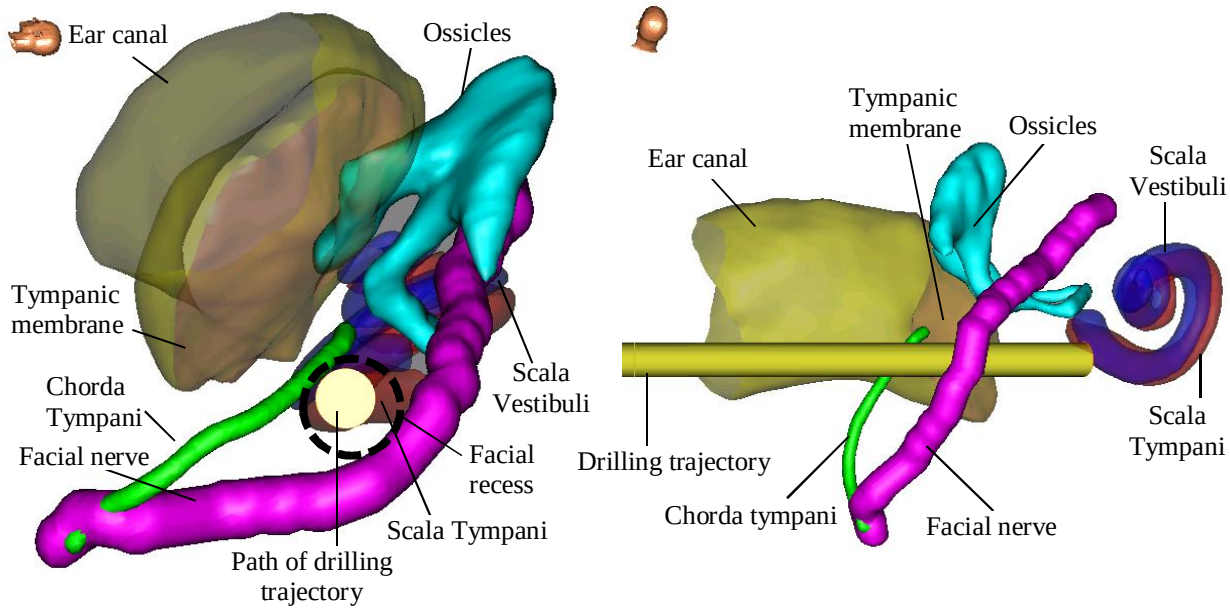


Figure1.1. Surfaces of ear anatomy and a safe drilling trajectory computed based on the surfaces, viewed in two orientations.

The PCI technique consists of two main planning steps: pre- and intra-operative planning steps.

(1) *Pre-operative planning:* Prior to surgery, a CT scan of a patient's head is acquired. Then, an optimally safe linear path is computed to target the cochlea, avoiding damage to vital ear structures. The algorithms we use for computing the PCI drilling trajectory necessitate the segmentation of several ear structures. The segmentation of the scala tympani (ST) and the scala vestibuli (SV), which are the two principal intra-cochlear structures, is achieved by a deformable shape model-based automatic method developed by our group for this purpose [16]. The segmentation of other ear structures such as the

ossicles, the ear canal, and the tympanic membrane is achieved by an automatic atlas-based segmentation method. These are structures that show high contrast in CT and for which a high degree of accuracy is not necessary. For some ear structures, such as the facial nerve and the chorda tympani, an atlas-based method is not adequate. This is because the facial nerve and the chorda tympani are thin structures (0.8-1.7 mm and 0.3-0.8 mm in diameter, respectively) and exhibit poor and inconsistent contrast with adjacent structures. Accurate facial nerve and chorda tympani segmentation is thus achieved by a model-based method developed by our group for this purpose [17].

(2) Intra-operative planning: On the day of surgery, three fiducial markers are implanted on the skull behind the patient's ear. Then an intra-operative CT scan of the head with the markers intact is acquired using a portable low-dose flat panel volume computerized tomography (fpVCT) machine (the xCAT ENT mobile CT scanner (Xoran Technologies, Ann Arbor, MI)) with voxel size of 0.3 mm × 0.3 mm × 0.3 mm. The markers serve as (a) fiducials to registers the patient's anatomy to the CT scan, and (b) a structure for securely attaching the Microtable legs. Then, the pre-operatively computed drilling trajectory is projected onto the intra-operative space using the rigid-body transformation that registers the pre- and intra-operative CTs. The registration is achieved by first manually aligning the CTs and then refining this alignment using an intensity-based rigid-body registration method that optimizes the mutual information between the two CT images [18]-[19]. The manual alignment can be performed either by manually translating and rotating the CTs or selecting three or more homologous points in each CT and then using the transformation that rigidly registers the homologous points to register the CTs.

Currently, we are working to extend the PCI concept to pediatric patients. This necessitates

(a) the segmentation of the structure of the ear in pediatric patients, and (b) the computation of optimally safe drilling trajectories using the structures' segmentations. There are anatomical differences between pediatric patients and adult patients that lead to poor segmentation quality when algorithms developed for adult ear anatomy segmentation are employed to segment ear structures in pediatric CT scans. Thus, we build a new model for pediatric ear anatomy and use it for segmenting pediatric CTs, employing the same segmentation methods used for adult patients with algorithm parameters that are optimized for pediatric patients.

One issue with the current PCI technique is that it requires a time-consuming manual intervention to achieve pre- to intra-operative CT registration. Manual adjustment is necessary because the difference in orientation and position between the pre- and intra-operative CTs is too large to be recovered by standard intensity-based registration methods. Automating the registration process is important because (a) manually initializing the registration requires someone who is expert in both temporal bone anatomy and in using the PCI surgery planning software to be present at every surgery; and (b) the registration step is a time critical process because it must be completed before the next step of the intervention can be undertaken. Since this is a critical bottleneck, manual intervention is often stressful as the extra time required to perform the registration step may prolong the surgical intervention. Thus, we developed algorithms that register the pre- to intra-operative CT fully automatically.

1.1.1. Challenges in Segmenting the Facial Nerve and Chorda Tympani

The facial nerve is a thin tubular structure that travels horizontally towards the lateral portion of the temporal bone. In the second-genu region, it bends and travels vertically in to the mastoid portion of the temporal bone, and exits through the stylomastoid foramen. Figure 1.2. shows the facial nerve and chorda tympani. To visualize the full length of the structures in panels (a) and

(b), the centerlines of the structures are mapped onto the coronal plane. This mapping is used to create a thin-plate-spline (TPS) transformation [20]. This transformation is then used to interpolate the CT images to the same plane. It can be seen in (a) and (b) that the facial nerve makes a sharper turn in pediatric patients than it does in adult patients. The chorda tympani typically branches from the vertical segment of the facial nerve approximately 1-2.5 mm superior to the stylomastoid foramen and runs at an angle to the tympanic membrane as shown in panel (c). During this study, it was observed that in some pediatric cases the chorda tympani enters the temporal bone near the stylomastoid foramen. Panel (a) in Figure 1.2 shows a chorda tympani of an adult patient that branches from the vertical segment of the facial nerve, whereas panel (b) shows a chorda tympani of an infant that originates near the stylomastoid foramen, instead of branching from the vertical segment of the facial nerve. It is also observed that the angle and the position at which the chorda tympani originates exhibit higher inter-patient variation in pediatric patients than it does in adult patients. We have also observed large variability in head size in the pediatric population. The head size is generally larger in the adult population.

Those anatomical differences between the pediatric and the adult patients lead to poor segmentation quality when a model of adult ear anatomy is employed to segment ear structures in pediatric CT scans. The techniques presented herein address the issue of anatomical differences and employ the same segmentation techniques developed for adult patients [17] for automatically segmenting the facial nerve and chorda tympani with a high degree of accuracy.

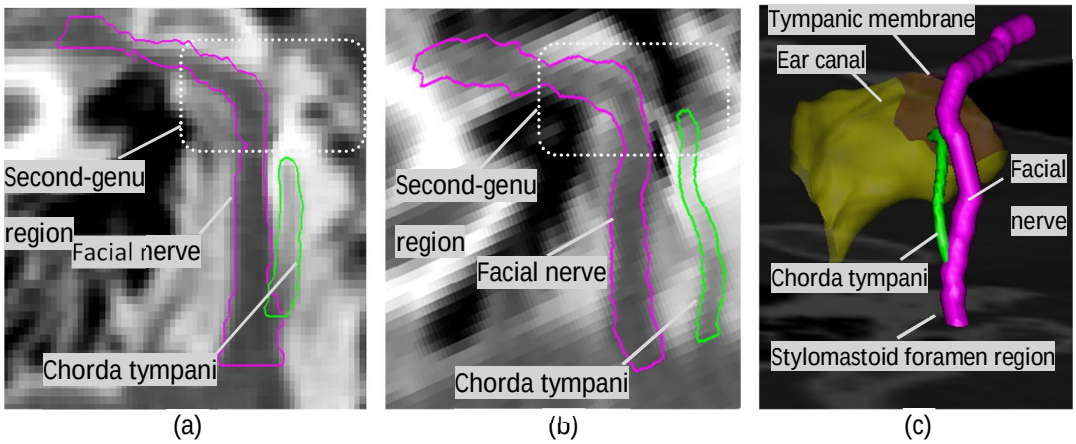


Figure1.2. Shown in (a) and (b) are surfaces of ST (red), SV (blue), and SG (green). In (b), a surface model of a CI electrode array inserted into ST is shown. In (c), contours of ST (red) and SG (green) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT, and in (d) contours of the SV (blue) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT. The bright structure in the post-implantation CTs is the CI electrode array

1.1.2. Challenges in Automating the Pre- to Intra-Operative CT Registration Process

Several properties of the intra-operative images acquired with the fpVCT complicate the automation of the pre- to intra-operative CT registration step of the PCI technique. While using an fpVCT machine is desirable because it is portable and acquires images with relatively low radiation dose, the images acquired are noisy and suffer from severe intensity inhomogeneity. This diminishes the capture range of standard intensity-based registration methods. Furthermore, the position, orientation, and field of view (FOV) of the patient's head in the intra-operative CT are unconventional and inconsistent. Thus, the variation in head orientation and position is much larger than the capture range of standard intensity-based registration methods. The inconsistent FOV results in the exclusion of standard orientation matching methods such as principal components analysis (PCA)-based alignment method. Thus, we developed a fully automatic method for pre- to intra-operative CT registration. We validated the method on several pairs of pre- and intra-operative CTs. It is fast and leads to results that are as accurate as those achieved using the manual initialization-based approach. These results suggest that the automatic approach

we developed can be used for PCI surgery.

1.2. Background on Image-guided Cochlear Implant Programming

After CI surgery is performed, cochlear implants are activated and programmed based on patient response to attempt to optimize hearing outcome. The standard techniques to do so assume that the electrodes are optimally placed in the scala tympani compartment of the cochlea. However, because surgeons must insert the electrode array blind to the intra-cochlear anatomy, the final position of the electrodes relative to the intra-cochlear anatomy is unknown. Research has shown that in 73% of CI surgeries the electrode array is placed fully within the scala tympani, while in the 27% of CI surgeries, the electrode array is fully within a neighboring cavity or is initially inserted into the scala tympani but crosses into a neighboring cavity [21]. 3D surfaces showing these structures as well as the implant are shown in Figures 1.3a and 1.3b. Examples of pre- and post-implantation CTs with overlaid structure contours are shown in Figures 1.3c and 1.3d.

We have recently developed a suite of algorithms that permit automatic determination of the electrode array position relative to intra-cochlear anatomy, using a pre- and a post-implantation CT [16], [22]-[25]. Currently, we are developing an image-guided CI programming (IGCIP) technique that uses this patient-specific spatial information to create customized sound processing strategies. Preliminary experiments performed with over thirty CI recipients indicate that this patient-customized CI programming technique can drastically improve hearing outcomes [26]-[27].

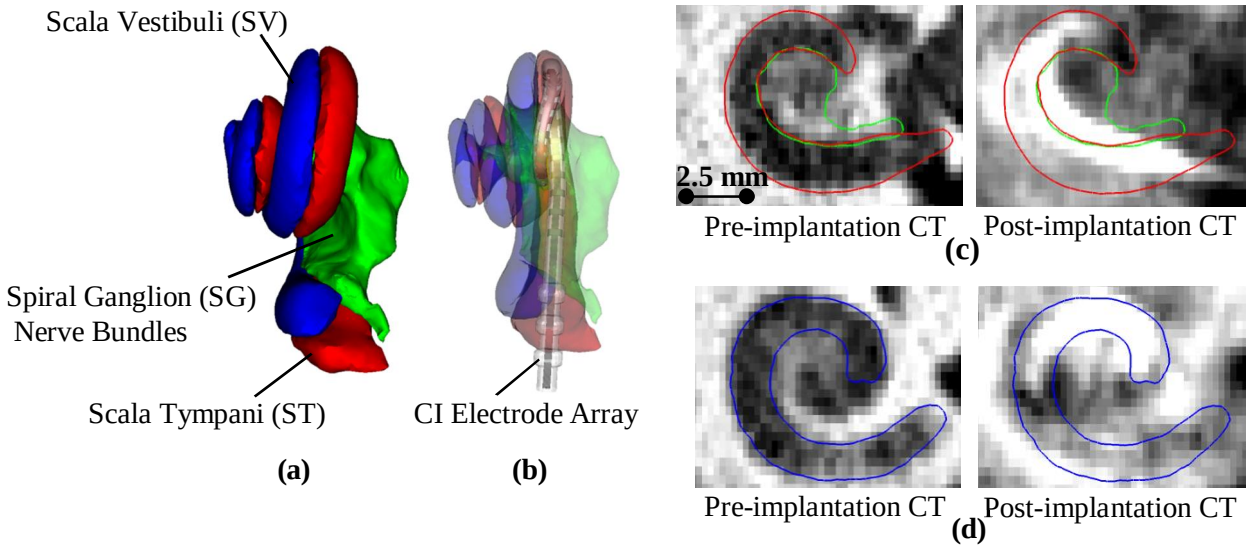


Figure1.3. Shown in (a) and (b) are surfaces of ST (red), SV (blue), and SG (green). In (b), a surface model of a CI electrode array inserted into ST is shown. In (c), contours of ST (red) and SG (green) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT, and in (d) contours of the SV (blue) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT. The bright structure in the post-implantation CTs is the CI electrode array.

The approach we previously developed for determining electrode array position relative to intra-cochlear anatomy involves several steps. First, we segment the intra-cochlear anatomy in the pre-implantation CT. Next, we identify the implanted electrodes in the post-implantation CT. Finally, we rigidly register the pre- and post-implantation CTs to determine the position of the implanted electrodes relative to the intra-cochlear anatomy. However, this approach cannot be used for many CI recipients because it requires a pre-implantation CT that is not always acquired. Thus far, the pre-implantation rather than the post-implantation CT has been used to identify the intra-cochlear anatomy because the cochlea is obscured by the metallic image artifacts introduced due to the implanted electrodes. In the work presented herein, we propose methods to extend our IGCIP strategies to the population of unilateral and bilateral CI recipients for whom a CT has not been acquired prior to implantation, thereby increasing the population of existing CI recipients who can benefit from the IGCIP strategies. The methods we developed permit automatic segmentation of the intra-cochlear anatomy in post-implantation CT of

unilateral CI recipient or bilateral CI recipients, despite the significant artifacts induced by the CI electrodes in those images.

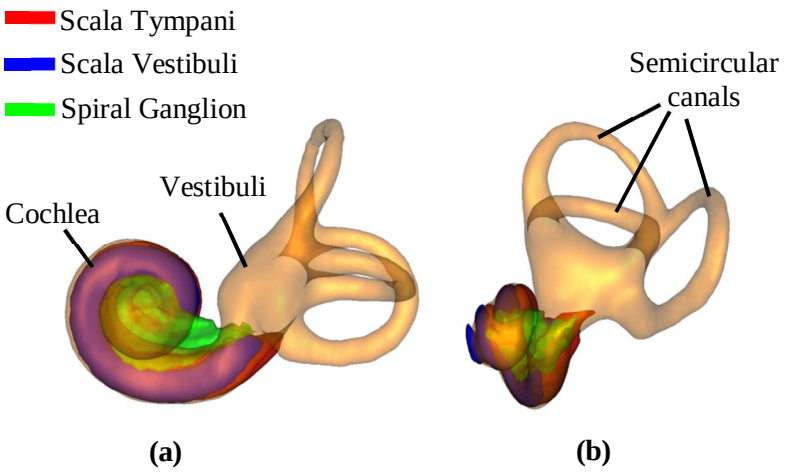


Figure1.4. Surfaces of the labyrinth (shown in transparent gold) and intra-cochlear anatomy (shown in transparent red (ST), transparent blue (SV), and transparent green (SG)) viewed in two orientations (a) and (b).

The methods proposed rely on first identify a landmark structure that (a) is in close proximity to the inner ear anatomy and (b) is unlikely to be entirely obscured by electrodes related image artifacts, and then use the position of the landmark structure to estimate the position of the inner ear anatomy. The structure we use as a landmark is the labyrinth, which is a structure that includes the semi-circular canals, the vestibuli, and the external wall of the cochlea. Figure 1.4 shows surfaces of the labyrinth and the intra-cochlear anatomy (also referred as inner ear anatomy) it externally bounds. To the best of our knowledge, there have been no methods proposed to automatically segment the inner ear anatomy in post-implantation CTs directly with a high degree of accuracy.

1.3. Goals and Contributions of the Dissertation

The goal of the dissertation is (1) to construct and use various models of pediatric ear anatomy to automatically segment the structures of the ear in pediatric CT, which we rely on to extend the PCI concept to pediatric patients, (2) to develop new algorithms for automatic pre- to intra-operative CT registration, which we use to automate the intra-operative registration step of the PCI technique, (3) to develop new algorithms for automatic segmentation of the inner ear anatomy in post-implantation CT of unilateral CI recipients, which we rely on to extend the IGCIP technique to the sub-population of unilateral CI recipients for whom a CT has not been acquired prior to cochlear implantation, and (4) to develop novel algorithms for automatic segmentation of the inner ear anatomy in post-implantation CT of bilateral CI recipients, which we rely on to extend the IGCIP technique to the sub-population of bilateral CI recipients for whom a CT has not been acquired prior to cochlear implantation.

Specifically the contributions of the dissertation are:

1) The construction and use of models of pediatric ear anatomy for the automatic segmentation of the structures of the ear in pediatric CT. The developed framework permits (a) automatic segmentation of the ossicles, tympani membrane, and ear canal in a new image by propagating models of the structures established on an atlas CT through the registration information that registers the atlas to the target CT, (b) automatic segmentation of the intra-cochlear structures using a deformable model-based segmentation method previously developed for segmenting the same anatomy in adult patients, (c) automatic segmentation of the facial nerve and chorda tympani using a model-based tubular structure segmentation method used for adults with algorithm parameters optimized for pediatric anatomy. Segmentation results

evaluated both qualitatively, by experts in ear anatomy, and quantitatively, by measuring distances to manually generated segmentations, indicate that the results are accurate enough to extend the PCI concept to pediatric population.

2) The development of a pose-invariant feature-based registration method for automatic registration of two head CTs that overlap only partially. The developed method permits automatic registration of the pre-operative CT, which we use to compute the PCI drilling trajectory and whose field of view only covers a slab of the whole head, to the intra-operative CT, which we use to transform the computed PCI drilling trajectory to the surgical space and whose field of view covers almost the whole head. The feature-based registration method we propose to register two CTs rely on first extracting the cortical surface in each CT and computing local shape features at selected points on each surface. Next, we establish a correspondence between the points in one cortical surface and the points in the other cortical surface by matching the set of shape features extracted on each surface. Finally, we compute the transformation matrix that registers the two CTs as the transformation matrix that registers the corresponding point sets, established in the previous step. Registration results validated by measuring the distance between the PCI drilling trajectory transformed using (a) the proposed automatic registration method and (b) the gold-standard semi-automatic registration method indicate that the proposed automatic registration method is accurate enough to automate the intra-operative registration step of the PCI technique.

3) The construction of a statistical shape model for the labyrinth and the development of artifact-robust, cochlear anatomy symmetry-based algorithms for automatic

segmentation of the intra-cochlear anatomy in post-implantation CT of unilateral CI recipients. The developed algorithms' approach is to acquire a post-implantation CT in which both ears are in the field of view, and to automatically segment the intra-cochlear anatomy in the implanted ear in those CT images using information extracted from the normal contralateral ear, exploiting the intra-subject symmetry in cochlear anatomy. Segmentation results evaluated quantitatively by comparing to gold-standard segmentations established on a corresponding pre-implantation CTs indicate that the developed algorithms are accurate enough to extend the IGCIP technique to unilateral CI recipients for whom a CT has not been acquired prior to implantation.

4) The construction of an inner ear anatomy shape library and the development of novel artifact-robust, shape library-based algorithms for automatic segmentation of inner ear anatomy in post-implantation CT of bilateral CI recipients. The developed algorithms' approach is to first identify a landmark structure that (a) is in close proximity to the inner ear anatomy and (b) is less likely to be entirely obscured by implant related image artifacts and then use the position of the landmark structure to estimate the shape of the inner anatomy. To identify the landmark structure, we developed a novel algorithm, which involves first mapping a shape chosen from a shape library based on its similarity to the shape in the target CT, and next refining the mapped shape using a statistical shape-model based segmentation method with a shape model that is specific to the target subject. Segmentation results evaluated quantitatively by comparing to gold-standard segmentations established on corresponding pre-implantation CTs indicate that the developed algorithms are

accurate enough to extend the IGCIP technique to bilateral CI recipients for whom a CT has not been acquired prior to implantation.

The remainder of the dissertation is organized as follows: Chapter II presents the creation and use of various models of pediatric ear anatomy for automatically segmenting ear anatomy in pediatric CT. Chapter III presents the shape feature-based registration method we propose to automatically register two CTs that overlap only partially. It presents the use of the developed method for automating the intra-operative registration step of the PCI technique. Chapter IV presents new algorithms we propose for automatically segmenting the intra-cochlear anatomy in post-implantation CT of unilateral CI recipients. It also details the studies we perform to establish intra-subject inner ear anatomy symmetry. Chapter V presents novel shape library-based algorithms we propose to automatically segment the intra-cochlear anatomy in post-implantation CT of bilateral CI recipients. It also details the process we use to create inner ear anatomy shape library. Finally, chapter VI summarizes the achievements of the research and discusses future works.

References

[1]. http://hearinghealthmatters.org/hearinginternational/2011/incidence-of-hearing-loss-around-the-world/, accessed on March 13, 2014.

[2]. Robert A. Dobie and Susan B. Van Hemel, "Hearing Loss: Determining Eligibility for Social Security Benefits", The National Academic Press, 2005.

[3]. Kiratzidis T., Arnold W., and Iliades T., "Veria operation updated. I. The trans-canal wall cochlear implantation," ORL 64:406–412, 2002.

[4]. Kronenberg J., Migirov L, and Baumgartner W., "The suprameatal approach in cochlear implant surgery: our experience with 80 patients," ORL 64:403–405, 2002.

[5]. Kronenberg J., Baumgartner W., and Migirov L., "The suprameatal approach: an alternative surgical approach to cochlear implantation," Otol Neurotol 25:41–45, 2004.

[6]. Mann W. and Gosepath J., "Technical note: minimal access surgery for cochlear implantation with MedEl devices," ORL 68:270–272, 2006.

[7]. Stratigouleas E.D., Perry B.P., King S.M. et al. "Complication rate of minimally invasive cochlear cochlear implantation," Otol. Head and Neck Surg., 135:383–386, 2006.

[8]. Wilson B.S., Dorman M.F., “Cochlear implants: Current designs and future possibilities,” J. Rehab. Res. Dev. 45(5): 695-730, 2008.

[9]. Rubenstein J.T., “How cochlear implants encode speech,” Curr. Opin. Otolaryngology Head and Neck Surg. 12(5): 444-8, 2004.

[10]. Wilson B.S., Dorman M.F., “Cochlear implants: Current designs and future possibilities,” J. Rehab. Res. Dev. 45(5): 695-730, 2008.

[11]. Fu Q.J., Nogaki G., “Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing,” J Assoc Res Otolaryngol. 6(1):19–27, 2005.

[12]. Boex C, de Balthasar C., Kos M.I., Pelizzone M., “Electrical field interactions in different cochlear implant systems,” J Acoust Soc Am 114:2049–2057, 2003.

[13]. Labadie, R. F., Balachandran, R., Noble, J. H., Blachon, G. S., Mitchell, J. E., Reda, F. A., Dawant, B. M. and Fitzpatrick, J. M. (2014), Minimally invasive image-guided cochlear implantation surgery: First report of clinical implementation. The Laryngoscope. doi: 10.1002/lary.24520.

[14]. J. H. Noble, O. Majdani, R. F. Labadie, B. Dawant, and J. M. Fitzpatrick, “Automatic determination of optimal linear drilling trajectories for cochlear access accounting for drill-positioning error,” Int. J. Med. Robot. Comput. Assist. Surg. 6(3), 281–290 (2010).

[15]. F. Labadie, J. Mitchell, R. Balachandran, and J. M. Fitzpatrick, “Customized, rapid-production microstereotactic table for surgical targeting: Description of concept and in vitro validation,” Int. J. Comput. Assist. Radiol. Surg. 4(3), 273–280 (2009).

[16]. Noble J.H., Labadie R.F., Majdani O., Dawant B.M., “Automatic segmentation of intra-cochlear anatomy in conventional CT”, IEEE Trans. on Biomedical. Eng., 58(9): 2625-32, 2011.

[17]. J. H. Noble, F. M. Warren, and B. M. Dawant, “Automatic segmentation of the facial nerve and chorda tympani in CT images using spatially dependent feature values,” Med. Phys. 35(12), 5375–5384 (2008).

[18]. F. Maes, A. Collignon, D. Vandermeulen, G. Mrchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” IEEE Trans. Med. Imaging 16, 187–198 (1997).

[19]. W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, “Multi-modal volume registration by maximization of mutual information,” Med. Image Anal. 1, 35–51 (1996).

[20]. Bookstein, Fred L., "Principal warps: thin-plate splines and the decomposition of deformations," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.11, no.6, pp.567, 585, Jun 1989.

[21]. Aschendorff, A., Kromeier, J., Klenzner, T., and Laszig, R., “Quality control after insertion of the nucleus contour and contour advance electrode in adults,” Ear Hearing, vol. 28, pp. 75S-79S, Apr. 2007.

[22]. Jack H. Noble, Theodore A. Schuman, Charles G. Wright, Robert F. Labadie, Benoit M. Dawant, “Automatic identification of cochlear implant electrode arrays for post-operative assessment”, Proc. SPIE 7962, Medical Imaging 2011: Image Processing, 796217.

[23]. Schuman TA, Noble JH, Wright CG, Wanna GB, Dawant B, Labadie, RF. “Anatomic Verification of a Novel, Non-rigid Registration Method for Precise Intrascalar Localization of Cochlear Implant Electrodes in Adult Human Temporal Bones Using Clinically-available Computerized Tomography,” The Laryngoscope, 120 (11): 2277-2283, 2010.

[24]. Wanna, G.B., Noble, J.H., McRackan, T.R., Dawant, B.M., Dietrich, M.S., Watkins, L.D., Rivas, A., Schuman, T.A., Labadie, R.F., “Assessment of electrode positions and audiological outcomes in bilateral cochlear implant patients,” Otology & Neurotology, 32(3):428-432, 2011.

[25]. Noble, J.H., Gifford, R.H., Labadie, R.F., Dawant, B.M., “Statistical Shape Model Segmentation and Frequency Mapping of Cochlear Implant Stimulation Targets in CT,” MICCAI 2012, 421-428, 2012.

[26]. Noble, J.H., Gifford, R.H., Labadie, R.F., Dawant, B.M., “Statistical Shape Model Segmentation and Frequency Mapping of Cochlear Implant Stimulation Targets in

CT," MICCAI 2012, 421-428, 2012.

[27]. Noble, J.H., Labadie, R.F., Gifford, R.H., Dawant, B.M., "Image-guidance enables new methods for customizing cochlear implant stimulation strategies," Neural Systems and Rehabilitation Engineering, IEEE Transactions on 21(5):820-829, Sept. 2013.

CHAPTER II

AUTOMATIC SEGMENTATION OF THE FACIAL NERVE AND CHORDA TYMPANI IN PEDIATRIC CT

Fitsum A. Reda[1], Jack H. Noble[1], Alejandro Rivas[2], Theodore R. McRackan[2],

Robert F. Labadie[2], and Benoit M. Dawant[1]

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232 USA

[2]Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University Medical Center, Nashville, TN 37235 USA

[This chapter has been published in *Medical Physics*, Volume 38, Issue 10, pp. 5590-5600, October 2011.]

Abstract

Cochlear implant surgery is used to implant an electrode array in the cochlea to treat hearing loss. We recently introduced a minimally invasive image-guided technique termed percutaneous cochlear implantation. This approach achieves access to the cochlea by drilling a single linear channel from the outer skull into the cochlea via the facial recess, a region bounded by the facial nerve and chorda tympani. To exploit existing methods for computing automatically safe drilling trajectories, the facial nerve and chorda tympani need to be segmented. The goal of this work is to automatically segment the facial nerve and the chorda tympani in pediatric CT scans. We have proposed an automatic technique to achieve the segmentation task in adult patients that relies on statistical models of the structures. These models contain intensity and shape information along the central axes of both structures. In this work, we attempted to use the same method to segment the structures in pediatric scans. However, we learned that substantial differences exist between the anatomy of children and that of adults, which led to poor segmentation results when an adult model is used to segment a pediatric volume. Therefore, we built a new model for pediatric cases and used it to segment pediatric scans. We employed the same segmentation method used for adults with algorithm parameters that were optimized for pediatric anatomy. A validation experiment was conducted on ten CT scans in which manually segmented structures were compared to automatically segmented structures. The mean, standard deviation, median, and maximum segmentation errors were 0.23, 0.17, 0.18, and 1.27 mm, respectively. The results indicate that accurate segmentation of the facial nerve and chorda tympani in pediatric scans is achievable, thus suggesting that safe drilling trajectories can also be computed automatically.

2.1. Introduction

Cochlear implantation (CI), a surgical technique, is routinely performed to restore hearing ability for patients that experience bilateral, severe hearing loss. In CI, an electrode array is surgically placed in the cochlea, via either a natural opening (the round window) or a surgical opening (cochleostomy), for electrical stimulation of the auditory nerve. The electrode array receives signals from externally worn components consisting of a microphone, a sound processor, a signal transmitter, and a signal receiver. The microphone senses sound waves. Then, the sound processor decomposes the sound waves, in a process that usually involves Fourier analysis, and converts them into sound signals that can be transmitted to the electrode arrays. Finally, the signal transmitter relays the sound signals to an internally implanted receiver that, in turn, transmits the electrical signals to the electrode array.

In traditional CI procedures, access to the cochlea is achieved by a wide excavation of the temporal bone region of the skull and manual accessing of the cochlea through the facial recess. Recently, we introduced a minimally invasive image-guided CI technique called Percutaneous Cochlear Implantation (PCI) [1]. PCI achieves access to the cochlea by drilling a single linear channel from the outer skull into the cochlea via the facial recess. The facial recess is a region approximately 2.5 mm wide bounded posteriorly by the facial nerve (controls ipsilateral facial mimetic motion) and anteriorly by the chorda tympani (controls ipsilateral taste to the tip of the tongue). The drilling trajectory is computed by algorithms that we developed to find a path that targets the cochlea and optimally preserves the safety of critical ear structures such as the ossicles, the ear canal, the facial nerve, and the chorda tympani [2]. Drilling is constrained to follow the computed trajectory by a patient-customized micro-stereotactic drill guide, called a Microtable, which was designed by our group [3], mounted on the patient's head. The PCI

approach involves the following four steps: (1) pre-operative planning, (2) intra-operative registration, (3) drill guide fabrication, and (4) drill guide mounting and drilling.

Step 1: Preoperative planning

A few days prior to CI surgery, a CT scan of the patient's head containing the ear region is acquired. Then, the ear structures are automatically identified and accurately segmented [4, 5]. Based on the segmented structures, a safe drilling trajectory is computed automatically [2].

Step 2: Intraoperative registration

On the day of surgery, three fiducial markers are implanted, typically at the most inferior (mastoid tip), posterior, and superior positions of the temporal bone. The marker consists of an anchor that is firmly attached to the bone, a metal sphere that serves as a fiducial marker, and a tubular extender that connects the two. A CT scan of the part of the head containing the markers and ear region is obtained using a CT scanner (e.g. xCAT ENT Mobile from Xoran Technologies, Ann Arbor, MI; voxel size $0.3 \times 0.3 \times 0.4$ mm^3). Next, the acquired intraoperative and preoperative CT scans are isotropically downsampled by a factor of four and rigidly registered using a 6 DOF (translation and rotation in three dimensions) transformation. Then, the regions of the ear are cropped from both images and subsequently registered using a 12 DOF (translation, rotation, scale and skew in three dimensions) affine transformation. The transformations are automatically estimated with an intensity-based registration method that maximizes the mutual information between the images [6, 7]. Usually, the preoperative image is acquired a few days before the surgery, but for cases where there is a substantial time gap between the preoperative CT and the surgery, this affine registration is necessary to account for local deformations caused by growth of the temporal bone. Using the compound affine

transformation, the drilling trajectory generated from the preoperative plan is transformed into the intraoperative image space, i.e., the space in which the fiducial markers are located. Finally, the centers of the markers are identified by a semi-automatic method developed by our group that starts with a user provided seed point [8, 9].

Step 3: Drill guide fabrication

The Microtable used as a patient specific drill guide is manufactured from a slab of Ultem (Quadrant Engineering Plastic Products, Reading, PA). The tabletop of the Microtable has four holes. In three of them, legs are affixed that connect it to the fiducial markers. The drill bit is guided through the fourth hole (targeting hole). Fabrication of the Microtable requires determining the location and depth of the four holes. These values are determined so that the targeting hole is collinear with the planned drilling trajectory. A component of the intra-operative software developed by our group is used to generate the command files that are used by a CNC machine (e.g. Ameritech CNC, Broussard Enterprise, Inc., Santa Fe Springs, CA) to manufacture the Microtable. The CNC machine takes less than three minutes to complete the fabrication of the Microtable.

Step 4: Drill guide mounting and drilling

Once the Microtable is fabricated, it is mounted on the marker spheres, and a drill press is attached to the targeting hole. Finally, drilling is performed lateral to the facial nerve with a wide bore drill bit (4mm diameter) and medially through the facial recess with a 1.5 mm diameter drill bit. The bit is guided through the targeting hole along the pre-operatively planned drilling trajectory and perpendicular to the tabletop of the Microtable. Figure 2.1 shows a Microtable mounted on a patient's head with a sham drill bit inserted during clinical validation testing.

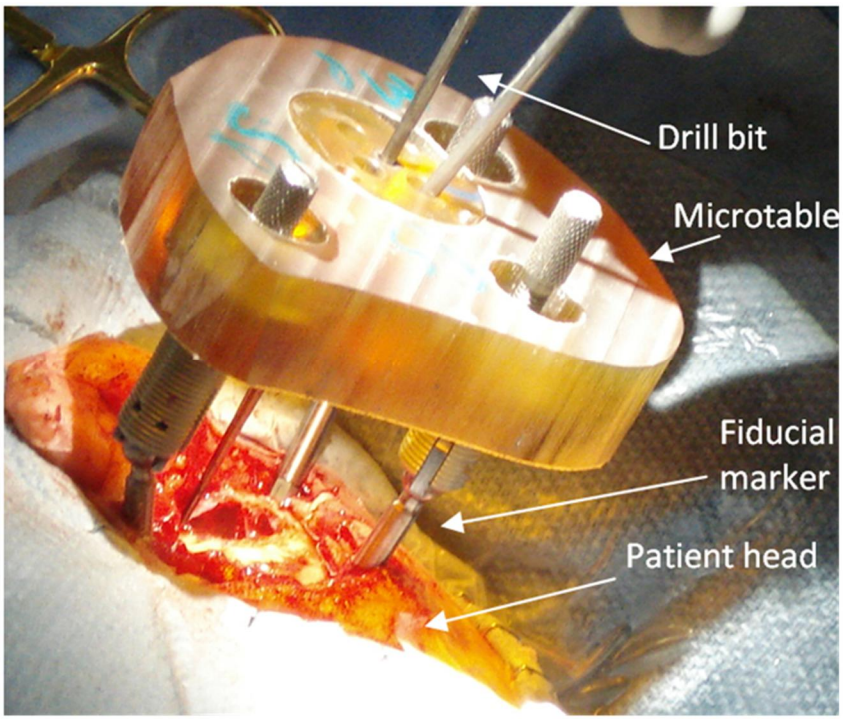


Figure 2.1. Microtable mounted on a patient head

The facial nerve, which controls the movement of the ipsilateral face, and the chorda tympani, which controls the sense of taste, are sensitive anatomical structures that are in close proximity to the desired CI drilling trajectory. Thus, to compute a safe insertion trajectory that will avoid damage to these structures, the facial nerve and chorda tympani need to be segmented. The effectiveness of traditional segmentation methods, such as atlas-based segmentation, is limited since the facial nerve and chorda tympani are thin structures (0.8-1.7 mm and 0.3-0.8 mm in diameter, respectively). They also exhibit poor contrast with adjacent structures and are surrounded by highly variable anatomy. To accurately segment these structures, we developed an automatic segmentation method that relies on a statistical model of the structures [4]. The models include intensity and shape information that varies with position along the medial-axis of the respective structures. We are now extending the PCI concept to pediatric patients. This requires segmenting the facial nerve and chorda tympani in pediatric CT scans. However, we learned that

substantial differences exist between the ear structures of adults and children. This led to poor segmentation of the facial nerve and chorda tympani when a model of adult anatomy is used to segment a pediatric CT.

The facial nerve is a thin tubular structure that travels through the temporal bone. In the second-genu region, the facial nerve bends and travels between the stapes and lateral semi-circular canal. The facial nerve then continues in the mastoid portion of the temporal bone and exits through the stylomastoid foramen. Figure 2.2 shows the facial nerve and chorda tympani. To visualize the full length of the structures in panels (a) and (b), we mapped the 3D centerlines of the structures onto the coronal plane. This mapping was used to create a thin-plate-spline (TPS)-based transformation that was then used to interpolate the CT images to the same plane. It is clearly seen in (a) and (b) that the facial nerve makes a sharper turn in pediatric patients than it does in adult patients. The chorda tympani typically branches from the vertical segment of the facial nerve approximately 1-2.5 mm superior to the stylomastoid foramen and runs at an angle to the tympanic membrane as shown in panel (c). During our study, we have observed that in some pediatric cases the chorda tympani enters the temporal bone near the stylomastoid foramen. Panel (a) in Figure 2.2 shows a chorda tympani of an adult patient that branches from the vertical segment of the facial nerve, whereas (b) shows a chorda tympani of an infant that originates near the stylomastoid foramen, instead of branching from the vertical segment of the facial nerve. We have also observed that the angle and position at which the chorda tympani originates in pediatric cases exhibit higher inter-patient variation than in adult cases.

To address those issues of anatomical differences, we have constructed a new model for pediatric patients, and we have employed the same segmentation algorithm that we used for adults with parameters optimized for pediatric populations. We report that, with this new model,

accurate and automatic segmentation of the facial nerve and the chorda tympani is achievable in pediatric patients. The chapter is structured into the following sections. Section 2.2 summarizes the method that was presented in our earlier work for adult patients [4] and describes how that method was adapted for application to pediatric patients. Results and discussions are presented in Section 2.3. Finally, conclusions and future work are detailed in Section 2.4.

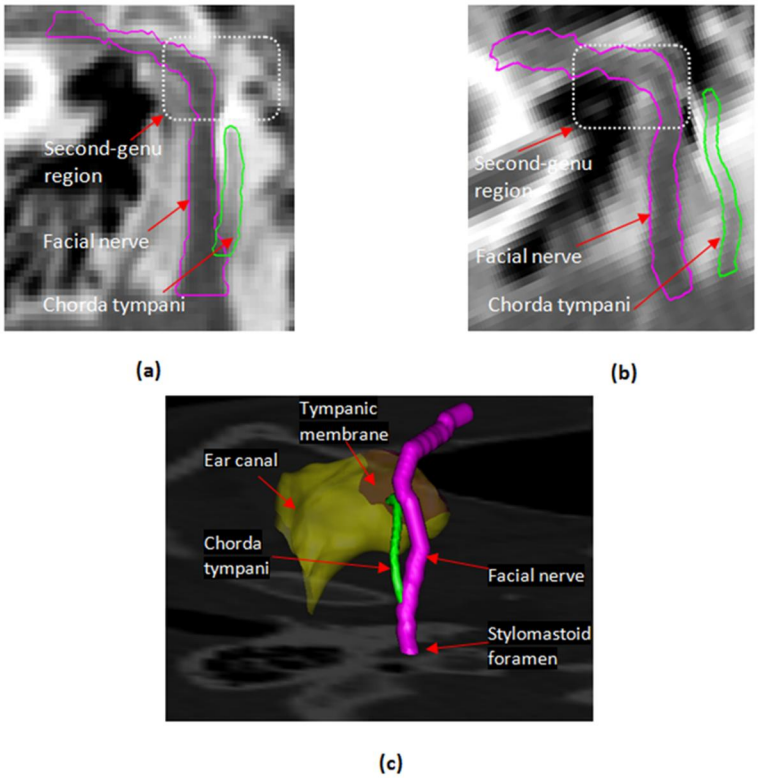


Figure 2.2. Comparison of facial nerve and chorda tympani structures in an adult and a pediatric CT scans. (a) Contours of facial nerve and chorda tympani in an adult CT. (b) Contours of facial nerve and chorda tympani in a pediatric CT. (c) 3D rendering of the anatomy

2.2. Methods

2.2.1. Data

A total of 22 pediatric scans, with age range of 11 months to 16 years, were used in this study. The images were acquired from different scanners. Typical scan resolution is 512 × 512 × 130 voxels of 0.3 × 0.3 × 0.4 mm^3 size. Out of the 22 scans, one was selected as a reference (atlas) volume, 11 were used as training volumes in order to generate the model, and the other 10 were used as test volumes. The choice for the atlas volume was done based on image quality, orientation, and size.

2.2.2. Segmentation Approach

The general approach we use to segment the structure in a target image involves extracting the centerline of the structure and then expanding it into the full structure using a standard level set method. In order to find the centerline of the structure in the target volume, we use a minimal cost path algorithm. To provide the minimum cost path algorithm with a *priori* intensity and shape costs, we create a model of the structure. The model is designed so that it can be aligned with the target volume.

2.2.3. Model Generation

The model is composed of statistical intensity and shape information associated with each voxel along the centerline of the structure of interest (SOI). The model centerline is defined as the centerline of the manually delineated SOI in the atlas volume. Each model centerline voxel is associated with statistical values of three features: *intensity*, *width*, and *unit orientation vector*. These values are computed as an ensemble average of the respective feature values measured from all the training scans' corresponding centerline points. The model is also used to create a

speed function for the level set algorithm, which expands the centerline into the full structure. The model data is stored only on the left ear of the reference volume. The right ear is modeled by reflecting the left ear model across the midsagittal plane, which is possible due to the symmetry of the human head.

The model generation process, outlined in Figure 2.3, consists of the following four steps: (A) the SOIs are manually segmented in the reference and training scans. The manual segmentations were created by a student rater (FAR) and later corrected by experienced physicians (AR and TRM). (B) The training images are aligned with the reference image by applying a series of three affine registrations. The first is a surface-based registration computed to correct the size difference between the reference and training images. This is necessary because the size of the head in the pediatric population varies substantially. This variability can be problematic for standard image registration algorithms, which are sensitive to initial position. Thus, the skull surfaces of the training volumes are first registered to the skull surface of the reference volume using the iterative closest point (ICP) algorithm [10]. This minimizes the sum of squared distances between all points from the reference surface to their closest points on the training surface. The surfaces of the skulls are extracted using the marching cubes algorithm, which creates triangle models of constant intensity surface from 3D image data [11]. For some volumes, manual initialization (rotation and translation) of the surface-based registration is necessary, because the images are acquired with the head in various positions (rotation and translation from the center of the field of view) due to patient sedation. Next, an intensity-based affine registration is applied to the images, after being downsampled by a factor of four in each dimension. Once global alignment of the training images with the reference images is achieved, the part of the scan with the ear anatomy is cropped from the training volumes using a bounding

box on the region of interest projected from the reference volume. Finally, the cropped images are registered using an intensity-based affine registration applied at full resolution. The intensity-based registrations use Powell's direction method and Brent's line search algorithm [12] to optimize the mutual information [6, 7] between the images and estimate a transformation matrix with 12 DOF (three rotations, translations, scales and skews). Applying registration on the whole volumes at a lower resolution followed by registration of the cropped regions at full resolution is computationally more efficient and, in our experience, leads to improved accuracy in the region of interest. (C) The manual delineations of the SOIs in the training volumes are projected onto the affinely registered reference space using the compound registration transformation. Then, centerlines of the manual segmentations are extracted using a topology preserving voxel thinning algorithm [13]. At each point along the extracted centerlines, values of structure width, intensity, and curve orientation are measured and stored. The orientation vector at each voxel is estimated using central differences, except for the first and last voxels where forward and reverse differences are used. (D) The reference volume is non-rigidly registered to each affinely registered training volume using an intensity-based non-rigid registration technique [14] that registers images by maximizing a normalized mutual information-based objective function [15]. Centerline points from the reference volume are projected onto each affinely registered training volume using the obtained non-rigid deformation field. A correspondence, based on the minimum Euclidean distance, is established between each reference centerline point and the closest point from each affinely registered training volume's centerline. This results in one corresponding point in each training volume for every reference point. Subsequently, statistical values of width, intensity, and orientation features at each point along the model

centerline are computed as the average of the measurements from its set of corresponding points. Features are measured on the affinely registered training images.

A. Manual segmentation of structures in the reference and training CT scans.

B. Affine registration of training CTs to the reference CT.

C. Extraction of centerlines of manually segmented structures from reference and affinely registered CTs, and measurement of the three features at each point along the centerlines.

D. Non-rigid registration of the reference CT to each of the affinely registered training CTs.

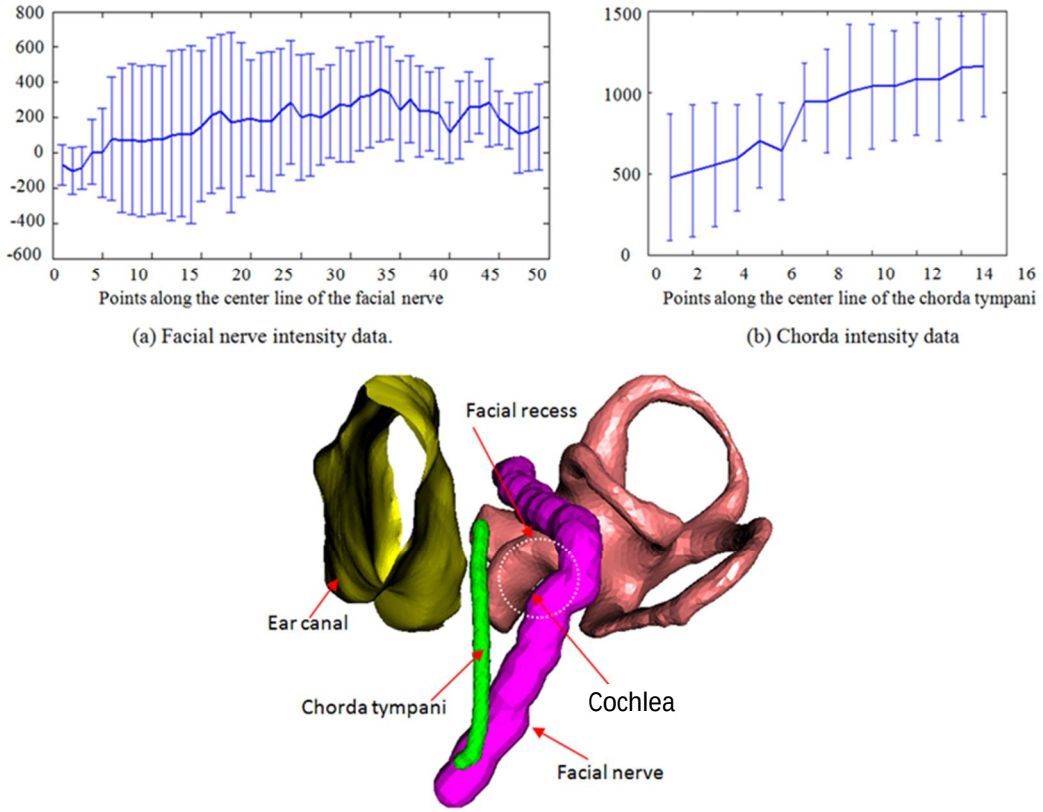
- D.1 Projecting centerline points from reference to each affinely registered training volumes.
- D.2 Establishing a correspondence between each projected point to one point in each training volume.
- D.3 Computing the average of the three features across each set of reference and corresponding points.

Figure 2.3. Model-generation steps

The above model generation process is applied to both the facial nerve and the chorda tympani to create models for these two structures. In Figure 2.4, panel (c) shows the surfaces of the facial nerve and chorda in the reference volume. The mean and standard deviation of the intensity (in Hounsfield units) along the facial nerve and the chorda tympani in the pediatric population is shown in (a) and (b), respectively. These panels clearly show that the intensity of the structures in CT scans changes along their length. It also shows that there is a considerable variation in intensity at each point along the length. This represents a challenge to segmentation methods that rely on a single threshold or even on intensity distributions.

Figure 2.5 demonstrates the differences in intensity and shape characteristics between adult and pediatric models. Panel (a) compares the intensity profiles of the facial nerve in the two populations. Although the two functions have a similar trend, the intensities along the pediatric facial nerve are consistently higher. The intensity profile of the chorda in the pediatric volumes, shown in panel (b), is consistently lower compared to the adult model. Surface models of the average adult and average pediatric facial nerves are shown in panel (c). It is clearly shown that

the facial nerve in the second-genu region makes a far more sharp turn for pediatric than adult individuals. In order to create this image, the pediatric model is projected onto the adult image space using the transformation matrix obtained by applying a series of affine registrations, as described in step (B) of the model generation procedure, to align pediatric and adult images. These intensity and shape differences explain the limited success we achieve when applying an adult model of the anatomy to segment pediatric images. Thus, we have constructed a new pediatric model for pediatric patients and employed the same segmentation algorithm we used for adults with parameters optimized for pediatric populations.



(a) Facial nerve intensity data.

(b) Chorda intensity data

Figure 2.4. Pediatric structures model data

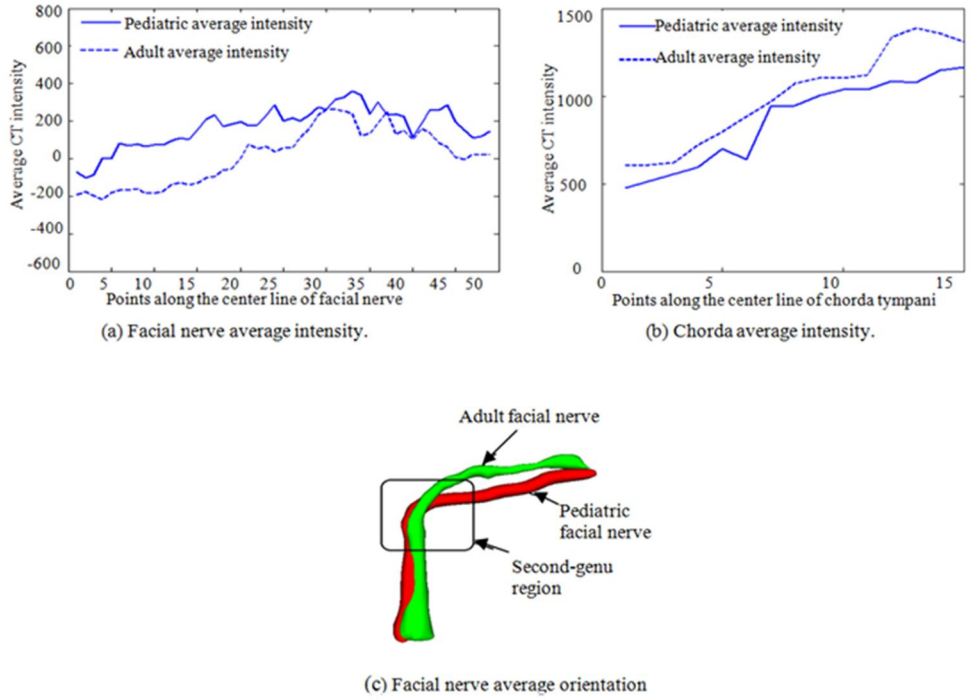


Figure 2.5. Adult and pediatric population statistical model

2.2.4. Structure Segmentation

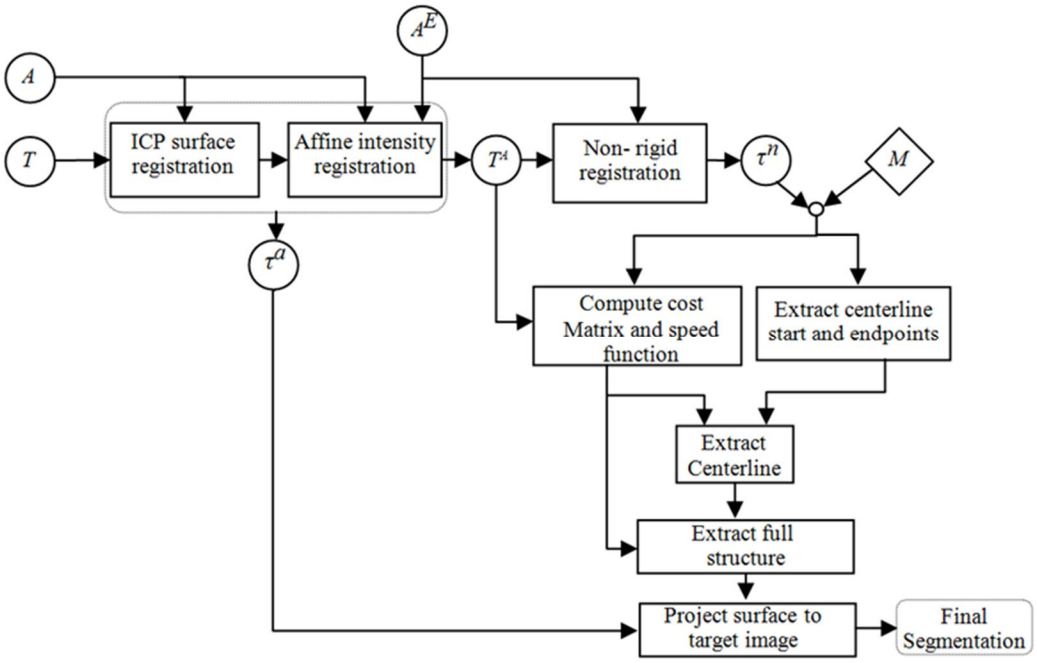


Figure 2.6. Structure segmentation flowchart.

Once created, the models can be used to segment the structures in a target image using the model-based segmentation algorithm we have developed, called the Navigated Optimal Medial Axis and Deformable-Model (NOMAD) algorithm [16]. The flow chart in Figure 2.6 shows the structure segmentation process in a target image. In the flow chart, a circle represents an image when the inside text is a Roman letter and a transformation when it is a Greek letter. Rectangles represent operations on images. The diamond represents model information. T is the target image we want to segment, A is the atlas image, and A^E is the ear region of the atlas image. T is first affinely registered to A and A^E in order to produce T^A. Then, A^E is non-rigidly registered to T^A using the adaptive bases non-rigid registration algorithm [14]. Subsequently, the model centerline points are projected onto T^A using τ^n, the non-rigid registration deformation field. Next, for each voxel in T^A, the closest point on the projected centerline is found, and a correspondence is established between that voxel and the model point. Based on the correspondence, a cost matrix, speed function, and the start and end points of the structure centerline are computed. The cost matrix and start and end points are supplied to a minimum cost path algorithm to extract structure centerline. A standard level set algorithm uses the speed function to expand the extracted centerline to full structure segmentation. Finally, the segmented surface is projected back onto T using τ^a, the affine transformation matrix.

The expressions of the cost terms associated with each feature are presented in Table 2.1. The parameters α and β are used to adjust the relative importance of each feature in the overall cost of a transition from one voxel to a neighboring voxel. The table reports the values of each of these parameters, followed by the sensitivity of the results to each parameter, as will be discussed later. A voxel in T^A is represented by $\boldsymbol{x}$. $M_I(\boldsymbol{x})$ and $\overrightarrow{M_o(\boldsymbol{x})}$ are model intensity and orientation values associated with the projected model point that is nearest to $\boldsymbol{x}$. $\vec{\boldsymbol{t}}$ is the curve

orientation at $\boldsymbol{x}$. *Nbhd(**x**)* is the set of 26 voxels that are in the neighborhood of voxel $\boldsymbol{x}$. Term 1 penalizes deviation of intensity from the intensity predicted by the model. This cost term is zero when the intensity at voxel $\boldsymbol{x}$ is equal to the intensity predicted by the model $M_I(\boldsymbol{x})$. We set the value of the normalization in this term to 2000, since, based on qualitative observations, the difference in intensity between a target and reference point is not more than 2000. Term 2 penalizes curve orientation in a direction different from the direction predicted by the model. Transitions in the direction of the predicted orientation have a cost of zero, while transitions in the opposite direction have a maximum cost. Term 3 is a model-independent term that favors voxels that are local intensity minima. The cost term is zero when the intensity at $\boldsymbol{x}$ is a local minimum and is highest when the intensity at $\boldsymbol{x}$ is a local maximum. The total cost associated with a transition from one point to a new point is the sum of term (1) and term (3) at the new point, and term (2) evaluated in the direction of the transition to the new point. This results in a 3-D cost matrix.

The start and end points of the facial nerve are identified as the center of mass of the most inferior 5% and the most anterior 5% of the projected facial nerve mask, i.e., the mask obtained after the manually delineated facial nerve mask in the atlas is projected onto T^A. For the chorda tympani, the start and end points were identified in the adult population also as the center of mass of the most inferior 5% and most superior 5% of the projected chorda mask. However, in the pediatric population this did not lead to an accurate segmentation of the chorda tympani because this structure is more variable in its starting position and orientation. To correct this, we delineated chorda tympani masks both in the reference and in each training image such that they extend 2-3 mm inferior to their true starting positions. This increases the chances for the

NOMAD algorithm to arrive at an accurate localization of the chorda tympani when it reaches the region of interest.

Table 2.1. Expression of cost terms for centerline extraction

	Cost term	Purpose	Facial nerve		Chorda tympani	
			α	β	α	β
1	$\beta[\lvert T^A(\boldsymbol{x}) - M_I(\boldsymbol{x})\rvert/2000]^\alpha$	Penalizes deviation from intensity predicted by model	1.7|30%	2.4|50%	3.4|80%	11.5|40%
2	$\beta\left[\left(1-\frac{\vec{\boldsymbol{t}}\cdot\overrightarrow{M_o(\boldsymbol{x})}}{\lvert\vec{\boldsymbol{t}}\rvert\lvert\overrightarrow{M_o(\boldsymbol{x})}\rvert}\right)/2\right]^\alpha$	Penalizes deviation from orientation predicted by model	2.4|60%	12.5|50%	4.0|30%	1.0|80%
3	$\beta\left[1-\left(\frac{\#\{\boldsymbol{y}\ in\ Nbhd(\boldsymbol{x}):\ T^A(\boldsymbol{y})>\ T^A(\boldsymbol{x})\}}{\#\{\boldsymbol{y}\ in\ Nbhd(\boldsymbol{x})\}}\right)^\alpha\right]$	Penalizes deviation from local intensity minima	4.0|40%	1.0|40%	1.2|70%	1.5|70%

Table 2.2. Speed function for level set expansion

	Speed function	Purpose	Facial nerve	
			δ	γ
1	$e^{[\gamma M_W(\boldsymbol{x})^\delta]}$	Slows the rate of propagation when the structure width is small	0.5|60%	1.0|60%
2	$e^{[-\gamma(\lvert T^A(\boldsymbol{x})-M_I(\boldsymbol{x})\rvert/2000)^\delta]}$	Slows the rate of propagation when the local intensity deviates from the predicted value	0.2|50%	1.1|70%

Once the starting points, the ending points, and the cost matrix are computed, the structure centerline is computed as the optimal path in the cost matrix using a minimum cost path finding algorithm [17]. Since the chorda tympani is a very thin structure, we complete its full segmentation by assigning a radius of 0.325 mm at each point of the centerline. Full structure segmentation of the facial nerve is accomplished by a standard geometric deformable model based on level sets. The generated centerlines are used to initialize the evolution. The rate of evolution at each centerline point is specified using the speed function, which is defined as the sum of the two expressions shown in Table 2.2. The parameters γ and δ are used to adjust the relative importance of each feature in the speed function. As is done in Table 2.1, the value of, and sensitivity of the algorithm to, each parameter is given in Table 2.2. The first term in the

speed function slows the rate when width at each centerline point is smaller, whereas the second term slows the rate when the local intensity deviates from the intensity predicted by the model. In a typical use of level set techniques, the process iterates until it convergence. Here, we fix the number of iterations to three because the lack of contrast between the structure and its surrounding leads to leakage. Thus, the values of δ and γ have been estimated on the training scans to lead to full structure segmentation in three iterations.

2.2.5. Segmentation Validation

We used two quantitative distance measures to evaluate our segmentation accuracy, which we call automatic-to-manual (AM) and manual-to-automatic (MA). To compute these distances, the surface voxels of the manual and automatic surfaces are identified. Once this is done, the MA error is computed as the Euclidean distance from each voxel on the manual surface to the closest voxel in the automatic segmentation. Similarly, the AM error is computed as the distance from each voxel on the automatic surface to the closest voxel in the manual segmentation. The non-symmetric AM and MA errors reduce to zero when the manual and automatic segmentations are in complete agreement.

The manual delineation of the structures on the testing scans was done by a student (FAR) and corrected by two experienced physicians (AR and TRM). The manual segmentations were generated only for the segments of the structures that are in close proximity to the drilling trajectory. For the facial nerve, those segments are the mastoid (vertical) and the tympanic (horizontal) segments. For the chorda, the region of interest is the segment that runs from the stylomastoid foramen to the tympanic membrane. Validation experiments were performed on the ten test volumes. A statistical model is created using the reference and the training volumes. Each test volume is then segmented using that statistical model. The parameter values used for

segmentation (see Tables 2.1 and 2.2) are selected using the procedure described in the following section.

2.2.6. Parameter Selection

The values of α and β in the cost terms (see Table 2.1) and the values of γ and δ in the speed function (see Table 2.2) were modified heuristically on the training scans until an acceptable value of the total maximum AM or MA error was found. Once good values of the parameters were obtained, each of these values were modified in 5% increments in the direction that decreased the maximum AM or MA error, until converging to a value for which the error clearly increased. The final parameter value was chosen in the generally flat region preceding the convergence value. To characterize the sensitivity of the algorithm to these parameter values, each parameter was then modified, also on the training scans, with 10% increments and decrements until the maximum AM or MA error increased by 0.5mm. The sensitivity of the algorithm to the parameter values was then measured as the percent deviation of the parameters at which this increase in error occurs. The final parameter values and the sensitivity of the algorithm to each parameter are presented in Tables 1 and 2.

2.3. Results

The result of the validation performed on the ten test volumes is presented in Table 2.3. The table presents the values of mean, standard deviation, median, and maximum errors in millimeters for both the facial nerve and the chorda tympani. The errors are measured using the two evaluation metrics, MA and AM, as discussed above. The mean and median errors for the structures are on the order of 0.2 mm (<1 voxel). Maximum distance is 1.3 mm for the facial nerve and 1.2 mm for the chorda tympani. Although the maximum distances are substantial, they

did not lead to unsafe trajectories in the set of volumes included in our validation study. To assist in the interpretation of the quantitative results, we present qualitative results of segmentation in 2D and 3D for three of the cases presented in Table 2.3 (see Figure 2.7). The cases with large chorda tympani and large facial nerve maximum errors (case 4 and 9, respectively) are presented in panels (a) and (b), respectively. In panel (c), we present the case with low chorda tympani and facial nerve maximum errors (case 10). Manual segmentation contours are in yellow while the automatic ones are in purple. In the 3D visualizations, the color encodes the distance to the closest point on the corresponding manual surface. Each 3D rendering shows a view of the segmentations along the path-of-flight of the planned drilling well (position marked with red circle) computed using the automatic segmentations. The largest localized chorda segmentation error can be seen at its superior endpoint in (a). The chorda and surrounding structures are highly variable and lack contrast in CT in this region adjacent to the tympanic membrane. The variability is so extreme that manual identification can be challenging. In (b), the largest facial nerve error is shown near the end of its horizontal segment. A case with low maximum errors is shown in (c).

Table 2.3. AM and MA segmentation errors on the ten test scans. Total refers to the mean, standard deviation, median, and max errors for all scans.

Volume	Ear	Facial Nerve							
		Mean		Std. Deviation		Median		Max	
		AM	MA	AM	MA	AM	MA	AM	MA
1	Left	0.347	0.275	0.197	0.15	0.347	0.234	0.84	0.694
2	Left	0.187	0.245	0.11	0.116	0.187	0.23	0.531	0.715
3	Right	0.198	0.24	0.137	0.122	0.198	0.222	0.857	0.888
4	Right	0.259	0.209	0.205	0.113	0.259	0.185	1.27	0.704
5	Left	0.132	0.184	0.103	0.085	0.132	0.175	0.505	0.425
6	Right	0.233	0.271	0.15	0.128	0.233	0.248	0.952	0.811
7	Left	0.223	0.247	0.156	0.124	0.223	0.227	1.135	0.842
8	Left	0.259	0.215	0.161	0.1	0.259	0.193	0.91	0.58
9	Right	0.305	0.259	0.19	0.127	0.305	0.234	1.132	0.769
10	Left	0.138	0.175	0.089	0.074	0.138	0.17	0.548	0.58
total		0.228	0.237	0.167	0.121	0.176	0.214	1.273	0.889

(a) Facial nerve segmentation errors measured in millimeters.

Volume	Ear	Chorda Tympani							
		Mean		Std. Deviation		Median		Max	
		AM	MA	AM	MA	AM	MA	AM	MA
1	Left	0.145	0.114	0.041	0.028	0.145	0.105	0.271	0.228
2	Left	0.101	0.091	0.032	0.046	0.101	0.072	0.237	0.347
3	Right	0.097	0.105	0.035	0.091	0.097	0.07	0.329	0.616
4	Right	0.318	0.295	0.282	0.278	0.318	0.128	1.25	1.07
5	Left	0.125	0.112	0.053	0.055	0.125	0.094	0.323	0.345
6	Right	0.087	0.068	0.017	0.016	0.087	0.061	0.17	0.133
7	Left	0.097	0.105	0.035	0.092	0.097	0.07	0.329	0.616
8	Left	0.155	0.136	0.042	0.045	0.155	0.123	0.301	0.317
9	Right	0.198	0.194	0.061	0.071	0.199	0.161	0.319	0.377
10	Left	0.086	0.07	0.019	0.02	0.086	0.06	0.165	0.162
total		0.141	0.129	0.097	0.099	0.095	0.077	1.241	1.07

(b) Chorda tympani segmentation errors measured in millimeters.

2.4. Conclusions

The percutaneous cochlear implantation surgery technique we have introduced requires the segmentation of the facial nerve and the chorda tympani to compute a safe drilling trajectory. In previous work presented by our group, the segmentation of these structures was achieved using an algorithm that relies on a statistical model generated from an adult population. We tested this algorithm on a pediatric population with limited success due to the substantial differences

between adult and pediatric anatomy. The differences are observed in the second-genu region of the facial nerve where the nerve makes a sharper turn in children than it does in adult patients. In addition, we observed variation in the starting position of the chorda tympani across patients. Typically, the chorda tympani branches from the vertical segment of the facial nerve. However, in pediatric patients it is not uncommon for it to exit from the stylomastoid foramen alone. In this work, to correct for the anatomical differences, a pediatric-specific statistical model is built and the same segmentation algorithm employed on adults is used for the segmentation of the facial nerve and the chorda tympani in the pediatric population.

In both the pediatric and the implementations (see Tables 1 and 2 of [4]) the algorithm is less sensitive to the speed function parameters $\{\delta, \gamma\}$ than those of the optimal path cost function $\{\alpha, \beta\}$. In the adult implementation, the orientation term is weighted the lowest for both the facial nerve and chorda tympani. In the pediatric implementation, the orientation term is weighted the lowest for the chorda tympani. In contrast to the adult implementation the orientation term is weighted the highest for the facial nerve in the pediatric implementation. We attribute this to the sharp turn of the facial nerve in pediatric individuals, which requires higher shape cost.

The automatic segmentation algorithm was evaluated on ten CT scans, resulting in mean, standard deviation, median, and maximum errors of 0.237, 0.121, 0.214, and 1.273 mm, respectively, for the facial nerve. These results are 0.141, 0.1, 0.1, and 1.241 mm for the chorda tympani. This level of accuracy is visually verified by experienced physician (AR) to be sufficient to automatically compute a safe insertion trajectory on the scans used in our study. The

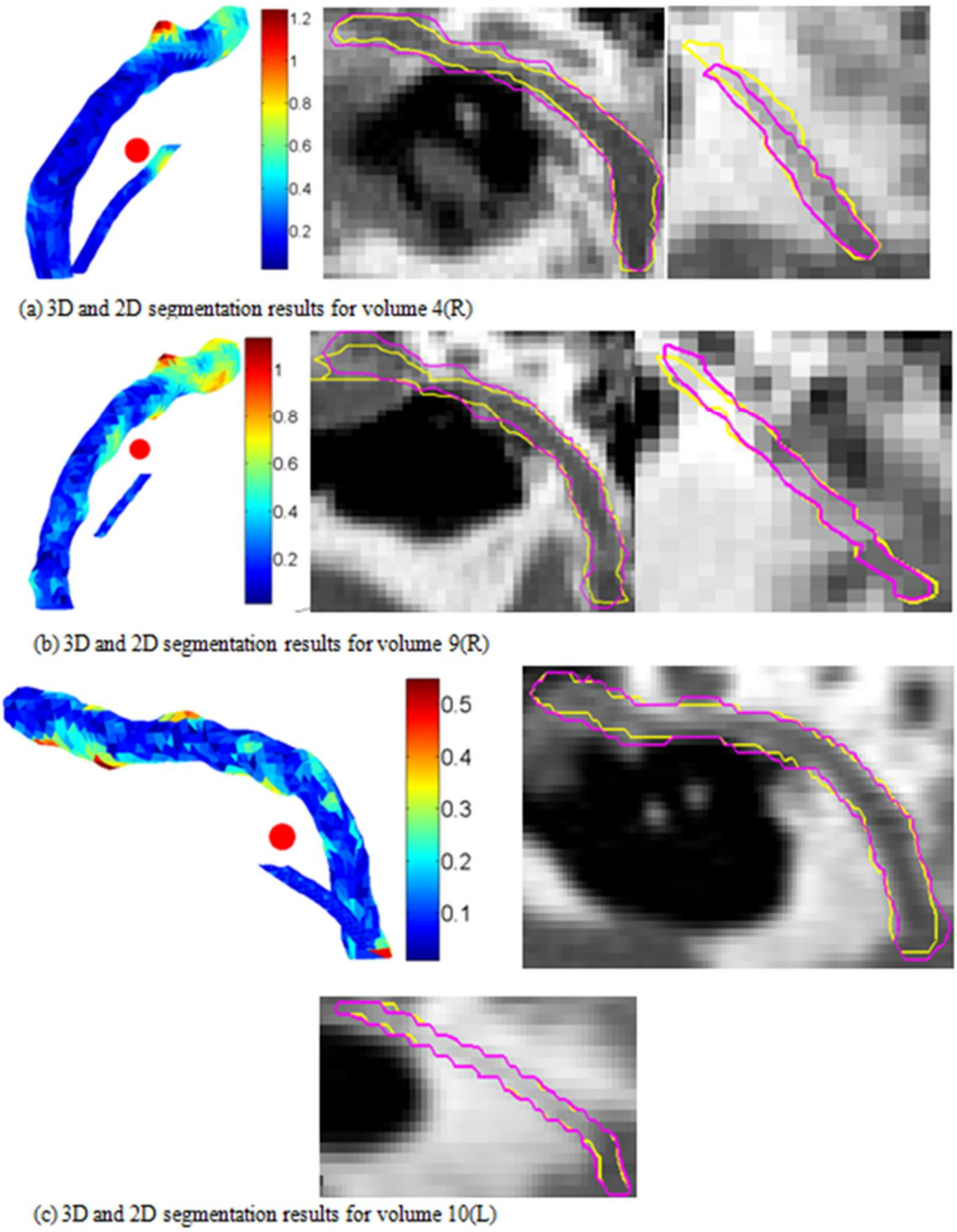


Figure 2.7. Segmentation result

results we have obtained thus suggest that the percutaneous cochlear implant approach is a viable approach for pediatric patients.

To date, this technique has been successfully used to perform PCI planning prospectively on three pediatric patients. Future work will include continuing clinical validation of the algorithm and its use as a tool for safely planning pediatric PCI surgeries.

2.5. Acknowledgments

This work was supported by NIH grants F31DC009791 and R01DC010184 from the National Institute of Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of these institutes. An early version of this article was presented at the SPIE Medical Imaging conference 2011.

References

[1]. Labadie, Robert F, Chodhury, Pallavi, Cetinkaya, Ebru, Balachandran, Ramya, Haynes, David S, Fenlon, Micahel R, Jusczyzck, Andrzej S, and Fitzpatrick, J Michael, "Minimally invasive, image-guided, facial-recess approach to the middle ear: Demonstration of the concept of percutaneous cochlear access in vitro," Otology and Neurotology, 26(4):557-562, 2005.

[2]. Noble, J. H., Majdani, O., Labadie, R. F., Dawant, B. and Fitzpatrick, J. M., "Automatic determination of optimal linear drilling trajectories for cochlear access accounting for drill-positioning error," The International Journal of Medical Robotics and Computer Assisted Surgery, 6(3): 281–290, September 2010.

[3]. Robert F. Labadie, Jason Mitchell, Ramya Balachandran, J. Michael Fitzpatrick, "Customized, rapid-production microstereotactic table for surgical targeting: description of concept and in vitro validation," International Journal of Computer Assisted Radiology and Surgery, 4(3):273-280, January 2009.

[4]. Jack H. Noble, Frank M. Warren, Benoit M. Dawant, "Automatic Segmentation of the Facial Nerve and Chorda Tympani in CT images using Spatially Dependent Feature Values," Medical Physics, 35(12):5375-5384, December 2008.

[5]. Noble JH, Dawant BM, Warren RM, Majdani O, Labadie RF, "Automatic Identification and 3-D Rendering of Temporal Bone Anatomy," Otology and Neurotology, 30(4):436-442, June 2009.

[6]. Maes F., Collignon, A., Vandermeulen, D., Mrchal G., Suetens, P. “Multimodality image registration by maximization of mutual information,” IEEE Trans. Medical Imaging, 16:187-198, 1997.

[7]. W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, “Multi-modal volume registration by maximization of mutual information,” Medical Image Analysis 1:35-51, 1996.

[8]. Labadie, Robert F., Balachandran, Ramya, Mitchell, Jason E., Noble, Jack H., et. al, “Clinical Validation Study of Percutaneous Cochlear Access using patient-Customized Microstereotactic Frames,” Otology and Neurotology, 31(1):94-99, January 2010.

[9]. Wang, M.Y. Maurer, C.R., Jr. Fitzpatrick, J.M. Maciunas, R.J., “An automatic technique for finding and localizing externally attached markers in CT and MR volume images of the head,” IEEE Transactions on Biomedical Engineering, 43(6):627-637, June 1996.

[10]. P. J. Besl and N. D. McKay, “A method for registration of 3-D Shapes”, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 14(2):239-256, February 1992.

[11]. William E. Lorensen, Harvey E. Cline, “Marching cubes: A high resolution 3D surface construction algorithm,” Computer Graphics, 21(4):163-169, July 1987.

[12]. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* , 2nd edition (Cambridge University press, Cambridge, 1992), pp. 412-419.

[13]. Bouix, S., Siddiqi, K., Tannenbaum, A., “Flux driven automatic centerline extraction,” Medical Image Analysis, 9:209-221, 2005.

[14]. Gustavo K. Rohde, Akram Aldroubi, Benoit M. Dawant, “The adaptive bases algorithm for intensity-based nonrigid image registration,” IEEE Transactions on Medical Imaging, 22(11):1470-1479, November 2003.

[15]. C. Studholme, D.L.G. Hill, and D.J. Hawkes, “An overlap invariant entropy measure of 3D medical image alignment,” Pattern Recognition 32(1):71–86, January 1999.

[16]. Jack H. Noble, and Benoit M. Dawant, “Automatic segmentation of the optic nerves and chiasm in CT and MR using the atlas-navigated optimal medial axis and deformable-model algorithm,” Proceeding of the SPIE Medical Imaging 725916, March 2009.

[17]. E. W. Dijkstra, “A note on two problems in connexion with graphs,” Numerische Mathematik 1: 269-271, 1959.

CHAPTER III

AUTOMATIC PRE- TO INTRA-OPERATIVE CT REGISTRATION FOR IMAGE-GUIDED COCHLEAR IMPLANT SURGERY

Fitsum A. Reda[1], Jack H. Noble[1], Robert F. Labadie[2], and Benoit M. Dawant[1]

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235 USA

[2]Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University Medical Center, Nashville, TN 37232 USA.

[This chapter has been published in *IEEE Transactions on Bio-medical Engineering*, Volume 19, Issue 11, pp. 3070-3077, November 2012.]

Abstract

Percutaneous cochlear implantation (PCI) is a minimally invasive image-guided cochlear implant approach, where access to the cochlea is achieved by drilling a linear channel from the skull surface to the cochlea. The PCI approach requires pre- and intra-operative planning. Computation of a safe linear drilling trajectory is performed in a pre-operative CT. This trajectory is mapped onto intra-operative space using the transformation matrix that registers the pre- and intra-operative CTs. However, the difference in orientation between the pre- and intra-operative CTs is too extreme to be recovered by standard, gradient descent based registration methods. Thus far, the registration has been initialized manually by an expert. In this work we present a method that aligns the scans completely automatically. We compared the performance of the automatic approach to the registration approach when an expert does the manual initialization on 11 pairs of scans. There is a maximum difference of 0.18 mm between the entry and target points of the trajectory mapped with expert initialization and the automatic registration method. This suggests that the automatic registration method is accurate enough to be used in a PCI surgery.

3.1. Introduction

Cochlear implantation (CI) is a procedure in which an electrode array is surgically implanted in the cochlea to treat hearing loss. The electrode array, inserted into the cochlea via either a natural opening (the round window) or a drilled opening (cochleostomy), receives signals from an external device worn behind the ear. The external device is composed of a microphone, a sound processor and a signal transmitter component. The microphone senses sound signals. The sound processor selects and arranges sound sensed by the microphone. The signal transmitter converts

the processed sound into electrical impulses and sends them to the internal receiver, which delivers the impulses to the electrodes in the implanted array. The electrodes send the electrical impulses to different regions of the auditory nerve. Conventionally, CIs require wide surgical excavation of the mastoid region of the temporal bone so that the surgeon can locate sensitive ear anatomy and achieve safe access to the cochlea. We have recently introduced a minimally invasive image-guided CI procedure referred to as percutaneous cochlear implantation (PCI) [1]. In PCI, access to the cochlea is achieved by drilling a linear channel from the outer part of the skull into the cochlea that passes within millimeters of, and avoids damage to, critical ear anatomy. The drilling trajectory is computed on a pre-operative CT scan prior to surgery using algorithms that we have developed to find a path that optimally preserves the safety of critical components of the ear including the ossicles, cochlea, external auditory canal, facial nerve, and chorda tympani [2]. The pre-operatively computed trajectory is guided by a customized micro-stereotactic frame, a device designed by our group that constrains the drill bit to follow the computed drilling trajectory to achieve safe access to the cochlea [3]. PCI is currently undergoing clinical trials at our institution.

The PCI approach consists of pre-operative planning, intra-operative registration, drill guide fabrication, and drilling, which are summarized as follows: (1) *Pre-operative planning*: A CT scan of the patient is acquired prior to surgery. Ear anatomy is automatically identified and accurately segmented using algorithms we have previously validated and reported on [4]-[7]. These algorithms rely on models of the anatomy defined on an atlas image. The algorithms start by automatically registering the atlas image to the pre-operative image. An optimally safe drilling trajectory is computed based on the segmented ear anatomy [2]. (2) *Intra-operative registration*: On the day of surgery, three fiducial markers are implanted on the region of the

skull behind the patient's ear –typically located at the inferior, posterior, and superior regions of the temporal bone. The marker consists of an anchor that is screwed into the bone, a metal sphere that serves as a fiducial marker, and a tubular extender that connects the two. Then, an intraoperative CT scan of the head with the markers in place is obtained using a flat panel volumetric computerized tomography (fpVCT) machine - the xCAT ENT mobile CT scanner (Xoran Technologies, Ann Arbor, MI) with voxel size of $0.3 \times 0.3 \times 0.3$ mm^3. The pre- and intra-operative images are manually brought into rough alignment. The manual alignment can be performed either by manually translating and rotating the images or selecting three or more homologous points in each scan. Subsequently, the images are automatically registered using an intensity-based rigid-body registration method that uses mutual information (MI) as the similarity measure [8], [9]. The marker centers are identified automatically [10]. Next, the pre-operatively computed drilling trajectory is transformed, using the obtained rigid body transformation, onto the intra-operative image space, and thus onto the same space as the identified fiducial markers. (3) *Drill guide fabrication*: The customized micro-stereotactic frame, which we refer to as a Microtable, is a patient specific drill guide that is manufactured from a slab of Ultem (Quadrant Engineering Plastic products, Reading, PA). Fabrication of the Microtable necessitates the determination of the location and depth of four holes. Three of the holes couple to the spherical extenders mounted on the bone-implanted markers and the fourth hole (targeting hole) is determined such that it is collinear with the drilling trajectory. In addition, the lengths of the three legs that connect the tabletop of the Microtable to the markers need to be specified. The intra-operative component of our proprietary software generates the command file required to manufacture the Microtable using a CNC machine (Ameritech CNC, Broussard Enterprise, Inc., Santa Fe Springs, CA). The CNC machine takes approximately four

minutes to fabricate one Microtable. (4) *Drilling*: After sterilization, the Microtable is mounted on the marker spheres, and a drill press is attached to the targeting hole. Finally, drilling is performed with a 1.5 mm diameter drill bit, which is guided along the pre-operatively planned drilling-trajectory through the targeting hole and perpendicular to the tabletop of the Microtable. Figure 3.1 illustrates the Microtable mounted on a patient's head with surgical drill attached.

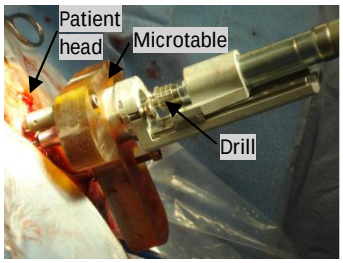


Figure 3.1. Drill attached to the Microtable, which is mounted on the patient head.

The manual registration in the intra-operative registration step of the process is typically performed by selecting three or more homologous points in each scan. The transformation matrix that registers these points is used to roughly align the scans. This article presents a method to automate the process. Automating the process is important because: (1) manually initializing the registration process requires someone who is expert in both temporal bone ear anatomy and in using the planning software to be present at every surgery; and (2) the registration step is a time critical process because it must be completed before the next step of the intervention – creation of a customized micro-stereotactic frame – can be undertaken. Since this is a critical bottleneck, manual intervention is often stressful as extra time required to perform this step may prolong the surgical intervention.

Several properties of the intra-operative images obtained with the fpVCT complicate automation of this process. While using a fpVCT machine is desirable because it is portable and acquires images with relatively low radiation dose, the images acquired are noisy and suffer from

severe intensity inhomogeneity. This diminishes the capture range of standard gradient descent-based registration techniques. Table 3.1 shows the capture ranges of the intensity-based registration for the image pairs used in this study. To compute the capture range the three translations and three rotation angles were modified in increments of 1mm and 1°, respectively, from the optimal solution until the error distance between the "entry" and "target" points along the drilling trajectory mapped using the optimal transformation and the new transformation is above 0.5 mm. The translation and rotation capture ranges are computed as the smallest of the three final translations and the three final rotation angles, respectively. As can be seen in the table, the capture range can be as small as 8 mm translation or 9° rotation. Furthermore, the position, orientation, and field of view (FOV) of the patient's head in the intra-operative CT are unconventional and inconsistent. Thus, the variation in head orientation and position is much larger than the capture range of the image registration algorithm. The inconsistent FOV results in exclusion of regions of the patient's head, which prevents the use of orientation matching techniques such as alignment of the pre- and intra-operative images by principal components analysis. Figure 3.2 shows a typical pre-registered intra-operative image (shown in bluescale) overlaid with a pre-operative image (grayscale) in axial, coronal, and sagittal views.

Table 3.1. Capture range of pre- and intra-operative image pairs.

	Translation	Rotation
1	26 mm	16°
2	14 mm	11°
3	20 mm	17°
4	13 mm	10°
5	18 mm	17°
6	12 mm	14°
7	19 mm	19°
8	36 mm	15°
9	27 mm	16°
10	23 mm	18°
11	20 mm	18°
12	8 mm	9°

We have recently presented a method for a coarse registration that is accurate enough to replace the manual initialization process currently used in the intra-operative registration step

[11]. This method relies on extracting corresponding features from each image and computes a transformation that best aligns these features. Although this method leads to results that are as accurate as the manual initialization-based approach, it cannot be used in the clinical workflow because it still requires some manual intervention and is too slow to be used in the operating room. In this work, we present a completely automatic registration method for pre- to intra-operative image registration. We have tested this approach on 11 pre- and intra-operative images. It is fast and leads to results that are as accurate as those achieved using the manual initialization-based approach. This suggests that the automatic approach we propose can be used for PCI surgery.

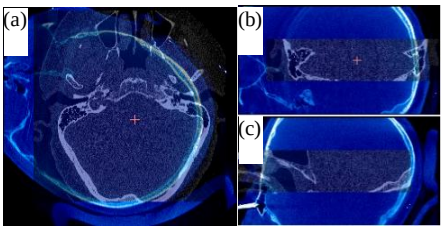


Figure 3.2. Intra-operative (blue and white) overlaid on pre-operative (black and white) CT image shown in axial (a), coronal (b) and sagittal (c) view.

3.2. Methods

3.2.1. Data

In this study, we conducted experiments on 11 pairs of pre- and intra-operative CT scans. In the planning processes, we also use a pre-operative atlas scan and an intra-operative reference scan. The scans are acquired from several scanners—GE BrightSpeed (GE Healthcare, Milwaukee, WI), Philips Mx8000 IDT, Philips iCT, and Philips Brilliance 64 (Philips Medical Systems, Eindhoven, the Netherlands) for pre-operative imaging, and a portable fpVCT machine (xCAT ENT, Xoran Technologies, Ann Arbor, MI) for intra-operative imaging. Each pair of testing images consists of pre- and intra-operative CT scans of the same patient. Typical scan

resolutions are 768 × 768 × 145 voxels with 0.2 mm × 0.2 mm × 0.3 mm voxel size for pre-operative images and 700 × 700 × 360 voxels with 0.3 mm × 0.3 mm × 0.3 mm for intra-operative images.

3.2.2. Overview

In this subsection we present an overview of the process we use to perform automatic registration of our pre- and intra-operative CTs. The approach will be detailed in following subsections. Our approach consists of two main steps. First, we perform a coarse registration using a scheme that is invariant to initial pose. Next, the registration is refined using a standard intensity-based registration. The coarse registration sub-routine is also a multistep process. Given a "target" pre-operative and "target" intra-operative CT that we wish to register, we first register the cortical surface, which is extracted using a level set segmentation scheme, of the target intra-operative image to the cortical surface of a reference volume, which we refer to as the intra-operative reference volume, using a pose invariant surface registration algorithm [12]. This reference volume is registered automatically offline to the target pre-operative CT. The final coarse registration between the pre- and intra-operative CTs is computed using the compound transformation. This multistep approach is used rather than performing a surface registration between the target pre- and intra-operative CTs directly because the surface registration algorithm is sensitive to differences in FOV and because the target pre-operative CT is typically limited in FOV to only the temporal bone region. Thus we have instead chosen to perform registration with a reference volume in which the entire cortical surface is included.

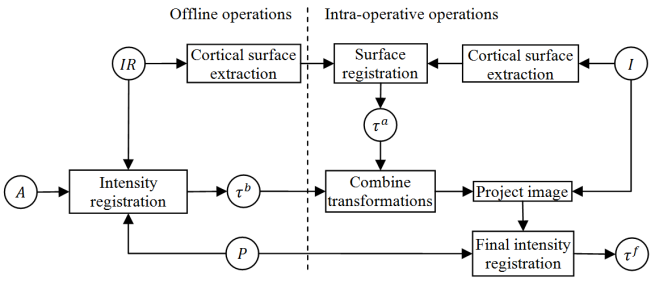


Figure 3.3. Registration flow chart.

A flow chart of the pre- to intra-operative CT registration process is shown in Figure 3.3. In this flow chart, a rectangle represents an operation on images, and a circle represents a transformation matrix when the text is in Greek letters and an image when the text is in Roman letters. *P* and *I* are the target pre- and intra-operative images we want to register. *IR* is the intra-operative reference image. *IR* is registered by hand once offline to a pre-operative atlas CT *A*, and *A* is automatically registered to *P* in the pre-operative planning stage using standard intensity-based techniques. Thus, using the compound transformation, offline registration between *IR* and *P* is achieved automatically prior to surgery. The cortical surface of *IR* is extracted with a procedure described in Section 3.2.3. The first step of the process that must be performed online intra-operatively is to apply the same surface extraction technique done offline on *IR* to *I* for cortical surface extraction of the target intra-operative CT. Then, the cortical surface of *I* is rigidly registered to the cortical surface of *IR* via a feature-based registration method called spin-image registration described in Section 3.2.4 [12]. Then, to define the coarse registration between *I* to *P*, we combine the transformation matrices obtained from the spin-image registration and the offline intensity-based registration, τ^a and τ^b. The final registration transformation τ^f is obtained by refining the coarse registration by performing a standard intensity-based rigid registration between the coarsely registered images.

3.2.3. Level Set Segmentation of the Cortical Surface

The cortical surface was chosen as the surface of interest for registration because it has surface features that are spatially distinct yet similar across subjects. To extract the 3D cortical surface in the intra-operative CT images, we use a level set segmentation method [13]. The level set evolves a surface using information from a high dimensional function. The high dimensional time-dependent function, usually defined as a signed distance map, is called the embedding function $\phi(\boldsymbol{x}, t)$, and the zero level set $\Gamma(\boldsymbol{x}, t) = \{\phi(\boldsymbol{x}, t) = 0\}$ represents the evolving surface. The evolution of the surface in time is governed by

$$\frac{\partial \phi}{\partial t} = -|\nabla \phi| \left[\alpha D(I) + (1 - \alpha) \nabla \bullet \frac{\nabla \phi}{|\nabla \phi|} \right]. \tag{3.1}$$

The term $D(I)$ specifies the speed of evolution at each voxel in *I*, and the mean curvature $\nabla \bullet \nabla \phi / |\nabla \phi|$ is a regularizing term that constrains the evolving surface to be smooth. We designed the speed term that guides the evolution of the surface using the result obtained after applying a "sheetness" filter to *I*, described in the following subsection. The level set method also requires the initial embedding function $\phi(\boldsymbol{x}, t = 0)$ to be defined. We initialize the embedding function automatically with a procedure described below. In the experiments we conducted, α is empirically set to 0.8.

Sheetness Filter

As will be described below, our speed function and our procedure for initialization of the embedding function rely on voxel "sheetness" scores computed by applying a sheetness filter to *I* [14]. The sheetness filter uses the eigenvalues of the local Hessian matrix to compute a sheetness score that is high for voxels that are near centers of sheet-like structures and low otherwise. To

compute this quantity, three ratios, $R_{\text{sheet}}, R_{\text{blob}}$, and R_{noise} defined below, are computed. For a given voxel, $\boldsymbol{x}$, $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ are the eigenvalues of the local Hessian matrix.

Sheet-like structures result in eigenvalue conditions $|\lambda_1| \approx |\lambda_2| \approx 0, |\lambda_3| \gg 0$, and the corresponding ratio (R_{sheet}) is zero for these structures. Blob-like structures result in eigenvalues conditions $|\lambda_1| \approx |\lambda_2| \approx |\lambda_3| \gg 0$, and the corresponding ration (R_{blob}) is zero for small aggregations of tissues. Then, the sheetness measure, S, which is defined as the maximum score over all scales σ at which the Hessian is computed, can be computed using the following equations:

$$R_{\text{sheet}} = \frac{|\lambda_2|}{|\lambda_3|} = \begin{cases} 0 & \text{for sheets} \\ 1 & \text{for tubes} \\ 1 & \text{for blobs} \end{cases}, \tag{3.2}$$

$$R_{\text{blob}} = \frac{|(2|\lambda_3| - |\lambda_2| - |\lambda_1|)|}{|\lambda_3|} = \begin{cases} 2 & \text{for sheets} \\ 1 & \text{for tubes} \\ 0 & \text{for blobs} \end{cases}, \tag{3.3}$$

$$R_{\text{noise}} = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2} = \begin{cases} \lambda_3 & \text{for sheets} \\ \sqrt{2}\lambda_3 & \text{for tubes} \\ \sqrt{3}\lambda_3 & \text{for blobs} \end{cases}, \tag{3.4}$$

and

$$S(\boldsymbol{x}) = \max_{\sigma \in \Sigma} S(\boldsymbol{x}, \sigma) = \begin{cases} 0 & \text{if } \lambda_3 > 0 \\ \left(e^{-\frac{R_{\text{sheet}}^2}{2\alpha^2}}\right)\left(1 - e^{-\frac{R_{\text{blob}}^2}{2\beta^2}}\right)\left(1 - e^{-\frac{R_{\text{noise}}^2}{2\gamma^2}}\right) & \text{if } \lambda_3 \leq 0 \end{cases}. \tag{3.5}$$

The values of α, β and γ, as suggested in [14], are chosen to be 0.5, 0.5, and 500. Σ refers to the scales of the Hessian matrix, which are chosen to be {0.5, 1.0,…, 3 voxels}. At a given scale, if $\lambda_3 > 0$, which occurs when the filter detects a dark structure with bright background, $S(\boldsymbol{x})$ is set to 0 because we wish to detect bone, which is bright in CT. When $\lambda_3 \leq 0$, the equation is

designed so that $S(\boldsymbol{x})$ is high when a sheetlike structure is detected. The overall sheetness score $S(\boldsymbol{x}) \in [0,1]$ will be high for bright sheetlike structures, which includes bone as well as some soft tissue structures. Figure 3.4b shows the resulting sheetness score *H* of the image *I* in Figure 3.4a.

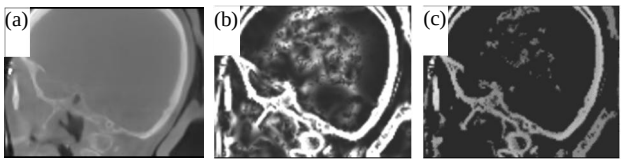


Figure 3.4. Images used in the level set initialization process. (a) Sagittal view of intra-operative CT, (b) *H*, the sheetness filter output, and (c) voxels used to estimate T_{bone}.

1. Compute a threshold, T_{bone}, that optimally separates the bone and soft-tissue.
2. Perform a coarse segmentation of the skull (bones).
   2.1. Threshold the images using T_{bone} to keep the bones.
   2.2. Dilate the resulting image from 2.1.
   2.3. Compute c_m.
3. Define the initial position of the evolving front as the surface of c_m.

Figure 3.5. Level set initialization method

Level Set Initialization

We initialize the embedding function as a signed distance map with zero level inside the surface that represents the interface between the skull and the brain and design our speed function to expand until reaching the brain/skull interface. Since some parts of the cortical surface have little contrast with surrounding structures, leaking of the level set could occur. To minimize the possibility of leakage, we have designed an approach in which we only propagate the evolving front for a fixed number of iterations (20 in our experiments). We initialize the evolving front such that its distance to the external cortical surface is approximately constant over its surface so that the required number of iterations is consistent. The procedure we use to identify this initialization surface consists of three main steps that are outlined in Figure 3.5: (1)

A threshold, T_{bone}, that optimally separates the bone from soft-tissue structures is computed based on the intensity histogram of the image using the Reddi's method [15]. However, instead of trying to compute a value for T_{bone} using the histogram of the whole image, which generally includes several peaks and valleys, we limit the histogram to contain information only from voxels that correspond to bone and sheetlike soft tissue structures, creating a histogram with one distinct valley, and thus simplifying the problem. Specifically, we use the intensity histogram of voxels with: (a) a sheetness score greater than 0.5, which removes information from noisy voxels that do not belong to bright sheetlike structures such as bone and sharpens the histogram so that the valleys are more distinct; and (b) intensity greater than -100, which removes extraneous valleys that exist at lower intensities. Figure 3.6 shows the intensity histogram of voxels that satisfy conditions (a) and (b) (shown as the red curve) and voxels that satisfy only (b) (shown as the

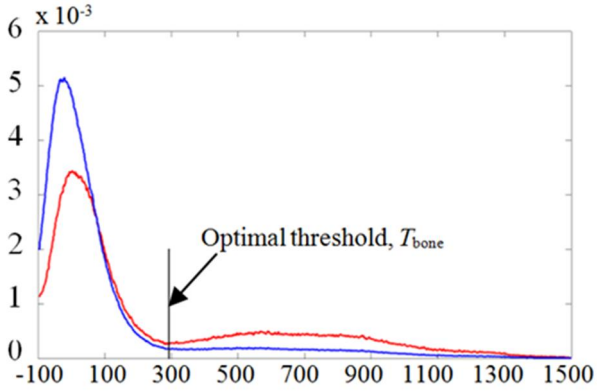


Figure 3.6. Shown in blue is the intensity histogram of voxels that have an intensity value greater than -100. Shown in red is the intensity histogram of voxels that have both an intensity value greater -100 and sheetness score greater than 0.5.

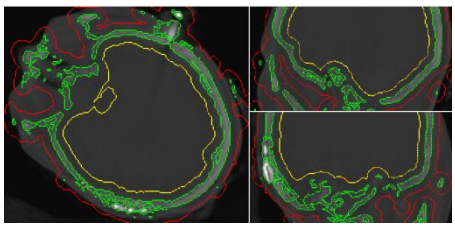


Figure 3.7. Shown in green are the contours of the binary skull segmentation. Shown in yellow and red are the contours of c_m and of $\{C\}$-c_m.

blue curve). (2) A coarse segmentation of the skull is performed by thresholding the image (for instance the target intra-operative image I) using T_{bone}. This results in a binary image I_{b} that contains the skull, some sheetlike soft-tissue structures, and some metal-related artifacts. Then, we dilate I_{b} with a spherical structuring element with a diameter of 6 mm. This is performed to enlarge the boundaries of the skull in I_{b}, and the resulting image is denoted as I_{db}. The contours of I_{b} are shown in green in Figure 3.7. In the same figure, parts of I_{db}'s contours located outside the skull are shown in red and the remaining parts are shown in yellow (the yellow contours are also the result of the next step). Next, to detect a set of initialization voxels that lie inside the cortical surface, we compute $c_m = \text{argmax}_{c\in\{C\}}(\sum_{i\in c} I(i))$, where $\{C\}$ is the set of all 26-connected components of the background of I_{db} and i is the coordinate of the voxel in I that belongs to c. Thus, c_m is the background component of the I_{db} that, when used to mask the original image I, results in the maximum sum of image intensities, and should correspond to the component that lies within the cortical surface. The contours of c_m computed for a volume are shown in yellow in Figure 3.7. Contours of all other components, $\{C\} - c_m$, are shown in red in Figure 3.7. The surface of c_m is used to define the initial position of the evolving front. While the binary skull segmentation I_{b} itself approximates the cortical surface well as seen in green in the figure, it alone is too noisy to identify and separate the cortical surface from other structures. Our technique is to apply an extreme dilation to that data. This removes noise and allows identification of a separable surface, the surface of c_m, which is close to the cortical surface and can be used to initialize the level set segmentation.

Level Set Segmentation

The speed function is set to $D = 1 - H$, where H is the sheetness score image, which ranges in value from 0 to 1. Instead of defining the speed function using the intensity or intensity

gradient type information, which would be very

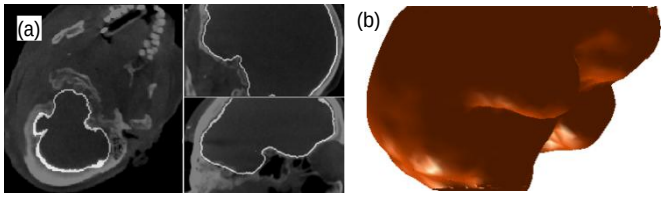


Figure 3.8. Result of level set segmentation. Shown in (a) in white are the contours of the cortical surface level set segmentation result. In (b) is a 3D triangle mesh representation of the resulting cortical surface.

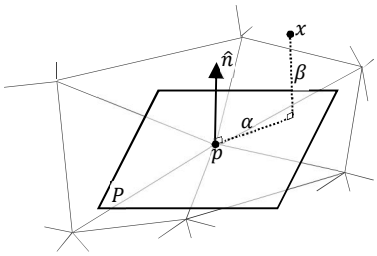


Figure 3.9. The distances **α** and **β** that are used for constructing the spin image at a vertex **p**.

noisy in this application, we use this sheetness score based approach, which consistently assigns low speeds to voxels where there are bones. Thus, the speed function will expand the evolving surface until the zero level set reaches the brain/skull interface where it will be slowed. Once the speed function is computed, the level set segmentation can be performed. An example segmentation result is shown in Figure 3.8a, and the 3D surface representation of the segmentation result is shown in Figure 3.8b.

3.2.4. Cortical Surface Registration

Once the cortical surfaces of the target and reference intra-operative CT are extracted using the technique described in the previous section, the next step is registration of the target cortical surface to the reference cortical surface. We perform the registration using a feature-based matching called the spin image method [12].

Spin Image Generation

The first step in this feature-based registration approach is feature extraction. For each vertex on the surface, features are measured in the form of a so-called "spin image," which captures the local shape of the 3D surface [12]. The spin image is a 2D histogram that describes the organization of neighboring vertices around a vertex in the surface. As shown in Figure 3.9, given a vertex p on the surface with unit normal vector $\hat{n}$ and a plane P passing through p and perpendicular to $\hat{n}$, two distances are computed from each other vertex x to the given vertex p: (1) the signed distance in the $\hat{n}$ direction, $\beta(p,x) = \hat{n} \bullet (x - p)$ and (2) the distance perpendicular to $\hat{n}$, $\alpha(p,x) = \|x - (p + \beta(p,x)\hat{n})\|$. These distances are then used in constructing the spin images, one for each vertex. A spin image is a 2D histogram with α on the *x*-axis and β on the *y*-axis. Each entry on the histogram represents the number of vertices in a neighborhood of the vertex for which the spin image is computed that belong to the entry.

Several parameter values are chosen for computing the spin image. These parameters are: (1) The size of the bins of the histogram. The bin size is usually determined as a multiple of the mesh resolution. The resolution of the surface mesh is the average length of the edges in the surface. The bin size affects how the vertex information is distributed in the spin image. (2) The height and width of the spin image. In our experiments, we have set the height and width of the spin image to be equal. The product of image width and bin size determines the support distance, which is the maximum distance a vertex can be from p and still contribute to p's spin image. (3) The support angle is defined as the maximum allowable absolute angle difference between $\hat{n}$ (the normal of p) and the normal of the contributing vertices. This is another mechanism for limiting contributing vertices to a local region. For this experiment the width and height of the spin image are set to 40, the bin size is set to half of the mesh resolution, and the support angle is set to 60°.

Establishing Spin Image Correspondence

To perform spin image registration, a correspondence between the vertices of the reference and target surfaces must first be established by matching the spin images associated with those vertices. To choose which reference volume vertices are included in the registration, at each vertex in the extracted reference surface, we compute a curvature measure and normalize it to range from 0 to 1 (see Figure 3.10) [16]. Only vertices on the reference surface with curvature value above 0.25 are used for the spin image registration. We do this because the regions of the cortical surface where the curvature is low are those that are flat, and their associated spin images are similar to those of their neighboring vertices. Thus, in the registration process we only include vertices in high curvature regions that are more likely to result in distinctive spin images. Spin image computation for these vertices is performed once and offline. Similarly, we only use 30% of the vertices from the target surface, which are chosen as the vertices with the highest curvature scores. Limiting the number of vertices included in the registration, and thus also limiting the number of spin images that need to be matched, improves the computation time of the subsequent search for correspondence.

Point correspondence is established by matching the spin image of each reference vertex to the spin image of the vertex in the target surface that maximizes a linear correlation-based similarity criterion. Correspondences are constrained such that if $C_1 = (r_1, t_1)$ and $C_2 = (r_2, t_2)$ are two sets of corresponding points between the reference and target surfaces, $\left\|\left(\alpha(r_1, r_2), \beta(r_1, r_2)\right) - \left(\alpha(t_1, t_2), \beta(t_1, t_2)\right)\right\| < \varepsilon$, i.e., the spin image coordinates from $r_1$to r_2 on the reference surface are approximately equal to spin image coordinates from $t_1$to t_2 on the target surface. This type of constraint is enforced between all sets of corresponding points to filter out point correspondences that do not obey rigid distance constraints. For more detailed description of the spin image matching process, please see [12]. Once correspondence is

established, a rigid-body transformation that best aligns the corresponding points is computed using the method of least squares fitting [17]. Figure 3.11 shows an example of three pairs of corresponding points and their associated spin images.

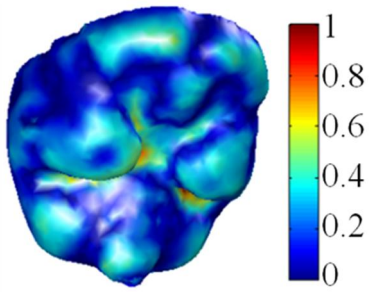


Figure 3.10. Reference intra-operative cortical surface. The color at each vertex encodes the curvature value.

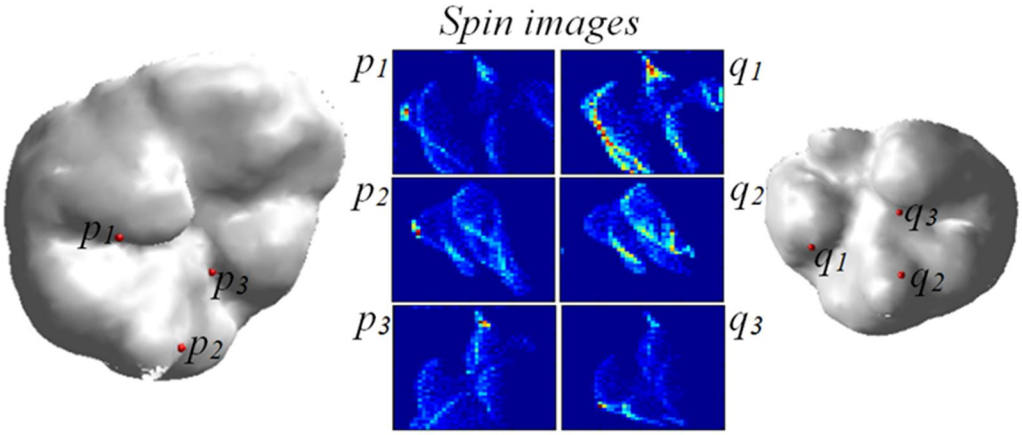


Figure 3.11. Cortical surface of reference and target intra-operative CT images. Three correspondences and their associated spin images are shown.

3.2.5. Validation

Each testing pre- and intra-operative image pair was registered using both the expert initialization-based and the automatic registration method we propose. We quantitatively validate our automatic approach by measuring the distance between the "entry" (a point along the typical surgical trajectory for PCI near critical ear anatomy) and "target" (cochlear implant insertion point) points computed using the automatic and manually initialized registration processes. Expert initialization has led to satisfactory results in the clinical trials we have been performing [3]. Thus, small errors between entry and target points will indicate that the automatic approach

is equally effective.

3.3. Results

Table 3.2 shows the error distances between the target and entry points generated using the fully-automated vs. the manual initialization-based registration approach. The maximum distance between points using the two approaches is 0.18 mm, and the average distances between entry and target points are 0.116 and 0.118 mm, respectively. These results suggest that the automatic registration we propose is accurate enough to perform a PCI surgery.

Table 3.2. Distance in millimeter from the "entry" and "target" points of the drilling trajectory that is mapped with expert initialized registration to the proposed automatic registration

Test pair	Entry Point	Target Point
1	0.030	0.029
2	0.169	0.180
3	0.102	0.094
4	0.157	0.161
5	0.124	0.123
6	0.127	0.129
7	0.091	0.093
8	0.108	0.123
9	0.153	0.148
10	0.096	0.095
11	0.117	0.117
Average	0.1159	0.1175

3.4. Conclusions

PCI surgery requires the registration of the pre- and intra-operative images to map the pre-operatively computed drilling trajectory onto the intra-operative space. The field of view and the orientation of the patient's head in the intra-operative CTs are inconsistent. These differences between the pre- and intra-operative CTs are too extreme to be recovered by standard, gradient descent-based registration methods. In this work, we present a completely automatic method for pre- to intra-operative CT registration for PCI. This approach relies on a feature-based registration method that, to the best of our knowledge, has not been used by the medical imaging

community. We found this technique to be efficient and accurate.

To quantitatively measure performance, we compared the target and entry points of an automatically registered trajectory to a trajectory mapped using the manually-initialized approach, which has been used in ongoing clinical validation studies [3], and we have found a maximum error distance of 0.18 mm. However, since both approaches use the same intensity-based registration approach as the final optimization step and converge to similar results, it is likely that both approaches produce equally accurate results. We are currently evaluating the automatic procedure prospectively to confirm this.

We recently presented another method for automating the manual initialization process that also relies on surface registration [11]. In that method, surface registration is performed by matching features on the skull surface. The drawbacks of that method are that the skull surface extraction requires manual intervention and the time required to perform surface extraction is ~20 min. The advantage of the proposed approach is that it eliminates all manual intervention, and it only requires 0.75 min, which is fast enough to be integrated into the PCI workflow since the manual initialization-based approach we currently use typically requires more than 2 min.

The proposed method could be applied to any registration problem with large misalignment. Although it will not replace existing accurate registration methods, it could be used for initialization. The accuracy of the method is limited because the spin images computed in local neighborhood vertices are not very different. This introduces ambiguity in the search for correspondence.

One limitation of the proposed registration approach is that it is not invariant to scale. Future work will focus on addressing this problem.

References

[1] R. F. Labadie, P. Chodhury, E. Cetinkaya, R. Balachandran, D. S. Haynes, M. R. Fenlon, A. S. Jusczyzck, and J. M. Fitzpatrick, "Minimally invasive, image-guided, facial-recess approach to the middle ear: Demonstration of the concept of percutaneous cochlear access in vitro," vol. 26, no. 4, pp. 557-562, Jul. 2005.

[2] J. H. Noble, O. Majdani, R. F. Labadie, B. M. Dawant, and J. M. Fitzpatrick, "Automatic determination of optimal linear drilling trajectories for cochlear access accounting for drill-positioning error," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 6, no. 3, pp. 281-290, Sept. 2010.

[3] R. F. Labadie, J. Mitchell, R. Balachandran, and J. M. Fitzpatrick, "Customized, rapid-production microstereotactic table for surgical targeting: description of concept and in vitro validation," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 4, no. 3, pp. 273-280, May 2009.

[4] J. H. Noble, F. M. Warren, R. F. Labadie, and B. M. Dawant, "Automatic segmentation of the facial nerve and chorda tympani in CT images using the spatially dependent feature values," *Med. Phys.*, vol. 35, no. 12, pp. 5375-5384, Dec. 2008.

[5] F. A. Reda, J. H. Noble, A. Rivas, T. R. McRackan, R. F. Labadie, and B. M. Dawant, "Automatic segmentation of the facial nerve and chorda tympani in pediatric CT scans," *Med. Phys.*, vol. 38, no. 10, pp. 5590-5600, Oct. 2011.

[6] J. H. Noble, B. M. Dawant, F. M. Warren, and R. F. Labadie, "Automatic identification and 3D rendering of temporal bone anatomy," *Otol. Neurotol.*, vol. 30, no. 4, pp. 436-442, Jun. 2009.

[7] J. H. Noble, R. F. Labadie, O. Majdani, and B. M. Dawant, "Automatic segmentation of Intracochlear anatomy in conventional CT," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2625-2632, Sept. 2011.

[8] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35-51, Mar. 1996.

[9] F. Maes, A. Collignon, D. Vandermeulen, G. Mrchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging*, vol. 16, pp. 187-198, Apr. 1997.

[10] M. Y. Wang, C. R. Maurer, J. M. Fitzpatrick, R. J. Maciunas, "An automatic technique for finding and localizing externally attached markers in CT and MR volume images of the head," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 6, pp. 627-637, Jun. 1996.

[11] F. A. Reda, B. M. Dawant, R. F. Labadie, and J. H. Noble, "Automatic pre- to intra-operative CT registration for image guided cochlear implant surgery," *In Proc. SPIE Conf. Med. Img.*, vol. 8316, pp. 83161E-83161E-10, Feb. 2012.

[12] A. E. Johnson, M. Hebert, "Surface matching for object recognition in complex three-dimensional scenes," *Image and Vision Computing*, vol. 16, no. 9–10, pp. 635-651, Jul. 1998.

[13] J. Sethian, *Level Set Methods and Fast Marching Methods*, 2nd ed. Cambridge University Press, Cambridge, 1999.

[14] M. Descoteaux, M. Audette, K. Chinzei, K. Siddiqui, "Bone enhancement filtering: Application to sinus bone segmentation and simulation of pituitary surgery," In: Duncan,

J.S., Gerig, G. (eds.) MICCAI 2005, Part I. LNCS, vol. 3749, pp. 9-16. Speinger, Heidelberg (2005).

[15] S. S. Reddi, S. F. Rudin, and H. R. Keshavan, “Optimal multiple threshold scheme for image segmentation,” *IEEE Trans. Syst., Man, Cybern.*, vol. 14, no. 4, pp. 661-665, 1984.

[16] P. Alliez, D. Cohen-Steiner, O. Devillers, B. Ležvy, and M. Desbrun, “Anisotropic Polygonal Remeshing,” *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 485-493, 2003.

[17] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 698-700, Sept. 1987.

[18] F. A. Reda, J. H. Noble, R. F. Labadie, and B. M. Dawant, “Fully-automatic pre-to intra-operative CT registration for cochlear implant,” In: B. M. Dawant et al. (eds.) WBIR 2012, LNCS, vol. 7359, pp. 89-98, Springer-Verlag Berlin Heidelberg, July 2012.

CHAPTER IV

AUTOMATIC SEGMENTATION OF INTRA-COCHLEAR ANATOMY IN POST-IMPLANTATION CT OF UNILATERAL COCHLEAR IMPLANT RECEIPIENTS

Fitsum A. Reda[1], Theodore M. McRackan[2], Robert F. Labadie[2], Benoit M. Dawant[1], and Jack H. Noble[1]

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235, USA

[2]Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University Medical Center, Nashville, TN 37232 USA.

[This chapter has been published in *Medical Image Analysis*, Volume 18, pp. 605-615, February 2014.]

Abstract

A cochlear implant (CI) is a neural prosthetic device that restores hearing by directly stimulating the auditory nerves using an electrode array that is implanted in the cochlea. In CI surgery, the surgeon accesses the cochlea and makes an opening where he/she inserts the electrode array blind to internal structures of the cochlea. Because of this, the final position of the electrode array relative to intra-cochlear anatomy is generally unknown. We have recently developed an approach for determining electrode array position relative to intra-cochlear anatomy using a pre- and a post-implantation CT. The approach is to segment the intra-cochlear anatomy in the pre-implantation CT, localize the electrodes in the post-implantation CT, and register the two CTs to determine relative electrode array position information. Currently, we are using this approach to develop a CI programming technique that uses patient-specific spatial information to create patient-customized sound processing strategies. However, this technique cannot be used for many CI users because it requires a pre-implantation CT that is not always acquired prior to implantation. In this study, we propose a method for automatic segmentation of intra-cochlear anatomy in post-implantation CT of unilateral recipients, thus eliminating the need for pre-implantation CTs in this population. The method is to segment the intra-cochlear anatomy in the implanted ear using information extracted from the normal contralateral ear and to exploit the intra-subject symmetry in cochlear anatomy across ears. To validate our method, we performed experiments on 30 ears for which both a pre- and a post-implantation CT are available. The mean and the maximum segmentation errors are 0.224 and 0.734 mm, respectively. These results indicate that our automatic segmentation method is accurate enough for developing patient-customized CI sound processing strategies for unilateral CI recipients using a post-implantation CT alone.

4.1. Introduction

A cochlear implant (CI) is a neural prosthetic device that restores hearing by directly stimulating the auditory nerve using an electrode array that is surgically implanted in the cochlea [1]. An external sound processor, typically worn behind the ear, processes sounds detected by a microphone into signals sent to the implanted electrodes. The CI sound processor is programmed after implantation by an audiologist. Based on patient response, the audiologist determines stimulation levels for each electrode and selects a frequency allocation table to define which electrodes should be activated when specific sound frequencies are detected [2]. The number of electrodes in a CI electrode array ranges from 12 to 22, depending on the manufacturer.

CI electrode arrays are designed such that when optimally placed in the scala tympani cavity of the cochlea, each electrode stimulates regions of the auditory nerve corresponding to a pre-defined frequency bandwidth [3]. However, because the surgeon threads the electrode array blind to internal cavities of the cochlea during the surgery, the final position of the electrode array relative to intra-cochlear anatomy is generally unknown. Research has shown that in 73% of CI surgeries the electrode array is placed fully within the scala tympani, while in the other 27% of CI surgeries, the electrode array is fully within a neighboring cavity or is initially inserted into the scala tympani but crosses into a neighboring cavity [4]. So far, the only option when programming the CI has been to assume the array is optimally placed in the cochlea and to use a default frequency allocation table. Currently, we are developing Image-Guided CI Programming (IGCIP) strategies that rely on patient-specific knowledge of the position of the electrodes relative to intra-cochlear anatomy, and we have shown that IGCIP strategies can drastically improve hearing outcomes [5]. IGCIP strategies are enabled by a number of algorithms we have developed that permit determining the spatial relationship between intra-

cochlear anatomy and the CI electrodes using a pre-implantation and a post-implantation CT [6]-[10]. The intra-cochlear Structures-Of-Interest (SOIs) are the scala tympani (ST), scala vestibuli (SV), and the spiral ganglion (SG), which is the ganglion of auditory nerve bundles. 3D renderings of these structures as well as the implant are shown in Figures 4.1a and 4.1b. Examples of pre- and post-implantation CTs with overlaid structure contours are shown in Figures 4.1c and 4.1d.

The approach we previously developed for determining electrode array position relative to the SOIs involves several steps. First, we segment the SOIs in the pre-implantation CT. Next, we identify the electrode array in the post-implantation CT. Finally, we rigidly register the two CTs to determine the position of the electrodes relative to intra-cochlear anatomy. However, this approach cannot be used for many CI recipients because it requires a pre-implantation CT, which is not always acquired prior to implantation. Thus far, the pre-implantation rather than the post-implantation CT has been used to identify the SOIs because the cochlea is obscured by image artifacts introduced by the metallic electrode array in the post-implantation CT (see Figure 4.1c and 4.1d). In this study, we propose methods to extend our IGCIP strategies to the population of unilateral CI recipients for whom a CT was not acquired pre-operatively, thereby increasing the portion of the population of existing CI recipients who can benefit from IGCIP strategies. The methods we present permit automatic segmentation of the SOIs in the post-implantation CT despite the significant artifacts induced by the CI electrodes in those images.

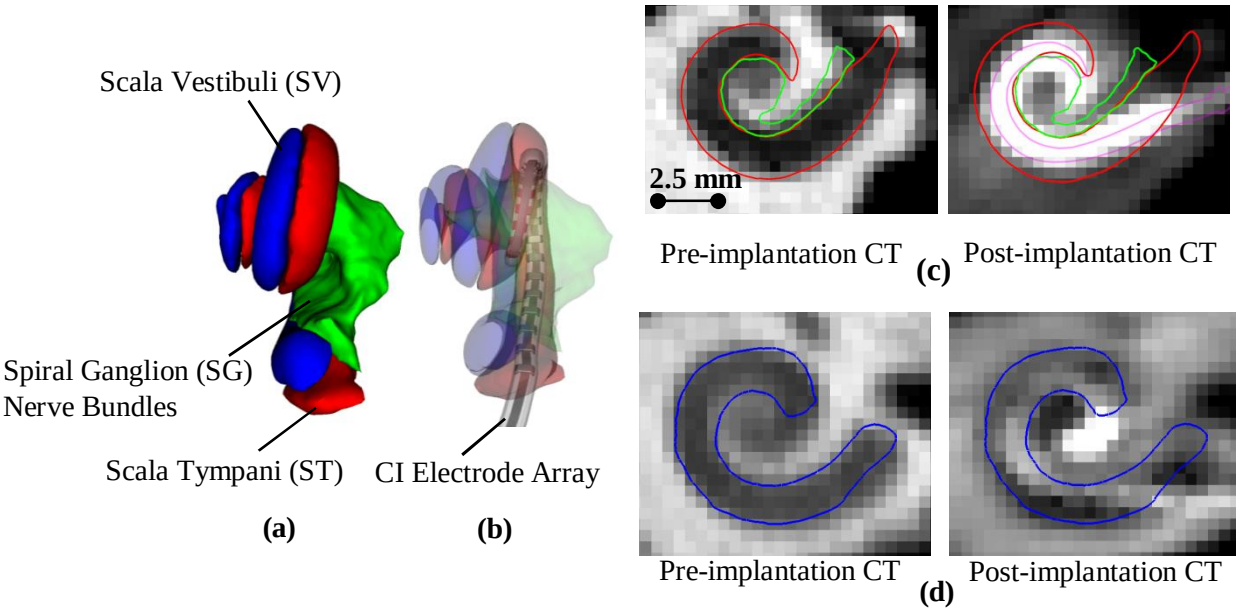


Figure 4.1. Shown in (a) and (b) are surfaces of ST (red), SV (blue), and SG (green). In (b), a surface model of a CI electrode array inserted into ST is shown. In (c), contours of ST (red), SG (green) and the electrodes (purple) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT, and in (d) contours of the SV (blue) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT. The bright structure in the post-implantation CTs is the artifact cause by the CI electrode array.

4.2. Methods

The method we propose for segmenting the intra-cochlear anatomy of unilateral CI recipients takes advantage of the intra-subject inter-ear symmetry we have observed. We acquire a post-implantation CT in which both ears are in the field of view and segment the intra-cochlear anatomy of the implanted ear using information extracted from the normal contralateral ear. That is, we first segment in the normal contralateral ear the ST, SV, SG, which are the SOIs, and the labyrinth. The labyrinth, which we use as a landmark structure, externally bounds the intra-cochlear anatomy and includes the three semicircular canals (see Figure 4.2). Next, we segment the SOIs in the implanted ear by projecting the SOI surfaces from the normal ear to the implanted ear. The transformation we use is the one that rigidly registers the mirrored labyrinth

surface from the normal ear to the labyrinth in the implanted ear. The labyrinth provides adequate landmarks for this registration because a portion of the labyrinth lies far enough from the implant that its image features are not drastically affected by the implanted electrode array and, as we will show, the position of the labyrinth well predicts the position of the SOIs.

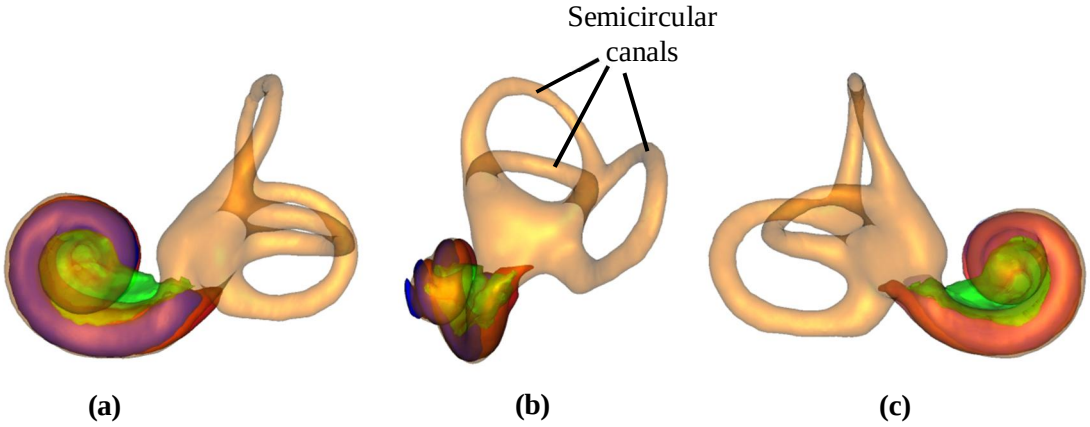


Figure 4.2. Surfaces of the labyrinth (shown in transparent gold) and intra-cochlear anatomy (shown in transparent red (ST), transparent blue (SV), and transparent green (SG)) viewed in three orientations (a), (b), and (c).

Our methods are detailed in the following subsections. In Section 4.2.1, we present details about our datasets. The registration processes we use at several steps throughout our work are detailed in Section 4.2.2. The study we perform to establish inter-ear symmetry of cochlear anatomy is presented in Section 4.2.3. The methods we use to segment both the labyrinth and the intra-cochlear anatomy in the normal ear are detailed in Section 4.2.4. Finally, in Section 4.2.5, we present the method we propose to segment the intra-cochlear anatomy in the implanted ear using information obtained from the normal ear.

4.2.1 Data

Table 4.1 summarizes the characteristics of the various sets of CT scans we have used. Age of subjects included in this study ranged from 18 to 90 years. The scans were acquired from several conventional scanners (GE BrightSpeed, LightSpeed Ultra; Siemens Sensation 16; and Philips

Mx8000 IDT, iCT 128, and Brilliance 64) and a low-dose flat-panel volumetric CT (fpVCT) scanner (Xoran Technologies xCAT® ENT). Conventional CT scans of 10 subjects were used for symmetry analysis as described in Section 4.2.3; conventional CT scans of 18 subjects were used for active shape model (ASM) creation as discussed in Section 4.2.4.1; fpVCT scans of 14 subjects were used for intensity gradient model (IGM) creation as discussed in Section 4.2.5.2; and 18 CT-fpVCT pairs of scans were used for segmentation validation as discussed in section 4.2.5.3. Typical scan resolution for conventional CT scans is 768×768×145 voxels with 0.2×0.2×0.3 mm^3 voxel size. It is 700×700×360 and 0.3×0.3×0.3 or 0.4×0.4×0.4 mm^3 for fpVCT scans.

Table 4.1. Datasets used in this study

Dataset #	Purpose	Dataset Size	Acquisition		# of CIs		
			Xoran fpVCT	Conventional	No CIs	One CI	Two CIs
1	Symmetry analysis	10		×	×		
2	ASM creation	18		×	×		
3	IGM creation	14	×		×		
4	Segmentation validation	6		×	×		
			×			×	
		12		×	×		
			×				×

Dataset 4 is used for segmentation validation. Each implanted ear in the dataset will be automatically segmented in a post-implantation CT using the algorithms proposed in this paper. For each of these ears, there is a pre-implantation CT that is used to generate gold standard segmentations to compare to the automatic segmentations for validation. Dataset 4 consists of two subgroups. The first (6 subjects) includes a set of conventional pre-implantation and low-dose post-implantation CTs of six unilateral CI recipients. The second (12 subjects) includes a set of conventional pre-implantation and low-dose post-implantation CTs of a group of 12 bilateral CI recipients. We use the second set to increase the size of our testing set without

having to scan more unilateral CI recipients prior to demonstrating the efficacy of our technique. To do so, we register the pre- and post-implantation CTs and use the contralateral side of the pre-implantation CT rather than the contralateral side of the post-implantation CT in our algorithm. Using this technique, we have 30 datasets (6 in the first group and 12x2 in the second group) on which we can test our unilateral segmentation algorithm. It should be noted that the second group of 24 ears from 12 subjects is not equivalent to ears from 24 subjects. While left and right ears generally have a different appearance in the post-implantation CT since the electrode array is positioned differently, differences in shape of anatomical structures between ears are not as large as inter-subject variations.

4.2.2. Image Registration Methods

Various processes we describe in the following subsections rely on image-to-image registration. In this subsection, we detail the affine and non-rigid registration methods we use. Given a "fixed" image, i.e., an atlas, and a "floating" image, i.e. the image we want to segment, we use a multi-step process outlined in Figure 4.3 to register them. First, we affinely register the entire but downsampled images. Next, we refine the registration by limiting the region of interest to a pre-defined region that encompasses the ear structure. At this stage, the affine transformation is estimated at full resolution. In both cases, we rely on an intensity-based technique that uses Powell's direction set method and Brent's line search algorithm [11] to optimize the mutual information [12]-[13] between the images. The registration within the region of interest is further refined with a non-rigid registration step using the adaptive-bases algorithm (ABA) [14]. ABA models the deformation field as a linear combination of a set of basis functions irregularly spaced over the image domain,

$$v(\boldsymbol{x}) = \sum_{i=1}^{N} \boldsymbol{c}_i \boldsymbol{\Phi}(\boldsymbol{x} - \boldsymbol{x}_i) \tag{4.1}$$

where $\boldsymbol{x}$ is a point in $\mathbb{R}^d$, with d being the dimensionality of images, the function $\boldsymbol{\Phi}$ is Wu's compactly supported positive definite radial basis function [15], and $\{\boldsymbol{c}_i\}_{i=1}^{N} \in \mathbb{R}^d$ is the set of basis function coefficients that are selected to optimize the normalized mutual information [16] between the images. The optimization process uses a gradient descent algorithm to determine the direction of optimization, and a line minimization algorithm to calculate the optimal step in that direction. The final deformation field is computed using a multiresolution and multiscale approach. Multiresolution is achieved by creating a standard image pyramid, and multiscale is achieved by modifying the region of support and the number of basis functions. A large region of support models a transformation at a large scale. The algorithm is initialized on a low-resolution image with few basis functions. Then, the region of support of the basis functions is reduced as the algorithm progresses to finer resolutions and smaller scales (larger number of basis functions). Using this approach, the final deformation field is computed as

$$v(\boldsymbol{x}) = \sum_{k=1}^{M} \boldsymbol{v}_k(\boldsymbol{x}), \tag{4.2}$$

with M being the total number of combinations of scales and image resolutions used.

1) Affine registration
    a. Affinely register the floating image to the fixed image, after downsampling both images by a factor of four in each dimension.
    b. Crop the ear region from the affinely registered floating image
    c. Affinely register the floating ear region to the fixed ear region image at full image resolution.
2) Non-rigid registration
    a. Non-rigidly register the floating ear region to the fixed ear region image.

Figure 4.3. Image registration process

4.2.3. Symmetry Analysis

To establish that the ST, SV, SG, and the labyrinth are symmetric across ears, we conduct experiments on the set of pre-implantation CTs in dataset 1 (see Table 4.1). We identify surfaces of the ST, SV, SG, and the labyrinth for both ears in each pre-implantation CT using methods that we describe in Section 4.2.4. Then, we register the surfaces of one ear to the corresponding surfaces of the contralateral ear using a standard point-based rigid-body registration method [17]. Finally, we measure distances between the points on each surface to the corresponding points on the registered surface.

4.2.4. Segmentation of the Normal Ear

To segment the ST, SV, and SG in the normal ear, we use an automatic active shape model (ASM)-based method we have developed previously [5]-[6]. The mean and maximum surface errors in segmenting the ST in fpVCTs are 0.18 and 0.9 mm. These are 0.22 and 1.6 mm for the SV, and 0.15 and 1.3 mm for the SG, respectively.

The method we have developed for the automatic segmentation of the labyrinth relies on an active shape model. The following subsections describe how we create the model, how we use

these models for segmentation purposes, and the study we have designed to test the accuracy of our results.

4.2.4.1. Labyrinth Active Shape Segmentation

We create an ASM of the labyrinth using the pre-implantation CTs in dataset 2 (see Table 4.1). We choose one of these pre-implantation CTs to serve as a reference volume, and we use the remaining CTs as training volumes. The active shape model creation process is outlined in Figure 4.4. This process has six main steps. First, the labyrinth is segmented manually in the reference volume by an experienced otolaryngologist (TRM). Second, the training volumes are registered to the reference volume using the multi-step registration techniques described in Section 4.2.2. Third, the labyrinth in each of the training volumes are pre-segmented by projecting the labyrinth surface from the reference volume onto each of the training volumes using the transformations computed in step 2. Fourth, the surfaces produced in step 3 are manually edited to correct for possible segmentation errors caused by mis-registration. These four steps produce both segmented surfaces and a one-to-one point correspondence between points on the reference surface and points on each of the training surfaces. The procedures described in these four steps are similar to the approach described by [18]. In the fifth step, all the training surfaces are registered to the reference surface with a 7-Degree-Of-Freedom (DOF) transformation (three rotations, three translations, and one isotropic scale) computed with a standard least squares fitting method [17]. We include isotropic scaling as a DOF so that inter-subject differences in labyrinth scale are normalized. Finally, in Step (6), we use eigenanalysis to build the ASM, which is composed of the mean $\bar{\boldsymbol{x}}$ and the eigenvectors $\{\boldsymbol{u}_j\}$ of the covariance matrix X of the registered shapes,

$$\{\lambda_j, \boldsymbol{u}_j\}_{j=0}^{M-2} : \lambda_j \boldsymbol{u}_j = X\boldsymbol{u}_j, \tag{4.3}$$

where M is the number of training shapes and $\{\lambda_j\}$ is the set of eigenvalues [19].

1) Manually segment the labyrinth in a reference volume.
2) Register the training volumes to the reference volume (see Section 2.2)
3) Pre-segment automatically the labyrinth in the training volumes by projecting the reference surface onto the training volumes with the transformations computed in step 2.
4) Manually adjust the surfaces produced automatically
5) Register the labyrinth surfaces in the training volumes to the reference surface

Figure 4.4. Active shape model generation process.

4.2.4.2. Segmentation of the Labyrinth Using the Active Shape Model

Once an ASM of the labyrinth is built, we use it to segment the labyrinth in a target volume using the segmentation process outlined in Figure 4.5. We start by registering the ASM reference volume to the target volume according to the procedure described in Section 4.2.2. We then project the ASM mean surface points onto the target volume and we fit the ASM to these projected points. This produces our initial segmentation that is then refined as follows. For each point on the ASM surface $\boldsymbol{y} = \{\boldsymbol{y}_i\}_{i=0}^{N-1}$ we find a new candidate point $\boldsymbol{y}_i'$ by searching for the point with the highest image intensity gradient within the interval [-1.35, 1.35] mm along the local surface normal $\hat{\boldsymbol{n}}_i$, equivalently,

$$\boldsymbol{y}_i' = \boldsymbol{y}_i + \Delta d \cdot k_{\max} \cdot \hat{\boldsymbol{n}}_i, \text{ where}$$

$$k_{\max} = \arg\max_k \left(I(\boldsymbol{y}_i + \Delta d \cdot (k+1) \cdot \hat{\boldsymbol{n}}_i) - I(\boldsymbol{y}_i + \Delta d \cdot (k-1) \cdot \hat{\boldsymbol{n}}_i) \right), \tag{4.4}$$

$$\text{for } k \in [-9, 9] \text{ and } \Delta d = 0.15 \text{ mm},$$

where $I(\bullet)$ is the image intensity at a given point. The approach of finding a point with the maximum gradient is similar to those investigated by [19]-[24]. Then, we fit the ASM to the new candidate point set $\boldsymbol{y}' = \{\boldsymbol{y}_i'\}_{i=0}^{N-1}$ to obtain an adjusted shape $\boldsymbol{y}'' = \{\boldsymbol{y}_i''\}_{i=0}^{N-1}$. To perform the fitting procedure, we first register $\boldsymbol{y}'$ to the ASM mean shape $\bar{\boldsymbol{x}}$ with a 7-DOF (three rotations, three translations, and one isotropic scale factor) transformation ψ. Then, we compute the adjusted point set $\boldsymbol{y}''$ using the equation

$$\boldsymbol{y}_i'' = \psi^{-1}\left(\bar{\boldsymbol{x}}_i + \sum_{j=0}^{K-1} b_j \boldsymbol{u}_{j,i}\right), \tag{4.5}$$

with K being the number of eigenshapes used, where

$$b_j = \boldsymbol{u}_j^T(\psi(\boldsymbol{y}') - \bar{\boldsymbol{x}}). \tag{4.6}$$

The magnitude of $\{b_j\}_{j=0}^{K-1}$ is chosen such that the Mahalanobis distance from the adjusted shape to the mean shape is less than 3:

$$\sqrt{\sum_{j=0}^{K-1}\left(\frac{b_j^2}{\lambda_j}\right)} \leq 3.$$

We iterate the adjustment step until the constraint $1/N\sum_{i=0}^{N-1}\|\boldsymbol{y}_i'' - \boldsymbol{y}_i\| < \varepsilon$ is satisfied, where N is the number of points, and ε is empirically set to 0.01 mm.

1) Automatically initialize shape
    a. Register reference volume to target volume
    b. Project ASM mean shape to the target volume space and fit the ASM to the projected points
2) Adjust shape
    a. Find a candidate point along each point's normal
    b. Fit the ASM to the candidate point set

Figure 4.5. Active shape segmentation process.

4.2.4.3. Labyrinth Segmentation Validation

To validate our labyrinth segmentation method, we fix the reference volume and use the method presented above to segment the remaining 17 training volumes in a leave-one-out approach. We measure distance between corresponding points on the automatic and manually generated surfaces to quantitatively evaluate the agreement between the two. Specifically, for each point on the automatic surface we measure the Euclidean distance to the corresponding point on the manual surface. Then, for each training volume, we measure the mean and maximum of these distances.

4.2.5. Segmentation of the Normal Ear

The process we use to segment the intra-cochlear anatomy in an implanted ear is outlined in Figure 4.6. In this process, we do not identify the intra-cochlear anatomy in the implanted ear directly. Rather, we identify the position of the labyrinth in the implanted ear and use it as a landmark structure to determine the position of the intra-cochlear anatomy. First, we estimate an initial position of the labyrinth in the implanted ear using a procedure described in Section 4.2.5.1. Next, we iteratively refine this estimation of the labyrinth position using a procedure described in Section 4.2.5.2. Finally, we determine the intra-cochlear anatomy in the implanted

ear by projecting the intra-cochlear surfaces segmented in the normal ear through the transformation that rigidly registers the labyrinth from the normal ear to the iteratively refined labyrinth in the implanted ear. The following subsections detail this approach.

1) Automatically initialize the position of the labyrinth
    a. Rigidly register the normal ear to the implanted ear
    b. Project the labyrinth surface from the normal ear to the implanted ear
2) Adjust the labyrinth surface
    a. Find a candidate point along each point's normal
    b. Assign a weight to each point
    c. Rigidly register initial point set to candidate point set
3) Iterate (2) until convergence
4) Project intra-cochlear surfaces from normal ear to implanted ear

Figure 4.6. Iterative intra-cochlear anatomy segmentation process.

4.2.5.1. Segmentation Initialization via Image-to-Image Registration

To estimate an initial position of the labyrinth in the implanted ear, we project the labyrinth surface from the normal contralateral ear to the implanted ear. The transformation we use for projecting the labyrinth is the transformation that registers the normal ear to the implanted ear. Figure 4.7 lists the steps we use to compute this transformation. We start by estimating a mirroring transformation through registration to a volume in which the mid-sagittal plane has been defined. Several approaches exist in the literature for accurate mid-sagittal extraction in MR images as well as other modalities [25]-[29]. The approach we have selected, while likely not as accurate as these dedicated methods, requires little extra processing because registration with a reference is already performed prior to segmentation and provides an estimation of the mirroring transformation that is accurate enough to initialize a subsequent refinement step. This is achieved by computing a rigid body transformation with an intensity-based method applied first to the

entire but downsampled images then to a region of interest but at full resolution. This is similar to the process we use in the first step of the process described in Figure 4.3.

1) Initialize a mirroring (inter-ear registration) transformation
    a. Rigidly register the target image to an atlas image where the mid-sagittal plane is pre-defined.
    b. Mirror the target image along the pre-defined plane.
    c. Project the mirrored image back onto the original target image space
2) Refine the mirroring transformation
3) Project the labyrinth surface from the normal ear to the implanted ear.

Figure 4.7. Inter-ear registration process.

4.2.5.2. Segmentation Refinement via Surface-to-Image Registration

To refine the position of the labyrinth, we iteratively adjust its position. We perform this by iteratively finding candidate positions for each point $\boldsymbol{y}_i$ on the labyrinth surface and rigidly registering the surface to those candidate points. This is similar to the iterative closest point surface registration algorithm introduced by [30]. At each iteration, we choose the candidate position $\boldsymbol{y}_i'$ for each point $\boldsymbol{y}_i$ as

$$\boldsymbol{y}_i' = \boldsymbol{y}_i + \Delta d \cdot k_{\min} \cdot \hat{\boldsymbol{n}}_i, \tag{4.7}$$

where $\Delta d = 0.15$mm, and we choose $k_{\min}$ to minimize the cost function

$$k_{\min} = \arg\min_k C_i(k) \; : \; k \in [-9,9]. \tag{4.8}$$

The cost function $C_i(\bullet)$ we have designed for candidate selection at each ith point is a function of an intensity-gradient model (IGM) of the image at that point. To build the IGM, we rely on a set of manual segmentations of the labyrinth in dataset 3 (see Table 4.1). For each ith point on the jth training surface, $\{\boldsymbol{x}_{ji}\}_{i=0}^{N-1}$, we extract an intensity-gradient profile $\boldsymbol{g}(\boldsymbol{x}_{ji})$ along the local surface normal $\hat{\boldsymbol{n}}_{ji}$ using the equation

$$\boldsymbol{g}(\boldsymbol{x}_{ji}) = \left[\left(I_j^{-10}(\boldsymbol{x}_{ji}) - I_j^{-8}(\boldsymbol{x}_{ji})\right), \left(I_j^{-9}(\boldsymbol{x}_{ji}) - I_j^{-7}(\boldsymbol{x}_{ji})\right), \ldots, \left(I_j^{8}(\boldsymbol{x}_{ji}) - I_j^{10}(\boldsymbol{x}_{ji})\right)\right]^T, \quad (4.9)$$

where

$$I_j^k(\boldsymbol{x}_{ji}) = I_j(\boldsymbol{x}_{ji} + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_{ji}), \quad (4.10)$$

$\Delta d = 0.15$ mm, and $I_j(\bullet)$ is the intensity of the jth training image at a given point. Finally, we define the IGM as the set of $\{\boldsymbol{g}(\boldsymbol{x}_{ji})\}_{i=0}^{N-1}$ for $j \in [0, 1, \ldots, M-1]$, where N is the number of points composing each training surface, and M is the number of training surfaces.

The cost we use for candidate point selection in Eqn. (4.8) above is then designed as

$$C_i(k) = \min_{j \in [0,1,\ldots,M-1]} \left\| \boldsymbol{g}(\boldsymbol{y}_i + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_i) - \boldsymbol{g}(\boldsymbol{x}_{ji}) \right\|, \quad (4.11)$$

which defines the cost for selecting $\boldsymbol{y}_i + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_i$ as a new candidate position for the ith point as the minimum Euclidean distance between the set of intensity-gradient profiles in the IGM and the intensity-gradient profile measured at that point. The standard approach is to compute the mean profile as well as the covariance of the profiles and determine candidate points by minimizing the Mahalanobis or Euclidean distance to the mean profile [19], [31]-[36]. However, pilot experiments we conducted indicated that our approach leads to superior final segmentation accuracy.

Finally, we compute the rigid body transformation T that registers the initial point set $\{\boldsymbol{y}_i\}$ to the candidate point set $\{\boldsymbol{y}_i'\}$ determined using Eqn. (4.7) using a weighted least-squares approach [37], formulated as

$$\arg\min_T \sum_{i=0}^{N-1} w_i^2 \|(T(\boldsymbol{y}_i) - \boldsymbol{y}_i')\|^2. \quad (4.12)$$

$\{w_i\}$ is a set of reliability weights that we assign to points using image intensity information derived from the images. Because the implant is very bright in the CT images, it obscures structure boundaries. Points that are near high intensity regions are thus assigned low weight values and points away from bright regions are assigned high weight values. To compute the weight values, we analyze the intensity distribution of the image over a subset of labyrinth boundary points that are known *a priori* to lie far away from the electrode (region shown in blue in Figure 4.8) and use this information to create a weight function that estimates the likelihood that each labyrinth surface point is located near an electrode. To do this, we first extract intensity profiles $\boldsymbol{r}(\boldsymbol{z}_i)$ at each ith point in the subset of surface points $\{\boldsymbol{z}_i\}_{i=0}^{N'-1} \subset \{\boldsymbol{y}_i\}_{i=0}^{N-1}$ that should lie far from the electrodes and is shown in blue in Figure 4.8, using the equation

$$\boldsymbol{r}(\boldsymbol{z}_i) = [I^{-10}(\boldsymbol{z}_i), I^{-9}(\boldsymbol{z}_i), \quad \cdots \quad , I^{10}(\boldsymbol{z}_i)]^T, \tag{4.13}$$

where

$$I^k(\boldsymbol{z}_i) = I(\boldsymbol{z}_i + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_{\boldsymbol{i}}), \tag{4.14}$$

$\Delta d = 0.15$ mm, and $I(\bullet)$ is the intensity of the target image at a given point. Then, at each iteration of the registration process, we compute a weight for each point as

$$w_i = \begin{cases} 1.0 & \max(\boldsymbol{r}(\boldsymbol{y}_i)) < R \\ e^{-\frac{(\max(r(y_i))-R)^2}{2\sigma^2}} & \max(\boldsymbol{r}(\boldsymbol{y}_i)) \geq R \end{cases}, \tag{4.15}$$

where the value of R is experimentally determined (see details below) to be the 68th percentile of the distribution of the maximum values of $\boldsymbol{r}(\boldsymbol{z}_i)$ measured at $\{\boldsymbol{z}_{\boldsymbol{i}}\}_{i=0}^{N'-1}$, and σ is the standard deviation of the same distribution, computed as

$$\sigma = \left(\frac{1}{N'}\sum_{i=0}^{N'-1}(\max(\boldsymbol{r}(\boldsymbol{z}_i)) - \mu)^2\right)^{1/2}.$$

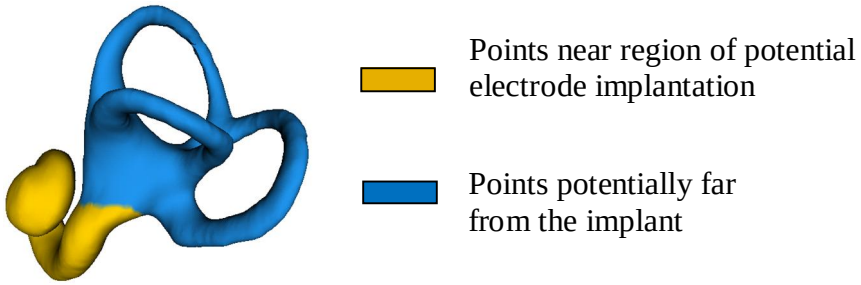


Figure 4.8. Points shown in blue are the points we use for computing R, the main parameter in our weight function. The remaining points of the labyrinth surface are shown in yellow.

The weight function in Eqn. (4.15) is designed such that a weight value of 1.0 is assigned to points with a maximum value in their intensity profile less than R, and weight values that exponentially decay from 1.0 are assigned to points with a maximum value in their intensity profile greater than R. By assigning weights in this way, we achieve our goal, which is to constrain the registration to rely more on points located in regions that are less affected by the image artifacts produced by the implant.

The value of R, as defined above, is customized for each target image because the intensity distribution in the images generated by the low-dose scanner used in this study vary across patients. To arrive at the value of R, we use the set of testing image pairs in dataset 4. First, we sample R in increments of 4% percentiles in the distribution of maximum values of intensities $\{\max(\boldsymbol{r}(\boldsymbol{z}_i))\}_{i=0}^{N'}$, and we measure the resulting SOI segmentation error on all testing image pairs. Next, we select the value of R as the value for which the overall mean segmentation error is the smallest. Figure 4.9 shows a plot of the overall mean error for $R \in \left[\min\left(\{\max(\boldsymbol{r}(\boldsymbol{z}_i))\}_{i=0}^{N'}\right), \max\left(\{\max(\boldsymbol{r}(\boldsymbol{z}_i))\}_{i=0}^{N'}\right)\right]$ in increments of 4 percentiles. As shown in this figure, R = 68th percentile leads to the smallest segmentation error. This is the value, computed for each volume that is used to produce the results presented herein.

We iterate the surface-to-image registration step formulated in Eqn. (4.12) until $1/N \sum_{i=0}^{N-1} \|T(\boldsymbol{y}_i) - \boldsymbol{y}_i\| < \varepsilon$ is satisfied, where ε is empirically set to 0.01 mm. In summary, at each iteration, we determine candidate points $\{\boldsymbol{y}_i'\}$ using Eqn. (4.7), we compute the weights $\{w_i\}$ using Eqn. (4.15), and we register the initial points $\{\boldsymbol{y}_i\}$ to the candidate points $\{\boldsymbol{y}_i'\}$ using Eqn. (4.12). Finally, we segment the intra-cochlear anatomy in the implanted ear by projecting the intra-cochlear surfaces from the normal ear to the implanted ear through the iteratively refined inter-ear labyrinth registration transformation.

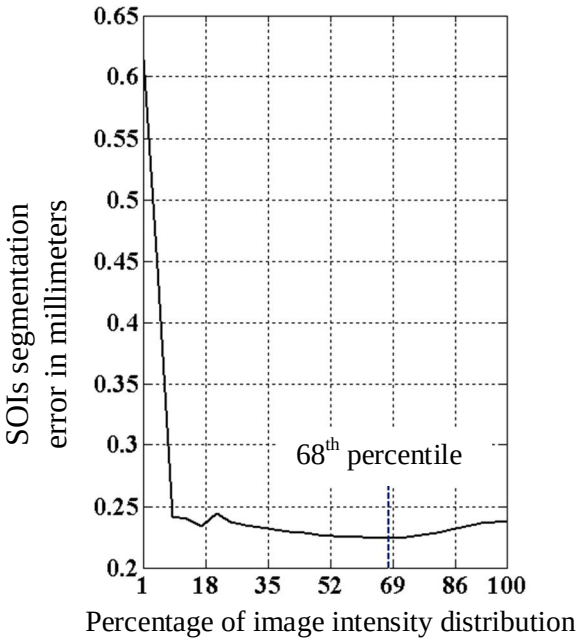


Figure 4.9. Mean error in the SOIs versus selection of *R* as a function of image intensity.

4.2.5.3. Validation

We validate our method by automatically segmenting the post-implantation volumes in dataset 4 and measuring the resulting segmentation errors. The gold-standard surfaces that we use for comparison were created in the pre-implantation volumes by manually editing surface points on segmentations that are automatically initialized by our pre-implantation CT

segmentation techniques. We measure distances between corresponding points on the automatic and gold standard surfaces to quantitatively evaluate the agreement between the two. Specifically, for each point on the automatic surface we measure the distance to the corresponding point on the gold standard surface. Then, for each volume, we measure the mean and maximum of the distances between all corresponding points. To assess how much improvement our proposed surface-to-image registration refinement step provides, we also measure segmentation errors achieved at initialization prior to performing the refinement step, which are the results that can be achieved using image registration alone. Finally we compare our technique to the best possible segmentation results that could be achieved using our contralateral ear registration method. These are obtained by registering directly the labyrinth surface in the normal ear to the gold standard labyrinth surface in the contralateral ear extracted from the pre-implantation CT with a 6-DOF (three rotations and three translations) point-based registration method that minimizes the RMS error between the two surfaces [17].

4.3. Results

4.3.1. Intra-cochlear Anatomy and Labyrinth Symmetry

To demonstrate intra-subject inter-cochlear symmetry, we measured the distance between the ST, SV, and SG surfaces rigidly registered across ears in dataset 1. These measurements are presented in Table 4.2. Figure 4.10 shows the ST, SV, and SG surfaces from one ear colormapped with the distance to the registered contralateral surface for subject one. These distance values are smaller than the segmentation error for these structures as reported in [5]-[6]. Distance maxima are located in the same areas segmentation error maxima occur, i.e., at both the

apical and basal ends of the cochlea. Segmentation errors occur at these locations due to the relative scarcity of local information available in the CT image to estimate the location of the intra-cochlear structures in these regions. This suggests that the small differences between the registered contralateral segmentations seen in Figure 4.10 are most likely due to segmentation error, and that the intra-cochlear anatomy is indeed highly symmetric.

Table 4.2. Distances in millimeter between rigidly registered ST, SV and SG surfaces

Subjects	Scala Tympani (ST)		Scala Vestibuli (SV)		Spiral Ganglion (SG)	
	Mean	Maximum	Mean	Maximum	Mean	Maximum
1	0.099	0.287	0.088	0.243	0.092	0.350
2	0.051	0.159	0.054	0.108	0.064	0.159
3	0.019	0.071	0.018	0.054	0.030	0.113
4	0.049	0.121	0.044	0.133	0.046	0.118
5	0.059	0.160	0.059	0.161	0.063	0.245
6	0.063	0.144	0.055	0.155	0.073	0.212
7	0.087	0.328	0.064	0.164	0.065	0.162
8	0.049	0.115	0.045	0.119	0.067	0.193
9	0.055	0.139	0.050	0.142	0.049	0.160
10	0.058	0.176	0.058	0.140	0.068	0.172
Overall	0.059	0.328	0.054	0.243	0.062	0.350

We performed similar experiments to demonstrate the existence of intra-subject symmetry in labyrinth anatomy. We measured the distance between the labyrinth surfaces rigidly registered across ears in dataset 1. These measurements are presented in Table 4.3, and they are smaller than the labyrinth segmentation error reported in Section 4.3.2. These results suggest that the labyrinth is also highly symmetric.

Table 4.3. Distances in millimeter between rigidly registered labyrinth surfaces.

Subject	1	2	3	4	5	6	7	8	9	10	Overall
Mean (mm)	0.100	0.064	0.082	0.073	0.051	0.094	0.071	0.053	0.120	0.039	0.075
Maximum (mm)	0.239	0.226	0.276	0.264	0.171	0.348	0.329	0.185	0.320	0.140	0.348

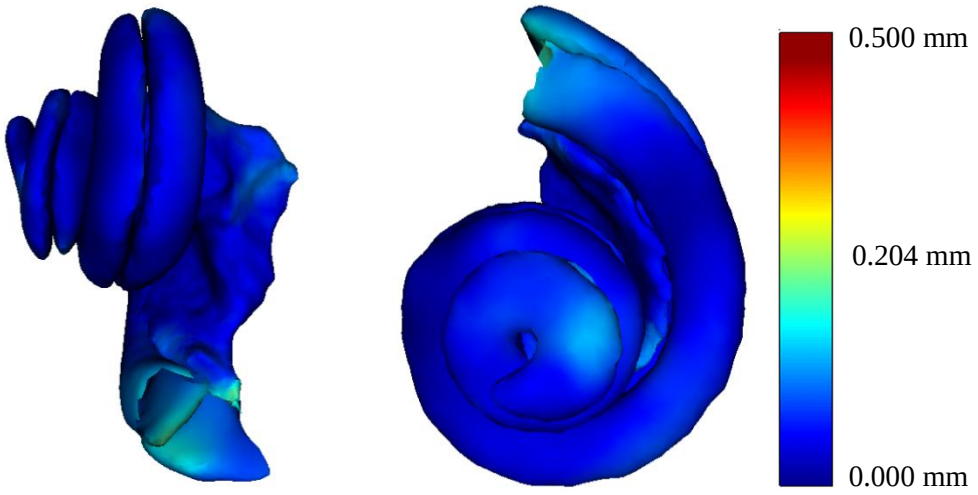


Figure 4.10. Subject one's ST, SV, and SG surfaces viewed in two different orientations. The color at each point encodes the distance in mm to the corresponding point on the registered contralateral surfaces.

4.3.2. Labyrinth Segmentation in the Normal Ear

We built the ASM of the labyrinth using 18 pre-implantation CTs (see dataset 2 in Table 4.1). A total of 9100 points compose each labyrinth shape. Table 4.4 presents the cumulative variations in percentage for the first nine principal components (eigenshapes). As shown in the table, the first nine principal components (eigenshapes) capture 90% of the shape variation in the training set. We used these eigenshapes in the ASM segmentation process. Previous studies suggest that the cochlea is fully formed at birth, and its size and shape does not change as an individual ages [38]. Thus, it is of note that the models we built are applicable for all age groups.

Table 4.4. Percent of labyrinth shape variations captured by the principal components of the shapes used for building the ASM of the labyrinth.

Principal component	1	2	3	4	5	6	7	8	9
Cummulated Variations %	25.07	42.98	55.02	66.64	74.31	78.95	83.21	87.08	90.00

Table 4.5 presents the mean and maximum errors measured as the distance from each point on the automatically generated surface to the corresponding point on the manually generated surface. To illustrate the segmentation improvement provided by the ASM-based segmentation

method, we also show errors for surfaces generated using an atlas-based segmentation method. In this approach we simply project the reference shape onto the target volume using the transformation that registers the reference volume to the target volume. The overall mean and maximum errors for the ASM-based segmentation method are 0.239 and 1.623 mm, respectively. These are 0.452 and 2.407 mm for the atlas-based method. The mean and maximum errors for the ASM-based method are smaller than the atlas-based method for all subjects.

Table 4.5. Mean and maximum labyrinth segmentation errors in mm for both atlas-based and ASM-based methods.

		Subjects								
		1	2	3	4	5	6	7	8	9
Atlas-based Method	Mean	0.442	0.408	0.395	0.370	0.458	0.414	0.534	0.616	0.489
	Maximum	2.153	2.156	1.757	1.450	2.140	2.009	2.407	2.299	1.536
ASM-based Method	Mean	0.254	0.335	0.236	0.280	0.240	0.256	0.174	0.232	0.238
	Maximum	0.963	1.623	1.017	0.679	1.123	1.223	0.664	1.398	0.993

		Subjects								
		10	11	12	13	14	15	16	17	Overall
Atlas-based Method	Mean	0.483	0.451	0.376	0.494	0.445	0.385	0.427	0.495	0.452
	Maximum	2.366	1.516	1.358	1.978	1.883	1.456	1.819	1.574	2.407
ASM-based Method	Mean	0.162	0.286	0.239	0.231	0.336	0.273	0.132	0.160	0.239
	Maximum	0.663	1.145	1.391	0.912	1.078	0.975	0.917	0.904	1.623

In Figure 4.11, we show renderings of the surfaces automatically segmented using both the ASM-based and atlas-based segmentation methods. These surfaces are colormapped with the segmentation error. The top row in this figure shows the labyrinth of the subject with the smallest mean error (subject 16), and the bottom row shows the labyrinth with the largest mean error (subject 2). As can be seen, the surfaces generated using the atlas-based method have unnatural deformations, whereas the surfaces generated using the ASM-based method are smooth and resemble, as expected, the structure surfaces included in the ASM. As can also be

seen in the figure, the mean errors for the ASM-based method are sub-millimetric over most of the labyrinth surface.

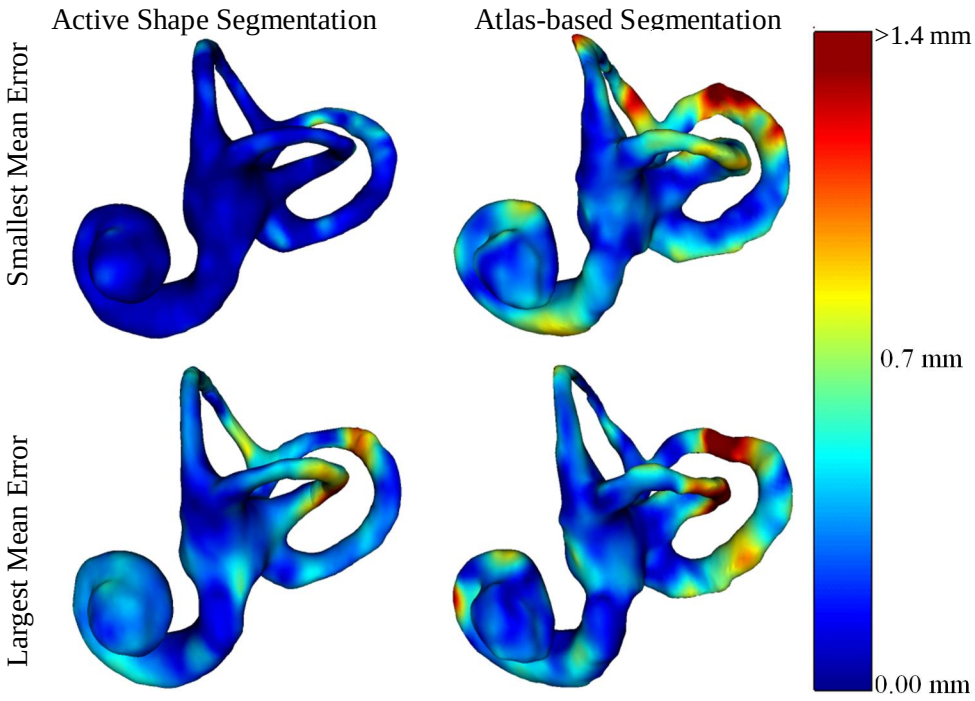


Figure 4.11. Automatically generated surfaces colormapped with errors in mm for subject 16 (top row) and subject 2 (bottom row). Left, surface of the labyrinth generated by the ASM-based method; right surface of the labyrinth generated by the atlas-based method.

4.3.3. Intra-cochlear Anatomy Segmentation in the Normal Ear

We compared quantitatively the gold-standard and automatically generated ST, SV, and SG surfaces for the 30 post-implantation ears in dataset 4. Figure 4.12a shows box plots of the mean error for each SOI, and Figure 4.12b shows the box plots of the maximum error for each SOI. In each box plot, the lower and upper bounds are the minimum and maximum values, respectively, the lower and upper whiskers are the first and third quartiles, respectively, and the red line is the second quartile or the median value. The overall mean and maximum errors for the proposed segmentation method are 0.224 and 0.734 mm, respectively. These results are comparable to those obtained by segmenting the SOIs in pre-implantation CT images using the methods

described in [5]-[6]. In Figure 4.13, we present qualitative results for the subject with the smallest maximum segmentation error (shown in green square on the box plots in Figure 4.12) and for the subject with the largest maximum segmentation error (shown in red square on the box plots in Figure 4.12). Gold standard contours are shown in red, blue, and green for ST, SV, and SG, respectively, and automatically generated contours are in yellow. For the subject with the smallest maximum error, there is excellent agreement between the gold-standard and automatic contours along the length of the structures. In the post-implantation CT, even though the structure boundary information is lost due to the presence of the implant, we are able to achieve sub-millimetric segmentation accuracy for all SOIs. For the subject with the largest maximum error, some disagreement between the gold-standard and automatic contours can be seen along the length of the structures. However, as shown in the surface visualization, these errors are still sub-millimetric. Figure 4.13 suggests that a number of voxels in the immediate proximity to the electrode array (bright voxels) do not lie within the segmentation-delineated borders of the scala tympani/scala vestibuli. This is caused by beam hardening and partial volume reconstruction artifacts that make the electrode appear larger in the images than it really is as shown in Figure 4.1.

Figure 4.14a (left) shows the mean and maximum error box plots for all SOIs (pooled ST, SV, and SG) segmented using only the inter-ear image registration-based initialization method described in Section 4.2.5.1. The overall mean and maximum errors are 0.639 and 1.545 mm, respectively. Figure 4.14a (middle) shows the mean and maximum error box plots for all SOIs segmented using the proposed method. As shown in the plots, using the proposed segmentation method leads to a 64.94% and 52.49% reduction in mean and maximum segmentation errors, respectively. Figure 4.14a (right) shows box plots of mean and maximum errors for the best

segmentation results that could be achieved using our proposed method, as described in Section 4.2.5.3. The overall mean and maximum errors are 0.166 and 0.657 mm, respectively. This shows that the segmentation results we achieve in post-implantation CT are close to the best that are achievable, despite the lack of contrast in these images due to artifacts induced by the implanted electrode array. Figure 4.14b shows the same information for the labyrinth.

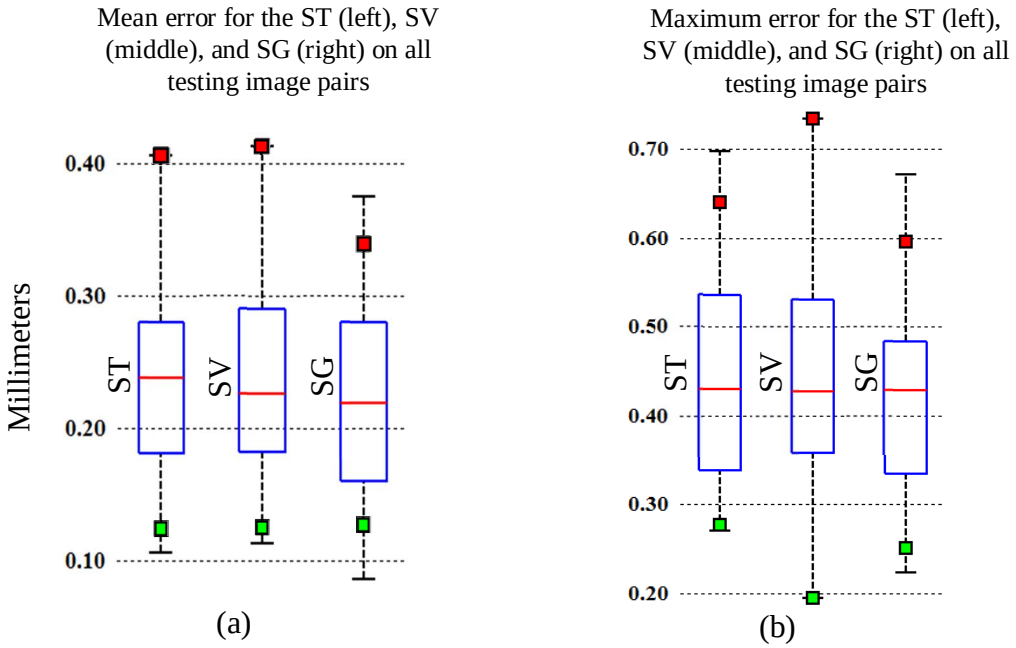


Figure 4.12. Quantitative results for the proposed segmentation method. The green squares on the box plots are quantitative results for the subject with the smallest maximum error, and the red squares are quantitative results for the subject with the largest maximum error.

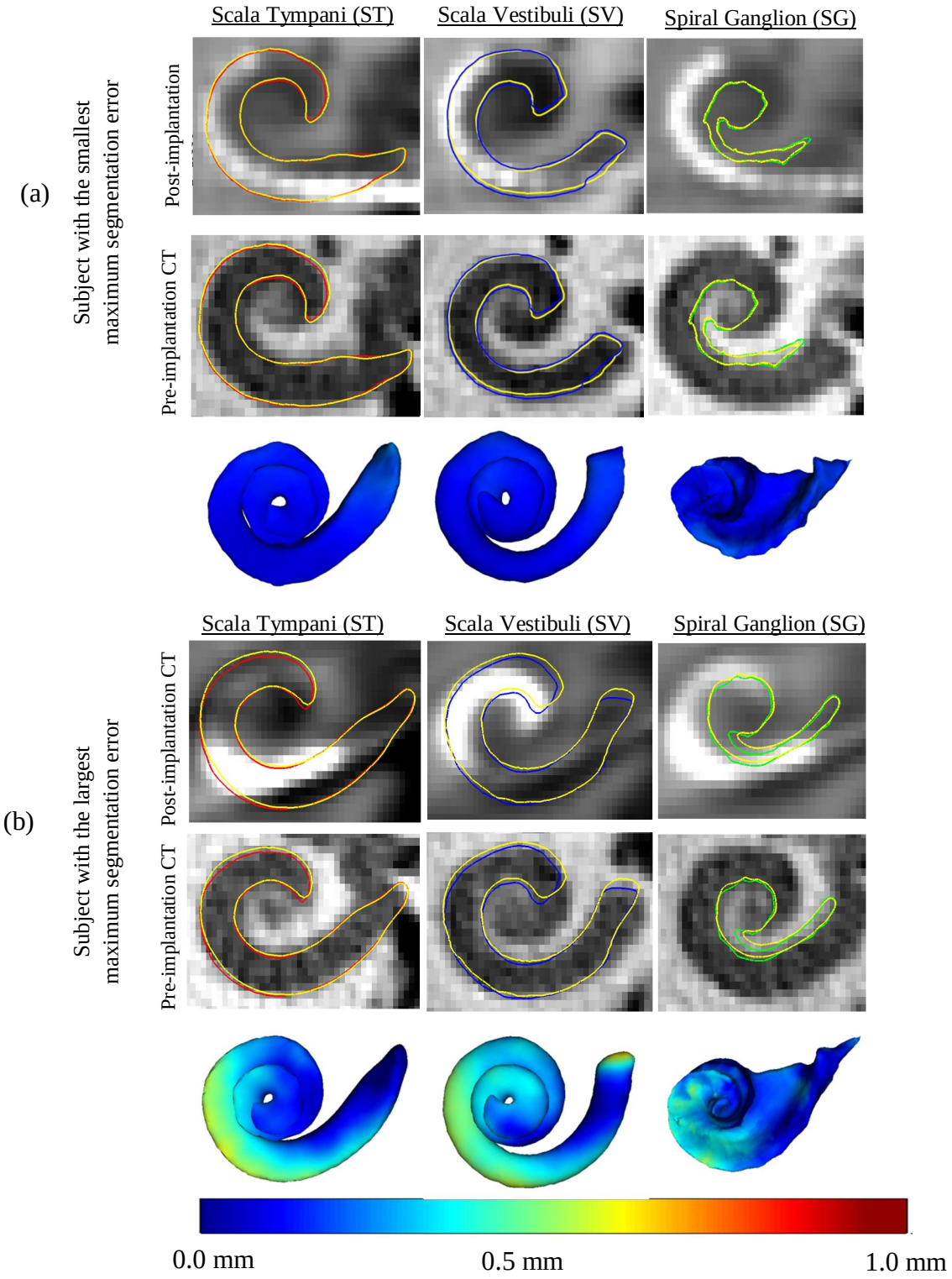


Figure 4.13. Qualitative segmentation results for the case with the smallest maximum segmentation error (shown in green box on Figure 12). The contours shown are the ST (left panel), SV (middle panel), SG (right panel). Structure contours for gold-standard ST (red), gold-standard SV (blue), gold-standard SG (green), and automatic contours for all structures (yellow) are shown in a slice of a post-implantation image (top row) and a corresponding pre-implantation image (middle row). On the bottom panels the structure surfaces colormapped with segmentation errors are shown. (b) Shows similar information for the subject with the largest maximum segmentation error (shown in red box on Figure 12).

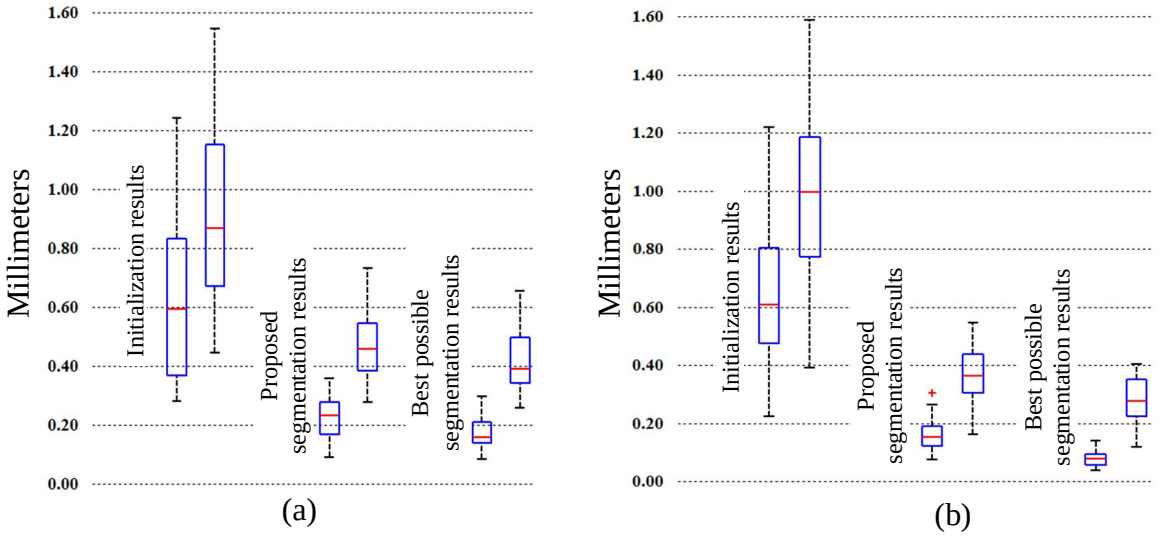


Figure 4.14. Various quantitative segmentation results for all 30 testing post-implantation CTs. (a) Mean and maximum error box plots for the SOIs segmented using the initialization method (left), using the proposed segmentation method (middle). On the right are the mean and maximum error box plots for the best possible SOIs segmentation results. (b) Shows the same information for the labyrinth.

4.4. Conclusions

Image-guided CI programming strategies like the ones we are currently developing require accurate localization of the position of implanted electrodes relative to intra-cochlear anatomy. Until now, it has been possible to segment the SOIs, localize the electrodes, and compute the distance between the electrodes and intra-cochlear anatomy only for CI recipients for whom a pre-implantation CT has been acquired. In this work, we have presented a method that does not require a pre-implantation CT. Our approach is to segment the labyrinth in the contralateral normal ear and use its position to segment the SOIs in the implanted ear by exploiting the intra-subject inter-ear symmetry. We performed symmetry analysis on ten subjects and the results suggest that both the SOIs and the labyrinth are highly symmetric.

To the best of our knowledge, there have been no methods proposed to automatically segment the labyrinth with a high degree of accuracy. We validate the ASM-based method we

propose on 17 ears using a leave-one-out approach. The overall mean and maximum errors are 0.239 and 1.623 mm, respectively. As shown in the SOI segmentation validation study on 30 subjects, this level of accuracy is sufficient to segment the SOIs with sub-millimetric accuracy. In previous work where we relied on a pre-implantation CT and achieved an average SOI segmentation error of 0.15 mm, we have obtained excellent programming results [5]. The segmentation accuracy we achieve with the method presented in this article that does not require a pre-implantation CT is slightly larger (0.22mm) but still small. We anticipate that this slight reduction in segmentation accuracy will not negatively affect the improvement in hearing outcomes we have observed in the more than 60 patients that have participated in our ongoing study (85% of these have reported substantial improvement in hearing). However, we will only be able to confirm this after we have assessed hearing improvements in subjects for whom a pre-operative CT has not been acquired and a programming plan has been created using only a post-implantation CT and the method described herein.

We note that the proposed approach does not permit to identify intra-cochlear anatomy for bilateral implant users for whom a pre-implantation CT has not been acquired. We are currently expanding our segmentation techniques [39] to make it possible and thus provide all cochlear implant subjects access to our image-guided programming method.

4.5. Acknowledgments

This work was supported by NIH grants R01DC008408, R21DC012620, and R01DC010184 from the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of this institute. We would also like to thank the reviewers for their detailed comments and suggestions, which have helped improve the quality of this article.

References

[1]. U.S. Food and Drug Administration PMA, Cochlear implantation, No. 84002446, 21 August 1995.

[2]. Wilson B.S., Finley C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K., Rabinowitz, W.M., “Better speech recognition with cochlear implants,” Nature 352, 236-238, 1991.

[3]. Wilson B.S., Dorman M.F., “Cochlear implants: Current designs and future possibilities,” J. Rehab. Res. Dev. 45(5): 695-730, 2008.

[4]. Aschendorff, A., Kromeier, J., Klenzner, T., and Laszig, R., “Quality control after insertion of the nucleus contour and contour advance electrode in adults,” Ear Hearing, vol. 28, pp. 75S-79S, Apr. 2007.

[5]. Noble, J.H., Labadie, R.F., Gifford, R.H., Dawant, B.M., “Image-guidance enables new methods for customizing cochlear implant stimulation strategies,” Neural Systems and Rehabilitation Engineering, IEEE Transactions on 21(5):820-829, Sept. 2013.

[6]. Noble, J.H., Labadie, R.F., Majdani, O., Dawant, B.M., "Automatic Segmentation of Intracochlear Anatomy in Conventional CT," Biomedical Engineering, IEEE Transactions on, 58(9):2625-2632, Sept. 2011a.

[7]. Jack H. Noble, Theodore A. Schuman, Charles G. Wright, Robert F. Labadie, Benoit M. Dawant, “Automatic identification of cochlear implant electrode arrays for post-operative assessment”, Proc. SPIE 7962(796217), Medical Imaging 2011b.

[8]. Noble, J.H., Gifford, R.H., Labadie, R.F., Dawant, B.M., “Statistical Shape Model Segmentation and Frequency Mapping of Cochlear Implant Stimulation Targets in CT,” MICCAI 2012, 421-428, 2012.

[9]. Schuman T.A., Noble J.H., Wright C.G., Wanna G.B., Dawant B.M., Labadie, R.F.

"Anatomic Verification of a Novel, Non-rigid Registration Method for Precise Intrascalar Localization of Cochlear Implant Electrodes in Adult Human Temporal Bones Using Clinically-available Computerized Tomography," The Laryngoscope, 120 (11): 2277-2283, 2010.

[10]. Wanna, G.B., Noble, J.H., McRackan, T.R., Dawant, B.M., Dietrich, M.S., Watkins, L.D., Rivas, A., Schuman, T.A., Labadie, R.F., "Assessment of electrode positions and audiological outcomes in bilateral cochlear implant patients," Otology & Neurotology, 32(3):428-432, 2011.

[11]. Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. Numerical Recipes in C, 2nd ed. (Cambridge University press, Cambridge, 1992), pp. 412–419.

[12]. Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S. and Kikinis, R. "Multi- modal volume registration by maximization of mutual information," Med. Image Anal., vol. 1, no. 1, pp. 35–51, Mar. 1996.

[13]. Maes, F., Collignon, A., Vandermeulen, D., Mrchal, G., and Suetens, P. "Multimodality image registration by maximization of mutual information," IEEE Trans. Med. Imag., vol. 16, no. 2, pp. 187–198, Apr. 1997.

[14]. Rohde, G.K, Aldroubi, A., and Dawant, B.M. "The adaptive bases algorithm for intensity-based nonrigid image registration," IEEE Trans. Med. Imaging, 22(11), 1470–1479 (2003).

[15]. Wu, Z., "Multivariate compactly supported positive definite radial basis functions," Adv. Comput. Math., vol. 4, pp. 283-292, 1995.

[16]. Studholme, C., Hill, D.L.G., and Hawkes, D.J. "An overlap invariant entropy measure of 3D medical image alignment," Pattern Recognition, 32(1):71-86, 1999.

[17]. Arun, K.S., Huang, T.S., and Blostein, S.D. “Least-squares fitting of two 3-D point sets,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 9, no. 5, pp. 698–700, Sep. 1987.

[18]. Frangi, W.J. Niessen, M.A. Viergever (2001). Three-Dimensional Modeling for Functional Analysis of Cardiac Images: A Review. IEEE Trans on Medical Imaging. 20(1):2-25.

[19]. Cootes, T.F., Taylor, C.J., Cooper, D. H., and Graham, J. “Active shape models - Their training and application,” Comp. Vis. Image Understanding, vol. 61, no. 1, pp. 39–59, 1995.

[20]. Kass, M., Witkin, A., Terzopoulos, D., Snakes: active contour models. International Journal of Computer Vision 1 (4), 321–331, 1988.

[21]. Staib, L.H., Duncan, J.S.,. Boundary finding with parametically deformable models. IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (11), 1061–1075, 1992.

[22]. Cohen, L., Cohen, I., Finite-element methods for active contours models and balloons for 2-D and 3-D images. IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (11), 1131–1147, 1993.

[23]. Chakraborty, A., Staib, L., Duncan, J.S.,. Deformable boundary finding in medical images by integrating gradient and region information. IEEE Transactions on Medical Imaging 15 (6), 859–870, 1996.

[24]. Sakalli, M., Lam, K.-M., Yan, H., A faster converging snake algorithm to locate object boundaries. IEEE Transactions on Image Processing 15 (5), 1182–1191, 2006.

[25]. Ruppert, G.C.S.; Teverovisky, L.; Chen-Ping Yu; Falaco, A.X.; Yanxi Liu, “A new symmetry based method for mid-sagittal plane extraction in neuroimages”, Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, On pages(s):

285-288.

[26]. Prima, S.; Ourselin, S.; Ayache N., “Computation of the mid-sagittal plane in 3-D brain images,” Medical Imaging, IEEE Transactions on, vol. 21, no. 2, pp. 122,148, Feb. 2002.

[27]. Liu, Y., Collins, R. T., and Rothfus, W. E., “Robust midsagittal plane extraction from normal and pathological 3D neuroradiology images,” IEEE Trans. on Medical Imaging, vol. 20, no. 3, pp. 175–192, Mar 2001.

[28]. Tuzikov, A.V., Colliot, O., Bloch, I., “Evaluation of the symmetry plane in 3D MR brain images,” Pattern Recognition Letters, vol. 24, no. 14, pp. 2219–2233, Oct 2003.

[29]. Smith, S. M. and Jenkinson, M., “Accurate robust symmetry estimation,” in Proc MICCAI ’99, London, UK, 1999, pp. 308–317, Springer-Verlag.

[30]. Besl, P.J., McKay, H.D., "A method for registration of 3-D shapes," Pattern Analysis and Machine Intelligence, IEEE Transactions on, 14(2):239-256, 1992.

[31]. Cootes, T.F., Taylor, C.J., Statistical models of appearance for medical image analysis and computer vision. In: Sonka, M., Hanson K.M. (Eds.), Proceedings of the SPIE Medical Imaging, vol. 4322, pp. 236-248, 2001.

[32]. Mitchell, S.C., Lelieveldt, B.P.F., van der Geest, R.J., Bosch, H.G., Reiber, J.H.C., Sonka, M.,. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. IEEE Transactions on Medical Imaging 20(5), 415-423, 2001.

[33]. Heimann, T., Wolf, I., Meinzer, H.-P., Active shape models for a fully automated 3D segmentation of the liver – an evaluation on clinical data. In: Larsen, R., Nielsen, M., Sporring, J. (Eds), Proceeding of MICCAI, LNCS, vol. 4191. Springer, pp. 41-48, 2006.

[34]. Heimann, T., and Meinzer, H-P., “Statistical shape models for 3D medical image

segmentation: A review", Medical Image Analysis, vol. 13(4), pp 543-563, August 2007.

[35]. Brejl, M., Sonka, M., 2000. Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples. IEEE Transactions on Medical Imaging 19 (10), 973–985.

[36]. Tobon-Gomez, C., Butakoff, C., Aguade, S., Sukno, F., Moragas, G., Frangi, A.F., Automatic construction of 3D-ASM intensity models by simulating image acquisition: application to myocardial gated SPECT studies. IEEE Transactions on Medical Imaging 27 (11), 1655–1667, 2008.

[37]. Sonka, M. and Fitzpatrick, J.M. "Medical Image Processing and Analysis," Handbook of Medical Imaging, vol. 2, chapter 8:369-70, 2000.

[38]. Jeffery, N. and Spoor, F., Prenatal growth and development of the modern human labyrinth. Journal of Anatomy, 204: 71–92, 2004.

[39]. Fitsum A. Reda, Jack H. Noble, Robert F. Labadie, and Benoit M. Dawant, "An artifact-robust technique for automatic segmentation of the labyrinth in post-cochlear implantation CT", Proc. SPIE 9034(9034-103), Medical Imaging 2014.

CHAPTER V

AN ARTIFACT-ROBUST, SHAPE LIBRARY-BASED ALGORITHMS FOR AUTOMATIC SEGMENTATION OF INNER EAR ANATOMY IN POST-COCHLEAR-IMPLANTATION CT

Fitsum A. Reda[1], Jack H. Noble[1], Robert F. Labadie[2], and Benoit M. Dawant[1]

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235 USA

[2]Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University Medical Center, Nashville, TN 37232 USA

Abstract

A cochlear implant (CI) is a device that restores hearing using an electrode array that is surgically placed in the cochlea. After placement, the CI is programmed to attempt to optimize hearing outcome. Currently, we are testing an image-guided cochlear implant programming (IGCIP) technique we recently developed that relies on knowledge of relative position of intra-cochlear anatomy to implanted electrodes. IGCIP is enabled by a number of algorithms we developed that permit determining the positions of electrodes relative to intra-cochlear anatomy using a pre- and a post-implantation CT. One issue with this technique is that it cannot be used for many subjects for whom a pre-implantation CT is not available. Pre-implantation CT has been necessary because it is difficult to localize the intra-cochlear structures in post-implantation CTs alone due to the image artifacts that obscure the cochlea. In this chapter, we present algorithms for automatically segmenting intra-cochlear anatomy in post-implantation CTs. Our approach is to first identify the labyrinth and then use its position as a landmark to localize the intra-cochlear anatomy. Specifically, we identify the labyrinth by first approximately estimating its position by mapping a labyrinth surface of another subject that is selected from a library of such surfaces and then refining this estimate by a standard shape model-based segmentation method. We tested our approach on 25 ears and achieved overall mean and maximum segmentation errors of 0.186 and 0.946 mm, respectively. The results we achieve suggest that our approach is accurate enough for developing IGCIP strategies based solely on post-implantation CTs.

5.1. Introduction

A cochlear implant is a device that restores hearing by directly stimulating the auditory nerve using an electrode array that is surgically placed in the cochlea. The CI device includes a sound processor component, typically worn behind the ear, which processes and converts sounds detected by a microphone into electrical signals sent to implanted electrodes. The CI sound processor is programmed by an audiologist who determines a number of processor programming parameters that specify the electrical signals sent to the implanted electrodes to attempt to optimize hearing outcome. The number of electrodes in a CI electrode array range from 12 to 22, depending on the manufacturer.

We recently developed and are currently testing image-guided cochlear implant programming (IGCIP) strategies that rely on patient-specific knowledge of the spatial relationship between implanted electrodes and intra-cochlear anatomy [1]. The intra-cochlear Structures-of-Interest (SOIs) are the scala tympani (ST), the scala vestibuli (SV), and the spiral ganglion (SG). The SG is a region that contains the group of nerves targeted for stimulation by the electrode array. Figure 5.1a shows the surfaces of the ST, SV and the SG. Figure 5.1b shows an example surface model of the electrode array inserted into the cochlea, and Figure 5.1c shows a surface of the active region (AR), which is the interface between (1) the SG and (2) the union of the ST and SV. The AR is the region where electrical activation of nerves is most likely.

IGCIP strategies are enabled by a number of algorithms we have developed that permit determining the positions of the electrodes relative to intra-cochlear anatomy using a pre- and a post-implantation CT [2]-[6]. In a preliminary study with over thirty subjects, we have shown that IGCIP strategies can significantly improve hearing outcomes [1]. One issue with our current technique is that it has not been possible to localize intra-cochlear structures in post-implantation

CTs directly due to the image artifacts caused by the electrode array that obscure the cochlea in the image (see Figure 5.1d and 5.1e). Thus far, the shape of the cochlea and intra-cochlear anatomy has been determined using a pre- and a post-implantation CT. The approach we previously developed for this consists of three steps. First, we segment the SOIs in the pre-implantation CT. Next, we localize the electrodes in the post-implantation CT. Finally, we register the two CTs to determine the relative position of the electrodes relative to intra-cochlear anatomy. For unilateral CI recipients where only a post-implantation CT with both ears in the field of view (FOV) is available, we have also developed a technique that permits estimating the shape of the cochlea in the implanted ear using information from the contralateral normal ear [7].

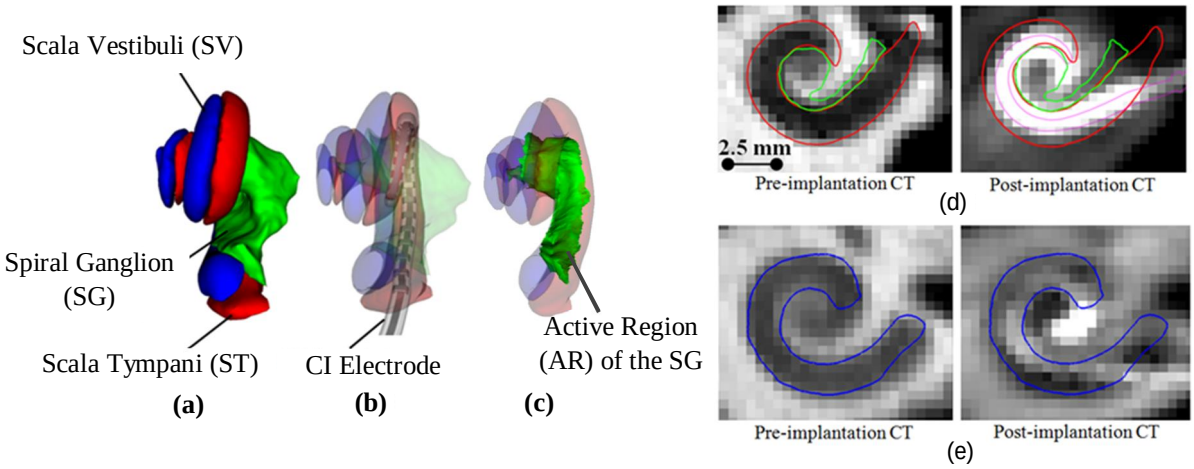


Figure 5.1. Shown in (a) and (b) are surfaces of ST (red), SV (blue), and SG (green). In (b), a surface model of a CI electrode array inserted into ST is shown. In (c), surfaces of AR (green), ST (transparent red), and SV (transparent blue). In (d), contours of ST (red), SG (green), the electrodes (purple) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT, and in (e) contours of the SV (blue) in the coronal view of a pre-implantation CT and a corresponding post-implantation CT.

However, the approaches we developed thus far cannot be used for many CI recipients for whom a pre-implantation CT of neither ear is available. In this study, we propose algorithms to extend our IGCIP strategies to the population of bilateral CI recipients for whom a CT has not been acquired prior to implantation, thereby making our IGCIP technique available for all CI recipients. The algorithms we present permit automatic segmentation of the SOIs in the post-

implantation CTs directly, without the need for a pre-implantation CT of either ear, despite the significant artifacts induced by the CI electrodes in those images. This new approach capitalizes on the physical relationship between the cochlear anatomy and the labyrinth, i.e., the rest of the inner ear.

5.2. Methods

Our approach consists of two main steps. First, (A) we localize the entire labyrinth, which we use as a landmark structure, by (1) coarsely estimating its position using a shape chosen from a library of labyrinth shapes, (2) automatically creating a statistical shape model that is specific to the subject, and (3) refining the coarse estimate by performing a statistical shape model-based segmentation. The labyrinth is a structure that shares the external wall of the cochlea with the intra-cochlear anatomy and that also includes the semi-circular canals. Figure 5.2a and 5.2b show a surface of the labyrinth and the intra-cochlear structures that it externally bounds. Next, (B) we segment the SOIs by fitting the subset of SOI model points, that represent the external wall of the cochlea (see Figure 5.2d and 5.2e), to the part of the labyrinth that represent the same (see Figure 5.2c). To do the fitting, we established offline a one-to-one point correspondence between the subset of SOI model points and the subset of labyrinth model points that represent the exterior of the cochlea. The exterior region of the cochlea, which we use to fit the SOIs model, is the only portion of the SOIs that (a) has contrast in CT and (b) can also be localized by the labyrinth.

The following subsections detail our approach. The datasets we use in this study are detailed in Section 5.2.1. The image-to-image registration process and active shape segmentation process we use at various steps throughout our study are detailed in Section 5.2.2 and 5.2.3, respectively. The process we use to create our shape library is detailed in Section 5.2.4. In

Section 5.2.5, we present the multi-step process we propose to localize the labyrinth and use its position to predict the position of the intra-cochlear anatomy. In Section 5.2.6, we present the process we use to show the feasibility of estimating the shape of the SOIs using the position of the labyrinth. The remaining sections detail our approaches for algorithm parameters selection and segmentation validation.

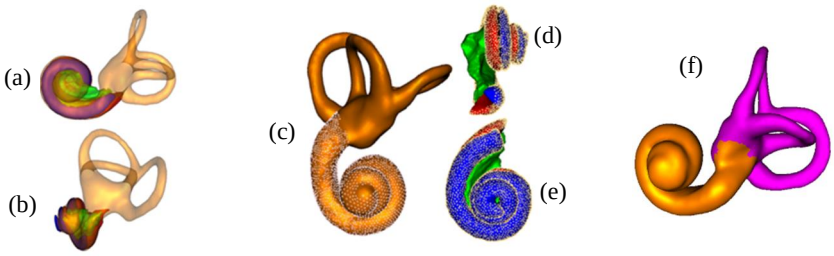


Figure 5.2. Shown in (a) are surfaces of a labyrinth (transparent orange) and of the intra-cochlear anatomy (ST (transparent red), SV (transparent blue), and SG (transparent green)). Panel (b) shows the same structures in a different orientation. In (c) the set of points that represent the external wall of cochlea and that are used to fit the SOI model to the labyrinth model is shown on the surface of the labyrinth. In (d) and (e) the same set of points is shown on the SOI surfaces. Panel (f) shows a labyrinth surface with near points in yellow and far points in purple.

5.2.1. Data

In this study, we use several groups of head CT scans, which are presented in Table 5.1. The scans were acquired from several conventional scanners (GE BrightSpeed, LightSpeed Ultra; Siemens Sensation 16; and Philips Mx8000 IDT, iCT 128, and Brilliance 64) and a low-dose flat-panel volumetric CT (fpVCT) scanner (Xoran Technologies xCAT® ENT). Conventional CTs of 70 subjects are used for creating a shape library of the ST, SV, SG, and the labyrinth as discussed in Section 5.2.4, conventional CTs of 25 subjects are used for creating a statistical shape model for the labyrinth as discussed in section 5.2.3, fpVCT scans of 14 subjects are used for creating an intensity model of the labyrinth as discussed in section 5.2.5, and CT-fpVCT pairs of 21 subjects are used for validating the segmentation results as discussed in Section 5.2.6. Typical voxel size for conventional CTs is $0.25\times0.25\times0.3$ mm^3; for flat-panel CTs it is

$0.4 \times 0.4 \times 0.4$ mm^3.

Our validation dataset (dataset 5) is constructed such that it allows us to (1) generate automatic segmentations on the post-implantation CTs using the approach we propose (2) rigidly register the post-implantation CTs to the corresponding pre-implantation CTs and (3) validate our results by comparing registered automatic segmentations to gold-standard segmentations established on the pre-implantation CTs. Dataset 5 consists of four subgroups. The first (12 subjects) includes a set of conventional pre- and low-dose post-implantation CTs of 12 unilateral CI recipients. The second (2 subjects) includes a set of conventional pre- and low-dose post-implantation CTs of 2 bilateral CI recipients. The third (5 subjects) includes a set of conventional pre- and post-implantation CTs of 5 unilateral CI recipients. The fourth (2 subjects) includes a set of conventional pre- and post-implantation CTs of 2 bilateral CI recipients. In total we have 25 pairs of pre- and post-implantation CTs (12+2×2 CT-fpVCT pairs in the first and second subgroup; 5+2×2 CT-CT pairs in the third and fourth subgroup) on which we can test our segmentation algorithm. Each pair corresponds to one implanted ear.

Table 5.1. Datasets used in this study.

Dataset #	Purpose	Dataset Size	Acquisition		CI electrodes		
			Xoran	Conventional	No CIs	One CI	Two CIs
1	Reference (atlas)	1		×	×		
2	Shape library creation	70		×	×		
3	Labyrinth ASM	25		×	×		
4	Intensity model	14	×		×		
5	Segmentation validation	12		×	×		
			×			×	
		2		×	×		
			×				×
		5		×	×		
				×		×	
		2		×	×		
				×			×

5.2.2. Image Registration Methods

In this subsection, we present the image-to-image registration process that we use at various steps throughout our study. Given a "fixed" image, i.e., an atlas or a reference image, and a "floating" image, i.e. the target image we want to segment, we use a multi-step process outlined in Figure 5.3 to register them. First, we affinely register the entire but downsampled images. Next, we refine the registration by limiting the region of interest to a pre-defined region that encompasses the ear structure. At this stage, the affine transformation is estimated at full resolution. In both cases, we rely on an intensity-based technique that uses Powell's direction set method and Brent's line search algorithm [8] to optimize the mutual information [9]-[10] between the images. The registration within the region of interest is further refined with a non-rigid registration step. The non-rigid registration method we use models the deformation field as a linear combination of radial basis functions with finite support [11], and computes the basis function coefficients that optimize the normalized mutual information [12] between the images through a steepest gradient descent optimization process.

1) Affine registration
    a. Affinely register the floating image to the fixed image, after downsampling both images by a factor of four in each dimension.
    b. Crop the ear region from the affinely registered floating image.
    c. Affinely register the floating ear region to the fixed ear region image at full image resolution.
2) Non-rigid registration
    a. Non-rigidly register the floating ear region to the fixed ear region.

Figure 5.3. Image registration process.

5.2.3. Active Shape Model (ASM)-based Segmentation

Various processes we describe in the following subsections rely on the creation of an active shape model (ASM) and performing active shape segmentation. Thus, in the following subsections we describe the general ASM framework that we use.

5.2.3.1. Active Shape Model (ASM) Creation

Given a reference surface and a set of training surfaces of a structure with a one-to-one point correspondence between the points on the reference surface and the points on each training surface, we perform the following steps to create an ASM of a structure. First, we register each training surface to the reference surface with a 7-Degree-Of-Freedom (DOF) (three translations, three rotations, and one isotropic scaling) transformation that minimizes the root-mean-squared (RMS) distance between the surfaces. Next, we use the registered surfaces to build the structures ASM according to the procedure described by Cootes in [13]. Finally, we store the ASM in the reference image space. The ASM is represented by the mean shape $\{\bar{\boldsymbol{x}}_i\}_{i=0}^{N-1}$, with N being the number of points in the shape, and L eigenvectors, $U = \{\boldsymbol{u}_l\}_{l=0}^{L-1}$, that are corresponding to the largest eigenvalues $\{\lambda_l\}_{l=0}^{L-1}$. Mathematically,

$$\{\lambda_l, \boldsymbol{u}_l\}_{l=0}^{L-1} : \lambda_l \boldsymbol{u}_l = X\boldsymbol{u}_l \quad , \tag{5.1}$$

where X is the covariance matrix of the registered surfaces.

5.2.3.2. Active Shape Segmentation

Active shape segmentation of the structure in a target image is performed by fitting the ASM to an initial estimate of the shape. This process consists of three main steps: *(i) Shape initialization:* We determine an initial coarse estimate of the shape $\{\boldsymbol{x}_i\}_{i=0}^{N-1}$ by projecting the mean shape $\{\bar{\boldsymbol{x}}_i\}_{i=0}^{N-1}$ from the reference image space to the target image space using the registration

transformation that registers the two images. This registration transformation is computed using the image-to-image registration process described in Section 5.2.2. *(ii) Shape adjustment*: We adjust the initial shape by iteratively finding a candidate position for each ith point in the initial shape and fitting the shape model to these candidate positions in a weighted least squares sense. The candidate position $\boldsymbol{x}_i'$ for each initial point $\boldsymbol{x}_i$ is determined along the surface normal $\widehat{\boldsymbol{n}}_i$ in the interval $[-1.5, 1.5]$ mm, equivalently,

$$\boldsymbol{x}_i' = \boldsymbol{x}_i + \Delta d \cdot k_{\min} \cdot \widehat{\boldsymbol{n}}_i \tag{5.2}$$

where $\Delta d = 0.15$ mm, and $k_{\min}$ is chosen as,

$$k_{\min} = \arg\min_k C_i(k) : k \in \{-10, -9, ..10\} , \tag{5.3}$$

i.e., the candidate position for the ith point is the position at which the cost function $C_i(\bullet)$ is the smallest cost value in the interval $[-1.5, 1.5]$ mm along $\widehat{\boldsymbol{n}}_i$. The cost function is tailored to the type of image we use as described in the following section. We then fit the shape model to the candidate points $\{\boldsymbol{x}_i'\}_{i=0}^{N-1}$ to obtain an adjusted shape $\{\boldsymbol{x}_i''\}_{i=0}^{N-1}$, given by

$$\boldsymbol{x}_i'' = \psi^{-1}\left(\overline{\boldsymbol{x}}_i + \sum_{l=0}^{L-1} b_l \boldsymbol{u}_{l,i}\right), \tag{5.4}$$

where ***b,*** a vector of parameters that define the shape, is given by

$$\boldsymbol{b} = [b_0, b_1, \ldots, b_{L-1}] = (U^T W^T W U)^{-1} U^T W^T W \boldsymbol{d} \tag{5.5}$$

and $\boldsymbol{d} = \{\boldsymbol{d}_i\}_{i=0}^{N-1}$, defined as

$$\boldsymbol{d}_i = \psi(\boldsymbol{x}_i') - \overline{\boldsymbol{x}}_i \text{ for } i \in \{0, N-1\} \tag{5.6}$$

is the residual between the mean shape and candidate points, after they are registered to the mean shape with a 7-DOF (three translations, three rotations, and one isotropic scaling) transformation ψ, computed as,

$$\psi = \arg\min_{\psi} \sum_{i=0}^{N-1} w_i^2 \|(\psi(\boldsymbol{x}_i') - \bar{\boldsymbol{x}}_i)\|^2. \quad (5.7)$$

We assign a reliability weight $w_i \in [0,1]$ for each ith candidate position. The reliability weight computation, as we will explain in the following sections, is tailored for the type of image we want to segment. The weight matrix $W = \text{diag}([\boldsymbol{w}_0, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_{N-1}])$, with $\boldsymbol{w}_i = [w_i, w_i, w_i]$, in Eqn. (5.5) is designed so that candidate points with high reliability have more influence on the least squares model fitting. *(iii) Iterate shape adjustment*: We iterate the shape adjustment process until the RMS distance between the adjusted shape and the initial shape at that iteration is small, specifically until $(1/N \sum_{i=0}^{N-1} \|\boldsymbol{x}_i'' - \boldsymbol{x}_i\|^2)^{1/2} < \varepsilon$ is satisfied, where ε is empirically set to 0.01 mm.

In summary, given an ASM of a structure and its initial shape estimate, we iteratively fit the ASM to segment the structure. At each iteration, we determine a candidate position for each ith point using Eqn. (5.2), we re-compute a weight for each ith candidate point, and finally we determine an adjusted shape by fitting the ASM to the candidate points in a weighted least squares sense using Eqn. (5.4).

5.2.4. Shape Library Creation

Our segmentation approach, as we will discuss in the next subsection, relies on first determining a coarse estimate of the labyrinth, which we use as a landmark, and of the intra-cochlear structures, which are the SOIs, by mapping surfaces of the labyrinth and SOIs chosen from a library of labyrinth and SOI surfaces to the subject's images. To enable this approach we create a library of internal ear structures segmented in a number of subjects' pre-implantation CTs.

Specifically, we use the set of CTs in dataset 2 (see Table 5.1) to create a library of surfaces that represent the labyrinth and SOIs of each subject's left or right ear, chosen randomly.

To produce a surface of the labyrinth in each CT in this dataset, we perform an ASM-based segmentation of the labyrinth in the CT using the active shape segmentation process described in Section 5.2.3.2. The labyrinth ASM we use in this step is created offline, using the reference CT in dataset 1 and the set of training CTs in dataset 3, according to the process we previously reported in [7] for the same purpose. When segmenting an image with this model, the cost $C_i(k)$ we use for candidate position selection in Eqn. (5.3) is given by

$$C_i(k) = -|I(\boldsymbol{x}_i + \Delta d \cdot (k+1) \cdot \hat{\boldsymbol{n}}_i) - I(\boldsymbol{x}_i + \Delta d \cdot (k-1) \cdot \hat{\boldsymbol{n}}_i)|, \tag{5.8}$$

where $I(\bullet)$ is image intensity in the CT at a given point. It is thus designed such that a candidate position for the ith point is chosen to be the position with the largest intensity gradient over the interval $[-1.5, 1.5]$ mm along $\hat{\boldsymbol{n}}_i$. We use two different approaches for selecting candidate positions, one for *contrasted* points $C \subset \{\boldsymbol{x}_i\}_{i=0}^{N-1}$, which are the subset of points we know a *priori* are well contrasted in CT, and one for *non-contrasted* points $C' = \{\boldsymbol{x}_i\}_{i=0}^{N-1} \setminus C$, which are the rest of the labyrinth surface points. Points that belong to C and C' have been labeled at the time the models were created. For each $\boldsymbol{x}_i \in C$, we determine a candidate position $\boldsymbol{x}_i'$ using Eqn. (5.2) and assign a reliability weight of $w_i = 0.99$, while for each $\boldsymbol{x}_i \in C'$ we use its original initial position determined via image registration as a candidate position and we assign a reliability weight of $w_i = 0.01$. A relatively high weight is thus assigned to the candidate positions for C so that the shape fitting is influenced more by those points with contrast in the CT. Although the results obtained with this technique are generally satisfactory, there are cases where mis-segmentation is observed. This is likely caused by the limited number of shapes we use to create our ASM, which may not be able to capture enough variability to segment

accurately all the images in our library. To deal with this issue, at each iteration, we determine the final adjusted point for the ith point using the equation

$$\boldsymbol{x}_i^a = \alpha \cdot \boldsymbol{x}_i'' + (1-\alpha) \cdot \boldsymbol{x}_i', \tag{5.9}$$

which is a weighted combination of the position of the fitted model position $\boldsymbol{x}_i''$, given by Eqn. (5.4), and the candidate position $\boldsymbol{x}_i'$, given by Eqn. (5.2), controlled by the weight parameter α. We set α to be 0.8 initially and we perform our iterative shape adjustment while decrementing α by 0.1, at the end of each iteration, for the first six iterations and use the final value of α for the remaining iterations. The value of α is set such that we largely rely on the model at the beginning. As we iteratively obtain better estimates of the shape, we gradually rely more on the candidate points which are likely to be positions with strong image gradient.

Finally, after the labyrinth is segmented algorithmically we manually adjust the segmentation to correct for any visually identifiable error. We then rely on the segmented labyrinth surface and an ASM of the SOIs, which we previously created and reported in [6], to segment the SOIs. To do this, we first establish offline a one-to-one point correspondence between the model points of the SOIs and the model points of the labyrinth. The SOI model points are then fitted to the corresponding points on the segmented labyrinth.

We produce the surfaces such that there is a one-to-one, across subject, point correspondence between the points composing the surfaces. For the purpose of segmentation strategy, which we will explain, we divide the points on each labyrinth included in the library into two groups: *near* points, which are points that may be close to implanted electrodes, and *far* points, which are the rest of the points. Figure 5.2f shows a surface of the labyrinth with the two point groups rendered with different colors.

5.2.5. Inner-Ear Anatomy Segmentation

Our approach for segmenting both the labyrinth and the SOIs in a target CT is to first determine a coarse estimate of the structures and then refine this coarse estimate. To identify a coarse estimate of the structures we map surfaces of the structures chosen, from our shape library, that best localize the structures in the target CT. We then create ASM models for the structures using the subset of shapes chosen, from our shape library, based on their similarity to the corresponding structure shapes in the target CT. Finally, we refine the coarse estimate using a standard weighted ASM-based segmentation method. The following subsections detail our shape library-based coarse shape estimation and our shape model-based segmentation refinement steps.

5.2.5.1. Shape Library-based Segmentation Initialization

We coarsely localize the labyrinth with a labyrinth surface chosen from our shape library. The surface is chosen such that its *far* points (see Figure 5.2f) best approximate the *far* points portion of the labyrinth in the target image. This process includes several steps. First, we determine the *far* points of the labyrinth in the target image by fitting the *far* points of the labyrinth ASM (see Section 5.2.3.1) following the segmentation process described in Section 5.2.3.2. The *far* points are likely to be far from implanted electrodes. They are thus unlikely to have been affected by implant-related artifacts in the image. Next, we register each labyrinth surface in our shape library to the target image using the transformation that minimizes the RMS distance between the *far* points on the library surface and the *far* points localized in the target image in the previous step. Finally, we compute a dissimilarity quantity $DS(k)$ for each registered surface k as the residual RMS. The registered surface with the smallest dissimilarity quantity k_s is used as the coarse segmentation, with k_s defined as

$$k_s = \arg\min_k \big(DS(k)\big) \; : \; k \in \{0, 1, \dots, M-1\} \quad , \tag{5.10}$$

$$DS(k) = \frac{1}{N^{\mathrm{f}}} \sum_{i=0}^{N^{\mathrm{f}}-1} \|\boldsymbol{x}_i - T_k(\boldsymbol{x}_{ki})\|^2 \tag{5.11}$$

in which M is the number of subjects in the library, $\{\boldsymbol{x}_i\}_{i=0}^{N^{\mathrm{f}}-1}$ is the set of *far* points localized in the image, $\{\boldsymbol{x}_{ki}\}_{i=0}^{N^{\mathrm{f}}-1}$ is set of *far* points in the kth shape in the library, N^{f} is the number points in the *far* portion of the labyrinth, and T_k is the 6-DOF (three rotations, three translations) transformation that registers the two *far* point sets, computed as,

$$T_k = \arg\min_T \frac{1}{N^{\mathrm{f}}} \sum_{i=0}^{N^{\mathrm{f}}-1} \|\boldsymbol{x}_i - T(\boldsymbol{x}_{ki})\|^2. \tag{5.12}$$

The value of the dissimilarity term is low when the shape represented by the *far* points localized in the image closely matches the shape represented by the *far* points in the kth surface. As we will show in the results section, the *far* portion of the labyrinth can be used as a good landmark for predicting the position of the labyrinth. A coarse segmentation of the SOIs is obtained by projecting the k_s-th subject's SOI surfaces to the target image through T_k.

In Section 5.2.6, we show the feasibility of estimating the shape of the SOIs using the position of the labyrinth. Specifically, we show that (a) the choice and the minimization of the dissimilarity quantity, defined in Eqn. (5.11), is a good strategy for selecting similar shapes in our library and (b) the dissimilarity quantity well correlates with SOIs localization error.

5.2.5.2. Shape Model-based Segmentation Refinement

To refine the coarse segmentations, we first segment the labyrinth by performing a weighted active shape segmentation and then segment the SOIs by fitting their ASM to the segmented labyrinth. This process is summarized in Figure 5.4.

1) Build a target-specific ASM for both the labyrinth and SOIs
2) Perform ASM segmentation of the labyrinth
   a. Initialize the shape
   b. Adjust the shape
   c. Iterate (b) until convergence
3) Segment SOIs by fitting their ASM to the labyrinth determined in (2).

Figure 5.4. Segmentation refinement process

First, we create two ASMs, one for the labyrinth and another for the SOIs using a subset of surfaces obtained from our shape library. These are chosen as the labyrinth and SOI surfaces of seven (a number chosen experimentally, see Section 5.2.7) subjects for whom the dissimilarity quantity computed using Eqn. (11) is the smallest. The ASMs we create are thus specific to each target image. Next, using this target-specific ASM, we refine the initial labyrinth segmentation as discussed in Section 5.2.3.2. In this process, we use the coarse labyrinth localized in section 5.2.5.1 as the initial shape. We then iteratively refine it by first finding candidate position $\boldsymbol{x}_i'$ (see Eqn. (5.2)) for each ith point $\boldsymbol{x}_i$ and then fitting the ASM to the candidate positions in a weighted least squares sense (see Eqn. (5.4)). The cost function $C_i(\bullet)$ we use for candidate position selection in Eqn. (5.3) is tailored to the type of image we want to segment. When the target CT is obtained from conventional scanners, $C_i(\bullet)$ is a function of the image gradient at that ith point in the target images, mathematically it is computed using Eqn. (5.8). When the target CT is obtained from low-dose scanners, $C_i(\bullet)$ is a function of an intensity model of the image at that ith point. To build the intensity model, we rely on a set of manually segmented labyrinth

surfaces obtained from dataset 4. For each jth training surface $\{\boldsymbol{x}_{ji}\}_{i=0}^{N-1}$ an intensity profile $\boldsymbol{p}(\boldsymbol{x}_{ji})$ is extracted at each ith point along the normal $\hat{\boldsymbol{n}}_{ji}$ using the equation

$$\boldsymbol{p}(\boldsymbol{x}_{ji}) = \left[I_j(\boldsymbol{x}_{ji} - \Delta d \cdot 9 \cdot \hat{\boldsymbol{n}}_{ji}), I_j(\boldsymbol{x}_{ji} - \Delta d \cdot 8 \cdot \hat{\boldsymbol{n}}_{ji}), \dots, I_j(\boldsymbol{x}_{ji} + \Delta d \cdot 9 \cdot \hat{\boldsymbol{n}}_{ji})\right]^T, \quad (5.13)$$

where $\Delta d = 0.15\,\text{mm}$, and $I_j(\bullet)$ is the intensity of the jth training image at a given point. The intensity model at the ith point is given by $\{\boldsymbol{p}(\boldsymbol{x}_{ji})\}_{j=0}^{M-1}$, where M is the number of training surfaces. Finally, the cost function is designed as

$$C_i(k) = \min_j \left\| \boldsymbol{p}(\boldsymbol{x}_i + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_i) - \boldsymbol{p}(\boldsymbol{x}_{ji}) \right\| : j \in \{0, 1, \dots, M-1\}, \quad (5.14)$$

which defines the cost for selecting $\boldsymbol{x}_i + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_i$ as the candidate position for $\boldsymbol{x}_i$ as the minimum Euclidean distance between the intensity profile at $\boldsymbol{x}_i + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_i$ and all the M intensity profiles contained in the set of model profiles at the ith point. The reliability $w_i \in [0, 1]$ we assign for each ith point is based on the intensity profile extracted at the ith point in I, and is given by,

$$w_i = \frac{\#\{k \in \{-10, -9, \dots, 10\} : I(\boldsymbol{x}_i + \Delta d \cdot k \cdot \hat{\boldsymbol{n}}_i) < R\}}{21} \quad (5.15)$$

where R is an intensity threshold that separates the bright metallic artifact from the rest of the structures. This weight is high when the set of intensity values in a given profile are below R, which indicates that the extracted profile is far from the image artifact in the image and is thus more likely to be reliable. To determine this threshold, the maxima along all the intensity profiles extracted along the surface normals at the points composing the initial shape are first computed. The threshold is then chosen experimentally to be the 90$^{\text{th}}$ percentile of the distribution of maxima. It is thus adapted to each image. Finally, we segment the SOIs by fitting the points on

the target-specific SOIs' ASM to their corresponding points on the segmented labyrinth, as discussed in Section 5.2.4.

5.2.6. Feasibility of the Initialization Method

To determine the feasibility of our initialization method, we use it to segment pre-implantation CTs for which gold-standard segmentations are available. Specifically, we relied on (a) the gold-standard segmentations of the labyrinth and SOIs established on the set pre-implantation CTs in dataset 5 (see Table 5.1) and (b) our shape library (see Section 5.2.4). We then perform the following steps on each pre-implantation CT in dataset 5. First, we identify the labyrinth and SOI surfaces using the initialization method described in Section 5.2.5.1. In this step, we use the *far* points of the gold-standard labyrinth established on the CT, rather than automatically determined *far* points. Next, we compare the identified SOI surfaces to the corresponding gold-standard SOI surfaces and measure surface distances from the points on the identified SOI surfaces to the corresponding points on the gold-standard SOI surfaces. The same distances are also measured for the labyrinth. Finally, we measure the correlation between the dissimilarity quantities, computed using Eqn. (5.11) for each shape in the library, and the errors for the SOIs localized by mapping each shape in the library.

5.2.7. Parameter Selection

To arrive at the size of the subset of shape samples we use for refinement, as described in Section 5.2.5.2, we use the set of testing image pairs in dataset 5 (see Table 5.1). For each testing image pair, we first sort the shapes in the library in ascending order of their dissimilarity with the shapes in that testing image pair, computed using Eqn. (5.11). Next, we vary the size in increments of one, from two most similar shapes to the total number of shapes in the library, and we measure the resulting SOIs mean segmentation error on all testing image pairs. Finally, we

select the size for which the overall SOIs mean segmentation error is the smallest. Figure 5.5a, shows a plot of SOIs mean segmentation error over 25 testing image pairs versus the size of the library shapes used in the refinement method, which ranges from two to 70 (the total number of samples in our shape library). As shown in the figure, the seven most similar samples lead to the smallest overall SOIs mean segmentation error. This is the size that is used for all testing image pairs to produce the results presented herein. Figure 5.5b and 5.5c shows the same plot for the 9 CT-CT testing image pairs and the 16 CT-fpVCT testing image pairs, respectively. The corresponding size for the CT-CT testing pairs is five; for the CT-fpVCT testing pairs it is seven. The first value in each plot (pointed by arrows in the figures) is the error for the SOIs localized using the initialization method alone, which uses one most "similar" subject selected from the shape library. The plot for both groups of testing pairs suggest that (a) using our refinement method generally leads to a reduction in the initial overall SOIs segmentation error, and (b) using the refinement method with a subset of the shape samples, rather that all shape samples, in the library results in smaller overall SOIs mean segmentation error.

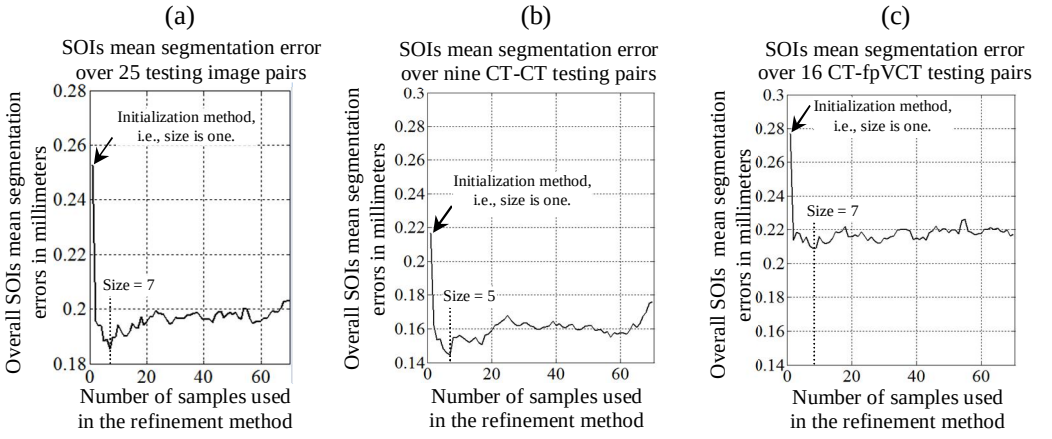


Figure 5.5. (a) Overall SOIs mean segmentation error for all 25 testing ears versus the number of most similar shapes sampled from the shape library to create target-specific ASMs. (b) and (c) present the same plot for the 9 CT-CT testing image pairs and the 16 CT-fpVCT testing image pairs, respectively.

We validate our method by automatically segmenting the ST, the SV, the SG, the AR and the labyrinth in the post-implantation CTs in dataset 5 (see Table 5.1) using the approach we propose

and by measuring the resulting segmentation errors. The gold-standard surfaces that we use for comparison were created in the corresponding pre-implantation CTs by manually editing surface points on segmentations that are automatically initialized by pre-implantation CT segmentation techniques we previously developed [2], [7]. For each structure, we measure a distance from each point on its automatically generated surface to the corresponding point on its gold-standard surface, and report the mean, standard deviation, median and maximum of the distances we measure over all points on the surface. To quantify the improvement afforded by our refinement method, we measure the same segmentation error when we only use the segmentation initialization step.

5.3. Results

5.3.1. Feasibility Study of the Initialization Method

Figure 5.6 shows the mean, median, standard deviation, and maximum error box plots for the SOIs and labyrinth localized on all 25 validation pre-implantation CTs using the segmentation initialization method as described in Section 5.2.6. Average SOIs and labyrinth localization errors are sub-voxel. In this process, the initialization method relies on the gold-standard position of the *far* points, rather than automatically detected *far* points, for mapping the library surfaces. Thus, the results we achieve (a) are the best results that can be achieved using the initialization method, and (b) show that, given the true position of the *far* points, the initialization method can localize the SOIs with a sub-voxel level of accuracy.

In Figure 5.7, we show bar plots for the correlation coefficients computed for all 25 validation pre-implantation CTs. For each pre-implantation CT, the correlation coefficient is computed between (a) the set of segmentation errors for the SOIs localized by mapping each shape in the library (see Section 5.2.6), and (b) the set of dissimilarity quantities computed

between the shape in that pre-implantation CT and each shape in the library. As shown in the figure, the correlation coefficient for all 25 pre-implantation CTs is in the interval [0.62, 0.84]. This result indicates that there is a good correlation between the dissimilarity quantity and the SOIs segmentation error and that the "dissimilarity" quantity we use for choosing "similar" samples in our library is a reasonably valid quantity.

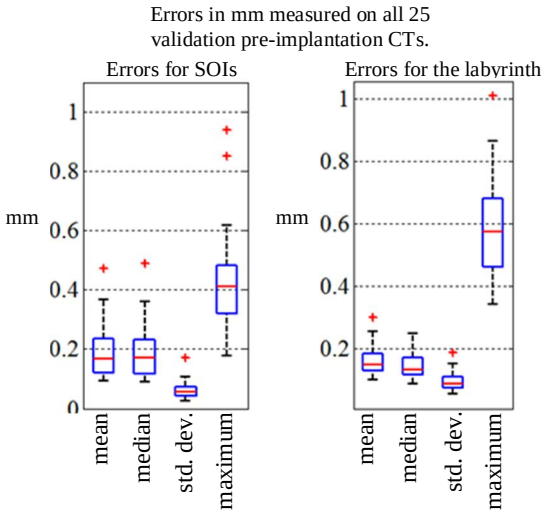


Figure 5.6. The left panel shows mean, median, standard deviation, and maximum distance errors box plots for the SOIs localized, on all 25 pre-implantation CTs, using the technique described in Section 5.2.6. The right panel shows the same information for the labyrinth.

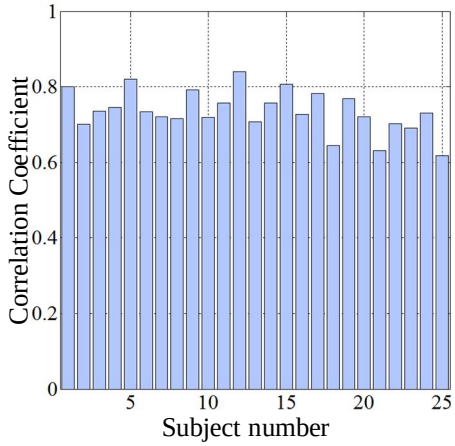


Figure 5.7. Correlation coefficient for each validation pre-implantation CT.

5.3.2. Segmentation of the Inner Ear Anatomy

We evaluate the accuracy of the proposed segmentation method on the set of pre- and post-implantation testing image pairs (see Dataset 5) by comparing the SOIs (the ST, the SV and the SG), the AR and the labyrinth segmentations automatically generated using the proposed method to the corresponding gold-standard segmentations established on the pre-implantation CTs. We present quantitative results for each structure generated on the 16 CT-fpVCT testing pairs, on the 9 CT-CT testing pairs, and on all 25 (16+9) testing image pairs. Figure 5.8a shows the mean, standard deviation, median and maximum surface error box plots in mm for the SOIs segmented on all 16 CT-fpVCT testing pairs using the initialization method (left panel) and using the refinement method (right panel). Figure 5.8b and 5.8c show the same results generated on all 9 CT-CT testing pairs and on all 25 testing pairs, respectively. Figure 5.9 shows the same information for the labyrinth. The results presented in Figure 5.8 and 5.9 show that the errors we achieve on CT-CT testing pairs are smaller than those that we achieve on CT-fpVCT testing pairs. This is likely because the post-implantation CTs in the CT-CT testing pairs are of good quality, which could lead to better accuracy in structures edges localization. The overall 25 ears mean, standard deviation, median and maximum errors are 0.253, 0.162, 0.209, 1.209 mm, respectively, for results achieved using the initialization method alone and 0.186, 0.110, 0.157, 0.946 mm, respectively, for results achieved using the refinement method. Overall SOI average segmentation errors in both groups of testing pairs are close to half the voxel size in the segmented CT and are all sub-millimetric (< 1mm).

A method we previously developed for segmenting the SOIs in pre-implantation CTs [2] resulted in mean and maximum segmentation errors of 0.15 and 1.6 mm, respectively. This shows that the segmentation errors we achieve in post-implantation CTs are close to those that

are achievable in pre-implantation CTs, despite the presence of metallic artifacts in the image that obscure the structures. Figure 5.10 shows qualitative results for two ears. The left panel is for a test ear with CT-fpVCT pair (7R) whose average error is close to the overall average error in the CT-fpVCT group of testing pairs, and the right panel is for a test ear with CT-CT pair (12R) whose average error is close to the overall average error in the CT-CT group of testing pairs. Both the gold-standard (yellow) and automatic (red for ST, blue for SV and green for SG) contours are overlaid on both the pre-implantation CT and the registered post-implantation CT. As can be seen in the figures, although the structures are obscured by the bright artifact introduced by the implant, there is a good agreement between the two contours along the length of the structures for both test ears.

We compared mean errors for the SOIs in each testing ear segmented using the initialization method (initial errors) and segmented using the refinement method (final errors). Figure 5.11 shows bar plots for the initial mean SOIs segmentation error (dark) and final mean SOIs segmentation error (light) for each test ear. As can be shown, for the 84% of cases (21 out of 25), the refinement method has led to a reduction in initial segmentation errors. However, for the four ears (three in the CT-fpVCT testing pairs group and one in the CT-CT testing pairs group) pointed by the arrows in the figure, the initial segmentation errors have gotten worse when we use the refinement method. In each of these ears, we observed that the most "similar" labyrinth shape chosen from the shape library does not localize the SOIs as well as it does the *far* points region of the labyrinth, i.e. the relative position of the *far* points and the *near* points in that particular subject is different than in the selected library shape. This is also the case for the subset of surfaces selected to build the target-specific ASMs for the refinement method. Because of this, the initial SOI segmentation errors are relatively large and they get worse when followed

by the refinement method. We conducted experiments for the four ears where we use the refinement method with a target-specific ASM built with a larger number of most similar shapes, rather than the seven (empirically determined) most similar shapes, in our library. The final SOI segmentation errors we obtain by doing so are smaller than the initial errors for two of these ears, for the other two ears the refinement method did not improve the initial results. This is most likely due to the image artifacts that inhibit the refinement method even when a relatively large number of shapes are used to build the target-specific ASMs. For those two ears where the refinement method reduced the initial errors, 28 and 23 shapes has been used to build the target-specific ASMs. The results from our experiment suggest that for the four ears our target-specific ASM did not capture the target structure shapes well. Although we can improve the initial segmentation errors for at least two of these ears using the refinement method with ASMs built with a larger number shapes, the results we get are not optimal overall. The cross-validation study we perform in Section 5.2.7 show that using the seven most similar shapes leads to the smallest overall SOIs mean segmentation error.

In Figure 5.12, we show renderings of segmented SOI surfaces for 10 representative ears (five with CT-fpVCT pairs (1L, 5R, 6L, 7R, 8R) and another five with CT-CT pairs (12R, 13R, 20L, 20R, 21L)) colormapped with final SOI segmentation errors in mm. As can be seen from the renderings, errors are sub-voxel for the majority of surface points for all ears, except for one (1L), which is one of the ears for which the refinement method did not lead to a reduction in initial segmentation error (see above).

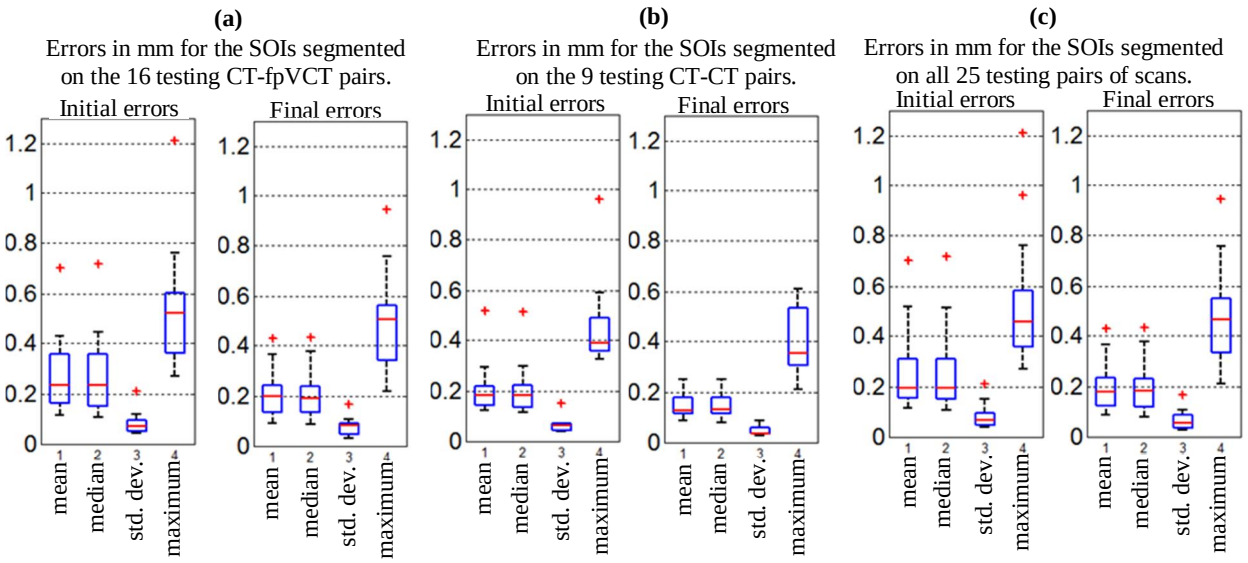


Figure 5.8. Various quantitative SOI segmentation results for test ears with low-dose post-implantation CTs (a), for test ears with conventional post-implantation CTs (b), for all 25 test ears (c). Shown for each group of test ears are (left to right), mean, median, standard deviation, and maximum error box plots for the SOIs segmented using the initialization method; mean, median, standard deviation, and maximum error box plots for the SOIs segmented using the refinement method.

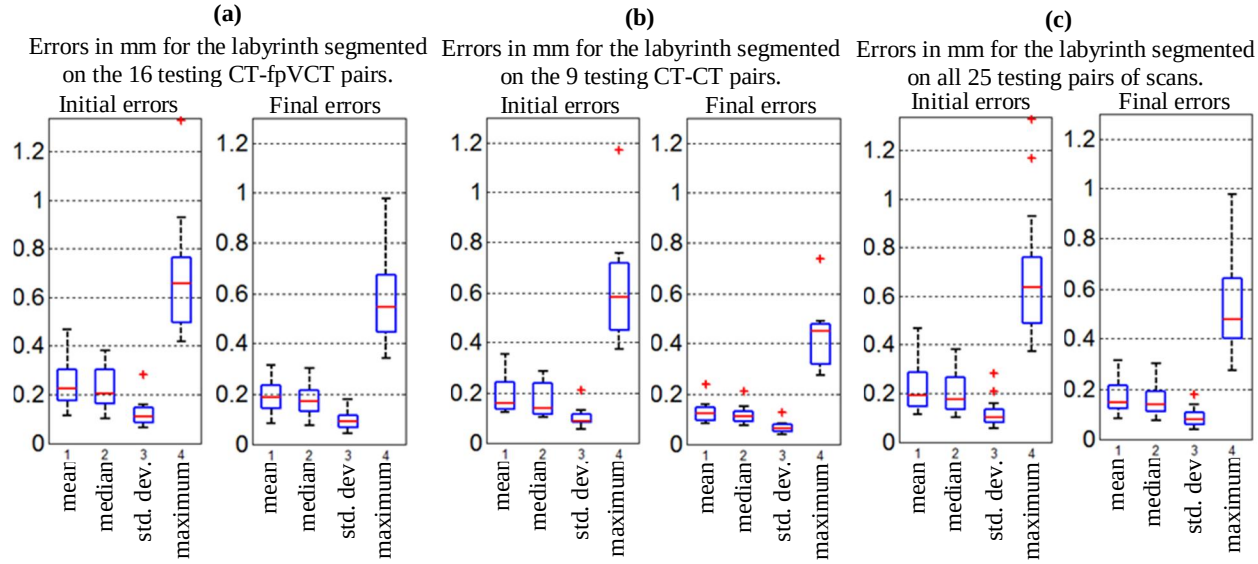


Figure 5.9. Various quantitative labyrinth segmentation results for test ears with low-dose post-implantation CTs (a), for test ears with conventional post-implantation CTs (b), for all 25 test ears (c). Shown for each group of test ears are (left to right), mean, median, standard deviation, and maximum error box plots for the labyrinth segmented using the initialization method; mean, median, standard deviation, and maximum error box plots for the SOIs segmented using the refinement method.

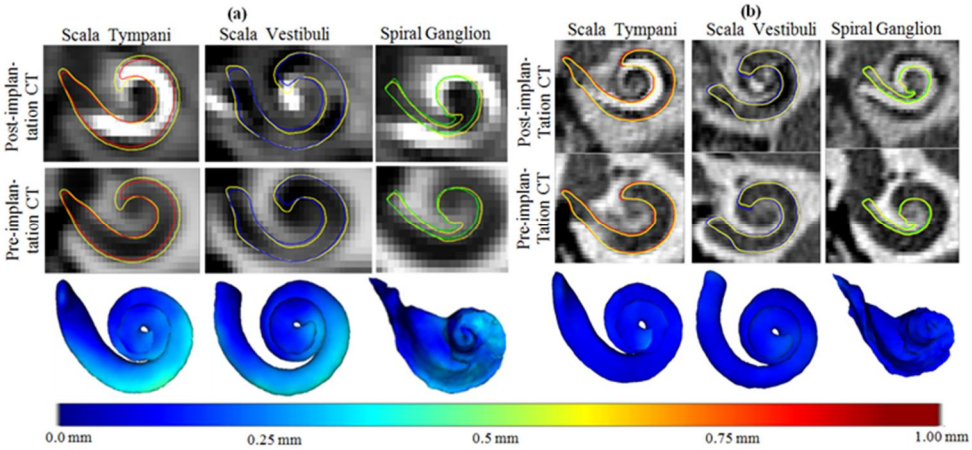


Figure 5.10. (a) Shows qualitative segmentation results for a testing conventional post-implantation CT with mean SOI segmentation error close to the overall mean error for the group of testing conventional CTs. The contours shown are the ST (left panel), SV (middle panel), SG (right panel). Structure contours for gold-standard ST (red), gold-standard SV (blue), gold-standard SG (green), and automatic contours for all structures (yellow) are shown in a slice of a post-implantation CT (top row) and a corresponding pre-implantation CT (middle row). On the bottom panels the structure surfaces colormapped with segmentation errors are shown. (b) Shows similar information for a testing Xoran post-implantation CT with mean SOI segmentation error close to the overall mean SOI segmentation error for the error close to the overall mean error for the group of testing Xoran CTs.

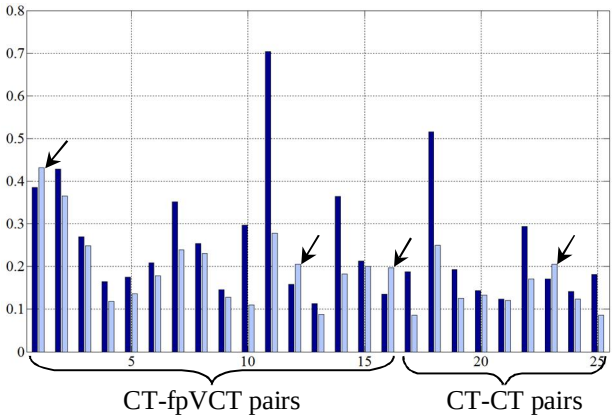


Figure 5.11. Mean error bar plots for the SOIs segmented using the initialization method alone (dark) and refinement method (light), for all 25 test ears.

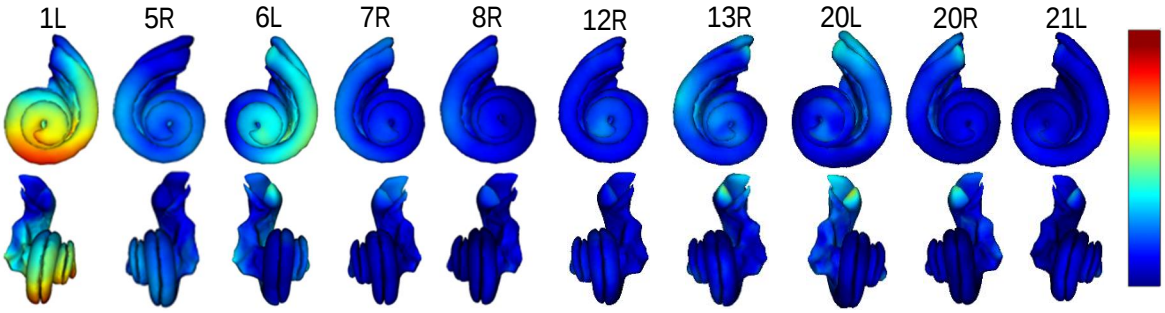


Figure 5.12. Surfaces of intra-cochlear structures colormapped with segmentation errors for representative five test ears with Xoran post-implantation CT (1L, 5R, 6L, 7R, 8R) and for representative five test ears with conventional post-implantation CT (12R, 13R, 20L, 20R, 21L).

Figure 5.13 presents the mean, standard deviation, median and maximum distance error box plots in mm for the AR, which is the part of the SOIs most likely to be stimulated by implanted electrodes and thus the most important surface for our programming application. The results in Figure 5.13a are for the AR segmented on all 16 CT-fpVCT testing image pairs using the initialization method (left panel) and using the refinement method (right panel). Figure 5.13b and 5.13c show the same results on all 9 CT-CT pairs and on all 25 (16+9) testing pairs, respectively. The results in Figure 5.13 show that the errors we achieve on CT-CT pairs are smaller than those that we achieve on CT-fpVCT pairs. This is again likely because the post-implantation CTs in the CT-CT testing pairs are of good quality, which could lead to better accuracy in structure edges localization. The overall mean, standard deviation, median, and maximum errors in segmenting the AR on all 25 testing ears are 0.181, 0.108, 0.153 and 0.946 mm, respectively. The corresponding errors we obtain in segmenting the AR using our initialization method alone, prior to performing our refinement method, are 0.256, 0.167, 0.209 and 1.079 mm, respectively. Figure 5.14 shows renderings of segmented AR surface for 10 representative ears (five with CT-fpVCT pairs and another five with CT-CT pairs) colormapped with the final errors in mm. As can be seen from these, errors are sub-voxel (< 0.4 mm) for the majority of AR for all representative ears, except for one (1L).

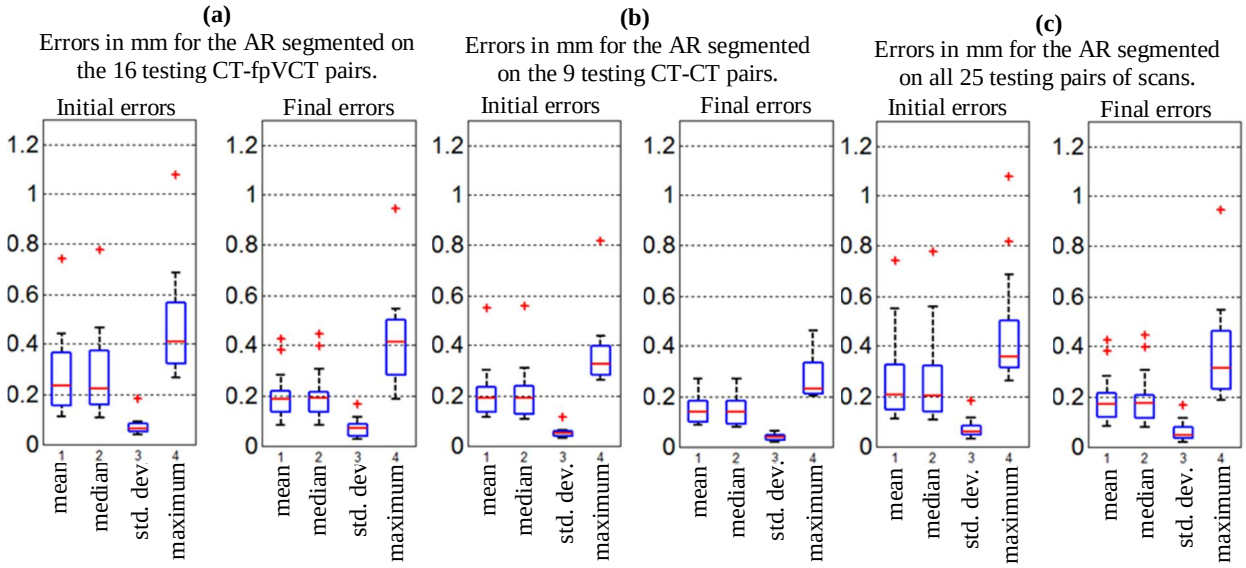


Figure 5.13. Various quantitative AR segmentation results for test ears with low-dose post-implantation CTs (a), for test ears with conventional post-implantation CTs (b), for all 25 test ears (c). Shown for each group of test ears are (left to right), mean, median, standard deviation, and maximum error box plots for the AR segmented using the initialization method; mean, median, standard deviation, and maximum error box plots for the SOIs segmented using the refinement method

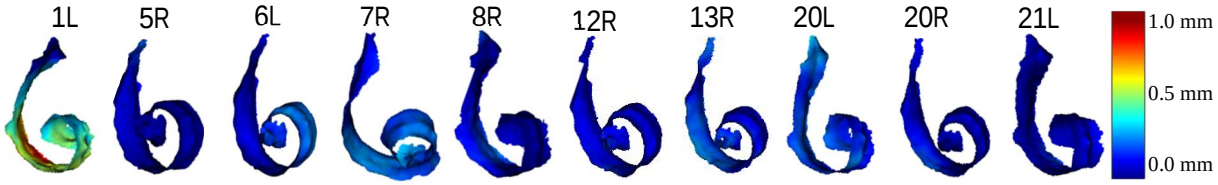


Figure 5.14. Surfaces of the active region colormapped with segmentation errors for a representative five test ears with fpVCT (1L, 5R, 6R,7R, 8L) and for a representative five test ears with conventional CTs (12R, 13R, 20L, 20R, 21L). L is for left ear and R is for right ear.

In Table 5.2, we present a summary of the mean, median, standard deviation and maximum errors for the SOIs, the AR and the labyrinth segmented on (a) the nine CT-CT testing image pairs, (b) the 16 CT-fpVCT testing image pairs and (c) all 25 testing image pairs. The presented results are for the structures segmented using (1) the initialization method and (2) the refinement method, and are the same quantitative results we presented as box plots in Figures 5.8, 5.9 and 5.13.

Table 5.2. Mean, standard deviation, median and maximum segmentation error for segmenting various structures segmented using the initialization method (top) and using the refinement method (bottom).

		Structure	Mean	Std. dev.	Median	Maximum
CT-fpVCT pairs	Initial Errors	AR	0.277	0.175	0.232	1.079
		SOIs	0.273	0.172	0.231	1.209
		Labyrinth	0.243	0.160	0.204	1.329
	Final Errors	AR	0.200	0.120	0.170	0.946
		SOIs	0.209	0.121	0.180	0.946
		Labyrinth	0.190	0.117	0.162	0.979
		Structure	Mean	Std. dev.	Median	Maximum
CT-CT pairs	Initial Errors	AR	0.217	0.142	0.184	0.819
		SOIs	0.217	0.136	0.185	0.962
		Labyrinth	0.190	0.136	0.156	1.690
	Final Errors	AR	0.146	0.071	0.136	0.461
		SOIs	0.144	0.072	0.132	0.612
		Labyrinth	0.126	0.083	0.106	0.736
		Structure	Mean	Std. dev.	Median	Maximum
All pairs	Initial Errors	AR	0.256	0.167	0.209	1.079
		SOIs	0.253	0.162	0.209	1.209
		Labyrinth	0.224	0.154	0.186	1.329
	Final Errors	AR	0.181	0.108	0.153	0.946
		SOIs	0.186	0.110	0.157	0.946
		Labyrinth	0.167	0.111	0.138	0.979

5.4. Conclusions

The IGCIP strategies we recently developed and are currently testing require accurate localization of the position of implanted electrodes relative to intra-cochlear anatomy. So far, we have made this possible for subjects for whom a CT has been acquired prior to implantation, where we segment the SOIs in the pre-implantation CT, identify the electrodes in the post-implantation CT, and register the two CTs to determine the spatial relationship between the implanted electrodes and the SOIs. We have also recently presented a technique that make IGCIP possible for subjects with no pre-implantation CT but who are implanted unilaterally. For this population of subjects, we determine the SOIs in the implanted ear using information extracted from the contralateral normal ear in the post-implantation CT. In this study, we have

presented a shape library-based algorithm that does not require a pre-implantation CT of either ear to segment the SOIs.

Our approach for segmentation relies on first approximating the shape of the labyrinth by mapping a labyrinth surface that is selected from a library of such surfaces, and then refining this shape by performing a weighted active shape segmentation with an ASM built to be specific for the target image. We then segment the SOIs by fitting their shape model to the external wall of the cochlea established on the segmented labyrinth. As the results we present show, we achieve sub-millimetric errors at all points on the surfaces, and overall SOIs segmentation error averages 0.144 mm over nine conventional post-implantation CTs, 0.209 mm over 16 low-dose post-implantation CTs, and 0.186 mm overall 25 post-implantation CTs. These are 0.146, 0.200, and 0.181 mm for the AR, the “important” part of the SOIs. These results, which we achieve on post-implantation CTs, are comparable to those results that are achievable on pre-implantation CTs and this indicates that our approach is accurate enough for use in position-based sound processing strategies. Our approach shows superior accuracy when employed on conventional CTs than on low-dose CTs. It is of note that our method can also be employed to segment SOIs on pre-implantation CTs.

The feasibility study we perform suggests that the initialization method is a feasible method for segmenting the inner ear structures. The segmentation results we achieve using the initialization method alone are very close to those results we achieve by refining the initial results using the refinement method. This shows that the initialization method alone can be used to achieve the task of segmenting the SOIs, particularly, in cases where the external wall of the cochlea in the image is completely obscured by the implant, which could prevent our refinement method from improving the initial results.

Future work will focus on exploring techniques for automatically determining when to use the initialization method alone to achieve the segmentation task. Future work will also include using the proposed segmentation method for determining electrode position-dependent programming strategies for CI users initially at our institution and, subsequently, at other institutions.

5.5. Acknowledgments

This work was supported by NIH grants R01DC008408, R21DC012620, and R01DC010184 from the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of this institute.

References

[1] Noble, J.H., Labadie, R.F., Gifford, R.H., Dawant, B.M., “Image-guidance enables new methods for customizing cochlear implant stimulation strategies,” Neural Systems and Rehabilitation Engineering, IEEE Transactions on 21(5):820-829, Sept. 2013.

[2]. Noble, J.H., Labadie, R.F., Majdani, O., Dawant, B.M., "Automatic Segmentation of Intracochlear Anatomy in Conventional CT," Biomedical Engineering, IEEE Transactions on, 58(9):2625-2632, Sept. 2011a.

[3]. Jack H. Noble, Theodore A. Schuman, Charles G. Wright, Robert F. Labadie, Benoit M. Dawant, “Automatic identification of cochlear implant electrode arrays for post-operative assessment”, Proc. SPIE 7962(796217), Medical Imaging 2011b.

[4]. Noble, J.H., Gifford, R.H., Labadie, R.F., Dawant, B.M., “Statistical Shape Model Segmentation and Frequency Mapping of Cochlear Implant Stimulation Targets in CT,” MICCAI 2012, 421-428, 2012.

[5]. Schuman T.A., Noble J.H., Wright C.G., Wanna G.B., Dawant B.M., Labadie, R.F. “Anatomic Verification of a Novel, Non-rigid Registration Method for Precise Intrascalar Localization of Cochlear Implant Electrodes in Adult Human Temporal Bones Using Clinically-available Computerized Tomography,” The Laryngoscope, 120 (11): 2277-2283, 2010.

[6]. Wanna, G.B., Noble, J.H., McRackan, T.R., Dawant, B.M., Dietrich, M.S., Watkins, L.D., Rivas, A., Schuman, T.A., Labadie, R.F., “Assessment of electrode positions and audiological outcomes in bilateral cochlear implant patients,” Otology & Neurotology, 32(3):428-432, 2011.

[7] Reda, F.A., McRackan, T.R., Labadie, R.F., Dawant, B.M., Noble, J.H., “Automatic segmentation of intra-cochlear anatomy in post-implantation CT of unilateral cochlear implant recipients”, Medical Image Analysis, 2014 (accepted).

[8]. Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. Numerical Recipes in C, 2nd ed. (Cambridge University press, Cambridge, 1992), pp. 412–419.

[9] F. Maes, A. Collignon, D. Vandermeulen, G. Mrchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” IEEE Trans. Med. Imaging 16, 187–198 (1997).

[10] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, “Multi-modal volume registration by maximization of mutual information,” Med. Image Anal. 1, 35–51 (1996).

[11] G. K. Rohde, A. Aldroubi, and B. M. Dawant, “The adaptive bases algorithm for intensity-based nonrigid image registration,” IEEE Trans.Med. Imag., vol. 22, no. 11, pp. 1470–1479, Nov. 2003.

[12]. Studholme, C., Hill, D.L.G., and Hawkes, D.J. “An overlap invariant entropy measure of 3D medical image alignment,” Pattern Recognition, 32(1):71-86, 1999.

[13]. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models - Their training and application,” Comp. Vis. Image Understanding, vol. 61, no. 1, pp. 39–59, 1995.

CHAPTER VI

Summary and Future Work

This dissertation presents several innovative algorithms for automatic segmentation of ear structures and registration of CT images on which we rely to automate or extend a previously developed image-guided cochlear implant surgery technique or image-guided cochlear implant programming technique to a larger population of patients. Specifically, automatic segmentation methods have been introduced for automatic segmentation of the facial nerve, the chorda tympani, the ossicles, the ear canal, the scala tympani, the scala vestibuli, the spiral ganglion and the labyrinth in CTs acquired before cochlear implantation. Methods have been developed for the automatic segmentation of the inner ear structures in CTs acquired before or after cochlear implantation. Methods have also been developed for automatic registration of head CTs that overlap only partially.

In chapter II, we present methods for automatic segmentation of the structure of the ear in pediatric CT. We have observed that there are anatomical differences between pediatric ear anatomy and adult ear anatomy that lead to poor segmentation quality when algorithms previously developed for segmenting adult ear anatomy are employed to segment pediatric ear anatomy. Specifically, we observed that (a) the facial nerve makes a sharper turn near the second genue in pediatric patients than it does in adult patients, (b) in some pediatric patient, the chorda tympani originates from the stylomastoid foramen region (soft tissue in the neck), while in adult patients the chorda tympani typically branches from the vertical segment of the facial nerve, (c)

there is large inter-patient head size variation the pediatric population. To address those issues of anatomical differences, we build new models for pediatric ear anatomy. These models are stored in a new (pediatric) atlas CT that we select to represent the pediatric population. To segment ear structures in pediatric CT, we employ various model-based methods previously developed for segmenting ear anatomy in head CTs of adult subjects. Specifically, to segment the ear canal, the ossicles, and tympanic membrane, which are structures that show high contrast with the surrounding structures and for which a high degree of accuracy is not necessary for our applications, we use single atlas-based segmentation methods. In this approach, models of the structures established on the atlas CT are propagated to the images we want to segment (target CT). The models are propagated through the registration transformation and deformation field that first affinely and then non-rigidly registers the atlas CT to the target CT. To segment the intra-cochlear structures (the scala tympani, the scala vestibuli, and the spiral ganglion), we use an active shape model-based segmentation method developed for segmenting the same intra-cochlear structures in adult CT. The shape model we use with this method is the same cochlear anatomy shape model built for the adult population, as there is evidence that the size and shape of the cochlea does not change as an individual grows [1]. To segment the facial nerve and chorda tympani, we use model-based tubular structure segmentation algorithms developed for segmenting the same tubular structures in adult CT with algorithm parameters optimized for pediatric population. Quantitative segmentation validation performed for the facial nerve and the chorda tympani on 10 ears show that mean, median, standard deviation, and maximum surface distance errors in segmentation are 0.237, 0.121, 0.214, and 1.273 mm for the facial nerve. These results are 0.141, 0.1, 0.1, and 1.241 mm for the chorda tympani. The voxel size of the validation image set ranges from 0.2 to 0.3 mm. Thus, the developed methods are able to achieve sub-voxel

mean error distance for both the facial nerve and chorda tympani. Although maximum segmentation errors are large, we have observed that they are highly localized. For instance, for the facial nerve, these errors typically occur near the end of its horizontal segment. Accuracy in this region is less vital for PCI as it is not immediately adjacent to the facial recess where the desired path of drilling trajectory passes through. Drilling trajectories were computed for each validation ear using the segmentations automatically generated by the methods developed herein. These trajectories were qualitatively assessed in the validation CTs by an experienced otolaryngologist and judged to be safe. The results we have obtained thus suggest that the percutaneous cochlear implantation technique is a viable approach for pediatric patients. To date, the developed methods have been used in conjunction with PCI path planning algorithms for performing pre-operative PCI planning for thirteen pediatric patients. For all thirteen patients, it was possible to pre-operatively plan a PCI drilling path that passes through the facial recess and avoids both the facial nerve and chorda tympani and targets the cochlea [2]. For four of these patients, intra-operative validation of the PCI technique could not be performed due to various medical and logistical reasons. For those nine patients that have undergone intra-operative validation, it was possible to pass a 1 mm diameter drill bit along the pre-operatively planned path without touching the facial nerve and chorda tympani. Distance from the drill bit to the facial nerve and chorda tympani were measured using endoscope images. The mean and standard deviation of the closest distance from the drill bit to the facial nerve were 1.1 ± 0.3 mm. These were 1.2 ± 0.5 mm for the chorda tympani [2]. A limitation of the approaches developed is that manual registration initialization is required for registering the atlas and target images, especially when there are substantial differences in head size, position and orientation between the images. We have attempted to correct these position and orientation differences by first extracting a

surface representation of the skull in both the atlas and target images and then roughly aligning the two images using the 7-DOF (three translations, three rotations, and one isotropic scale) transformation that registers the extracted skull surfaces. However, we have observed that this approach is not always sufficient. Future work for robust initial registration could include the use of the shape feature-based registration methods we developed to achieve the Specific Aim 2 of this dissertation (see Chapter III). These methods, which we also summarize below, rely on extracting a surface of an anatomical structure in the images and aligning the images using the spin-image registration method [3]. The anatomy of interest is selected such that its surface representation has enough distinct local shape features that are also consistent across subjects. Finally, we have attempted to use the developed segmentation methods to automatically segment the structures of the ear in head CTs of neonatal subjects. We have observed that a careful and time-consuming manual atlas-to-target image registration is necessary for the developed methods to generate segmentations that are of acceptable quality. This manual alignment is necessary because there are large differences in the shape of the ear canal, ossicles and tympanic membrane compared to the shape of these structures in our pediatric atlas images. On the other hand, we have observed that the shape of the labyrinth is very similar both in size and shape to the labyrinth in our pediatric atlas images. As part of our future work, we suggest creating new atlas images for neonatal subjects and building new models of ear anatomy that are specific to those populations of subjects.

In chapter III, we present a method for automatic registration of two head CT scans that overlap only partially. The PCI technique requires the registration of the pre- and intra-operative images to map the pre-operatively computed drilling trajectory onto the intra-operative space. The pre-operative CT, typically acquired with a high quality scanner, has a FOV that only covers

a slab of the whole head, while the intra-operative CT, acquired with a low-dose portable scanner, has a FOV that covers almost the whole head. Because of operating room logistical constraints, the two CTs can also be acquired with different head orientations. The variation in head orientation and position is much larger than the capture range of standard intensity-based image registration methods. The inconsistent FOV results in the exclusion of regions of the patient's head, which prevents the use of standard orientation matching techniques such as alignment of the images by principal components analysis. So far, the intra-operative registration has been performed by manually translating and rotating one of the scans or by selecting three or more homologous points on the scans and then aligning the scans using the transformation that registers the homologous points. Automating the registration is important because (a) it eliminates the need for someone who is expert in ear anatomy and in using our PCI planning software to be available for every surgery, and (b) the intra-operative registration is a time-critical process that must be completed before the next step of the PCI technique can be undertaken. Since this is a critical bottleneck, manual intervention is often stressful as the extra time required to perform this step may prolong the surgical intervention. The method we developed herein permits automatic alignment of the CTs, despite the large differences in FOV, position and orientation of the images. The method relies on a spin image (shape feature)-based registration method that, to the best of our knowledge, has not been used by the medical imaging community. We found the spin image technique to be efficient and accurate. Our approach has two main steps. First, we perform a coarse spin image feature-based registration. Next, we refine this by performing a standard intensity-based registration. The coarse registration subroutine is also a multistep process. First, we automatically register the intra-operative CT to a reference CT using the transformation that registers the cortical surface of the intra-operative CT, extracted

automatically, to the cortical surface of the reference CT, extracted offline. The reference CT has also been registered offline to the pre-operative CT. The cortical surfaces are registered by matching the set of spin-images extracted on the subset of surface points for which the local curvature is high. We perform the spin image-based registration via an intermediate reference CT in which almost the full head is inside the FOV, rather than directly to the pre-operative CT, because the spin image-based registration method is sensitive to differences in FOV. The approach of registering the two CTs via an intermediate reference CT is a new approach that, to the best of our knowledge, has not been presented before. Registration results are quantitatively validated by comparing trajectories mapped using the proposed automatic method to corresponding trajectories mapped using the semi-automatic methods, which has been used in ongoing clinical validation. The maximum distance between trajectories is 0.18 mm. As both mapping methods use the same intensity-based registration approach as the final registration step and converge to similar results, it is likely that both approaches produce equally accurate results. To date, this technique has been successfully used to accomplish intra-operative registration to perform one adult PCI surgery. We compared the time in minutes the proposed automatic registration approach takes to that of the registration approach when an expert performs the manual initialization on a set of four pre- and intra-operative image pairs. Each pre- and intra-operative image pair is registered by two experts and the automatic approach we propose. The automatic approach consistently takes less time to perform the registration compared to the registration initialized by the two experts. On average the automatic approach takes 0.78 minutes while the first and second expert take 2.05 and 1.18 minutes, respectively, on an Intel dual quad-core 2.93 GHz Xeon processor, Windows Server 2008 based machine. The results indicate that the proposed automatic registration method on average reduces the intraoperative registration

time by 61.9% and 33.9% when compared to the first and second expert, respectively. One challenge we encountered with our approach is that in some rare cases the lateral portion of the cortical surface (the surface of the skull/brain interface) is not inside the field-of-view of the intra-operatively acquired CTs. This prevents us from obtaining a surface of the brain/skull interface, as the missing portion of the brain/skull interface leads to leakage when we use our level set-based cortical surface extraction method. This could result in poor or incorrect initial registration results. To deal with this particular issue, we suggest, as future work, to focus on aligning the volumes by extracting features on directly on the images. For instance, extracting intensity-based spin-image features [4]. Finally, the proposed automatic registration method could be applied to any registration problem with large misalignment. The developed method could also be used to coarsely align volumes that overlap only partially. Although it will not replace existing accurate registration methods, such as methods that are based on optimization of mutual information or normalized mutual information between image pairs [5]-[7], it could be used for initialization. As discussed above, the proposed automatic registration method could also be tailored to further automate the atlas-to-target image registration which is a necessary first step to achieve automatic segmentation of pediatric ear anatomy in a target image. One limitation of the spin image method is that it is not robust to scale difference between shapes. Future direction will include the investigation of techniques that address this issue as well as the quantitative or qualitative validation of the developed method for general registration problems.

In chapter IV, we present new algorithms for automatic segmentation of the intra-cochlear anatomy in post-implantation CT of unilateral cochlear implant recipients. The image-guided cochlear implant programming strategies we developed require the determination of the spatial relationship between the implanted electrodes and intra-cochlear anatomy for individual

CI recipients. Until now, it has been possible to segment the intra-cochlear structures in a pre-implantation CT, localize the electrodes in a post-implantation CT, and register the two CTs to determine the spatial relationship between the electrodes and the intra-cochlear anatomy. The algorithms we propose herein do not require a pre-implantation CT. The approach is to segment the labyrinth, which we use as a landmark, in the contralateral normal ear and use its position to segment the intra-cochlear anatomy in the implanted ear by exploiting the intra-subject symmetry in inner ear anatomy. Specifically, we first segment the intra-cochlear anatomy (the scala tymapni, scala vestibuli, and the spiral ganglion) and the labyrinth in the normal ear by fitting their ASMs to the normal ear images. Next, we project the segmented surfaces from the normal ear to the implanted ear using the transformation that registers the two ear region images. Finally, we segment the implanted ear by first iteratively registering the projected labyrinth surface to the implanted ear images, and then using the compounded transformation to map the projected intra-cochlear surfaces. At each iteration of the surface-to-image registration process, we first find an optimal position for each point on the surface by finding the position, along the surface normal at that point, for which the 1-D gradient profile extracted, along the surface normal at that position, is the closest to the set of exemplar 1-D gradient profiles of that point, created offline from a multitude of training images. Next, we project the surface using the transformation that rigidly registers the surface to the optimal positions. To constrain the surface-to-image registration process to rely more on points that are far from implant related image artifacts, we designed a novel function to assign a reliability weight to each point on the surface. The cochlear implant appears very bright in CT, thus those points near this high intensity region are unreliable and are assigned a low weight. The weight assigning function is updated at each iteration to assign weights to points based on the updated position of the labyrinth. Symmetry

studies we performed suggest that the inner ear anatomy is highly symmetric, and thus substantiate our approach for segmentation of the implanted ear using segmentations obtained from the contralateral normal ear. To the best of our knowledge, this is the first attempt to segment the intra-cochlear anatomy in post-unilateral-implantation CTs using information in those images alone. It is also of note that this is the first time that contralateral mapping is proposed for segmentation of anatomical structures with a high degree of accuracy. We believe that our approach can be used for segmentation of general structures that exhibit bilateral symmetry, for instance the eye or the pelvis. We validated the segmentation results obtained using the algorithms presented herein on 30 ears for which both a pre- and post-implantation CT are available. We achieve an average intra-cochlear anatomy segmentation error of 0.22 mm. In addition to the accurate localization of intra-cochlear structures, the electrodes-position dependent programming strategies we developed require the accurate determination of the position of implanted electrodes. We achieve this task using parametric Gradient Vector Flow snakes-based algorithms we previously developed [8]. We tested these algorithms on seven post-implantation CTs and overall mean and maximum errors in localizing CI electrode arrays are 0.195 and 0.826 mm, respectively [8]. This level of accuracy is, as we describe below, accurate enough for our image-guided programming technique. We thus use these algorithms in conjunction with intra-cochlear anatomy segmentation algorithms to create programming plans that are specific to individual CI recipients. In previous work where we used both a pre- and a post-implantation CT for determining the spatial relationship between implanted electrodes and intra-cochlear anatomy, we achieved an average intra-cochlear anatomy segmentation error of 0.15 mm, and have obtained excellent programming results for more than 60 patients that have participated in our ongoing clinical valuation study (85% of those recipients have reported

substantial improvement in hearing). The intra-cochlear anatomy segmentation error we obtained using the method presented herein that does not require a pre-implantation CT is slightly higher. But, we anticipate that this slight reduction in segmentation accuracy will not negatively impact the improvements in hearing outcomes we obtain with using electrodes position-based CI programming strategies. The developed algorithms can now be used to create programming plans for unilateral CI recipients for whom a CT has not been acquired prior to implantation, thereby extending our electrodes position-dependent programming techniques to a larger population of CI recipients. Future work will involve assessing hearing improvements for programming plans created using only post-implantation CTs and the algorithms presented herein.

In chapter V, we present algorithms that permit the automatic segmentation of the intra-cochlear anatomy in post-implantation CT images directly without requiring a pre-implantation CT of either ear in those images. Our approach for segmentation relies on first approximating the shape of the labyrinth by mapping a labyrinth surface that is selected from a library of such surfaces, and then refining this shape by performing a weighted active shape segmentation with an ASM built specifically for the target CT. We then segment the intra-cochlear anatomy by fitting their shape model to the external wall of the cochlea established on the segmented labyrinth. Proof-of-concept experiments we performed show that our approach for coarse segmentation, which is to map library surfaces, is feasible and that our criterion for choosing similar shapes in our library is highly correlated with errors we obtain in segmenting inner ear anatomy by mapping the chosen similar shapes. To the best of our knowledge, this is the first attempt to segment the intra-cochlear anatomy in post-bilateral-implantation CTs directly without requiring a pre-implantation CT of either ear. Our shape library-based segmentation

framework is similar to multi-atlas-based segmentation methods. These methods typically involve registration of each atlas volume to the target volume and computation of a chosen similarity measure, based on the registration information, to determine the similar atlases. As the approach we developed herein does not require atlas-to-target volume registrations to be performed, it has computational advantage over standard multi-atlas segmentation methods. To the best of our knowledge, our similar subject selection approach has not been presented before. The approach is to first identify a portion of the structure-of-interest in a target image and then to compute a shape similarity measure between the identified portion and the corresponding portion of the structure-of-interest established on each library subject (atlas). The similar subjects (atlases) are then selected as those (subjects) atlases for which the similarity to the target subject is the largest. We have evaluated the developed method by comparing automatically generated intra-cochlear segmentations to gold-standard intra-cochlear segmentations using two groups of testing image pairs. The first is a group of 9 conventional pre- and post-implantation CT pairs; the second is a group of 16 conventional pre- and low-dose post-implantation CT pairs. We achieve sub-millimetric errors at all points on the segmentation surfaces, and intra-cochlear anatomy segmentation error averages 0.144 mm over nine conventional post-implantation CTs, 0.209 mm over 16 low-dose post-implantation CTs, and 0.186 mm overall 25 post-implantation CTs. These are 0.146, 0.200, and 0.181 mm for the AR, the "important" part of the intra-cochlear structures. The results we achieve on the post-implantation CTs are comparable to those that are achievable on pre-implantation CTs and this indicates that the approach we propose herein is accurate enough for use in position-based sound processing strategies. Our approach shows superior accuracy when employed on conventional CTs than low-dose CTs. In our validation study, the overall segmentation results we achieve using the initialization method alone are very

close to those that we achieve by refining the initial results using the refinement method. This shows that the initialization method alone can be used to achieve the task of segmenting the intra-cochlear structures. The refinement method generally improves the already good results we achieve using the initialization method. Specifically, for 84% of our validation set (21 out of 25), the refinement method has led to a reduction in initial segmentation error. However, for the other four validation cases, superior results have been obtained when using the segmentation initialization method alone. We have observed that for these four cases, the similar shapes selected from our library do not localize the target intra-cochlear structures as well as they do the "far" region of the labyrinth (the semi-circular canals and vestibuli). Future work will focus on addressing this issues as well as exploring techniques for automatically determining when to use the initialization method alone to achieve the segmentation task. Future direction could also include the expansion of our shape library, which is likely to lead to higher chance of finding similar shapes. Future work will also include using the proposed algorithms for determining electrode position-dependent programming strategies for bilateral CI recipients with no pre-implantation CT initially at our institution and, subsequently, at other institutions.

It is of note that the shape library-based intra-cochlear anatomy segmentation method can also be used to segment intra-cochlear anatomy in post-implantation CT of unilateral CI recipients. Although both the contra-lateral mapping (the algorithms developed for the third aim of this dissertation) as well as the shape library-based methods can localize the intra-cochlear anatomy with mean surface distance error of ~0.2 mm, we cannot say which method is better before we study the performance of the methods on a large validation dataset. The algorithms and systems developed to achieve the aims of this dissertation focus on CT images. However, we believe that the developed frameworks can be employed to achieve the same goals on images

acquired with other modalities, for instance Magnetic Resonance. As the gray-level appearance of the structures-of-interest in images varies across modalities, intensity or gradient models of the structures-of-interest we created need to be re-generated for other modality images. Finally, although the algorithms and methods presented in this dissertation are not final solutions, we believe that the proposed approaches have made substantial contributions towards automating the segmentation and identification of the structures of the ear and automating registration of CT images, which we believe are valuable for automating and extending the use of image-guidance for otologic surgery and cochlea implant programming to a larger population of patients.

References

[1]. Jeffery, N. and Spoor, F., Prenatal growth and development of the modern human labyrinth. Journal of Anatomy, 204: 71–92, 2004.

[2]. Balachandran R1, Reda FA, Noble JH, Blachon GS, Dawant BM, Fitzpatrick JM, Labadie RF, “Minimally Invasive Image-Guided Cochlear Implantation for Pediatric Patients: Clinical Feasibility Study”, Otolaryngology Head Neck Surg. 2014 Jan 21 (**In Press**).

[3]. A. E. Johnson, M. Hebert, “Surface matching for object recognition in complex three-dimensional scenes,” Image and Vision Computing, vol. 16, no. 9–10, pp. 635-651, Jul. 1998.

[4]. Lazebnik, S.; Schmid, C.; Ponce, J., "A sparse texture representation using affine-invariant regions," Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on , vol.2, no., pp.II-319,II-324 vol.2, 18-20 June 2003 doi: 10.1109/CVPR.2003.1211486.

[5]. W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, “Multi-modal volume registration by maximization of mutual information,” *Med. Image Anal.*, vol. 1, no. 1, pp. 35-51, Mar. 1996.

[6]. F. Maes, A. Collignon, D. Vandermeulen, G. Mrchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Trans. Med. Imaging*, vol. 16, pp. 187-198, Apr. 1997.

[7]. C. Studholme, D.L.G. Hill, and D.J. Hawkes, “An overlap invariant entropy measure of 3D medical image alignment,” Pattern Recognition 32(1):71–86, January 1999.

[8]. Jack H. Noble, Theodore A. Schuman, Charles G. Wright, Robert F. Labadie, Benoit M. Dawant, “Automatic identification of cochlear implant electrode arrays for post-operative assessment”, Proc. SPIE 7962(796217), Medical Imaging 2011.