



Published in final edited form as:

Med Image Comput Assist Interv. 2012 ; 15(Pt 2): 421–428.

Statistical Shape Model Segmentation and Frequency Mapping of Cochlear Implant Stimulation Targets in CT

Jack H. Noble¹, René H. Gifford², Robert F. Labadie³, and Benoît M. Dawant¹

Jack H. Noble: jack.h.noble@vanderbilt.edu; René H. Gifford: rene.h.gifford@vanderbilt.edu; Robert F. Labadie: robert.labadie@vanderbilt.edu; Benoît M. Dawant: benoit.dawant@vanderbilt.edu

¹Dept. of Elect. Eng. and Comp. Sci., Vanderbilt University, Nashville, TN 37235, USA

²Dept. of Hearing and Speech Sci., Vanderbilt University, Nashville, TN 37235, USA

³Dept. of Otolaryngology – Head & Neck Surg., Vanderbilt University, Nashville, TN 37235, USA

Abstract

Cochlear implant (CI) surgery is considered standard of care treatment for severe hearing loss. CIs are currently programmed using a one-size-fits-all type approach. Individualized, position-based CI programming schemes have the potential to significantly improve hearing outcomes. This has not been possible because the position of stimulation targets is unknown due to their small size and lack of contrast in CT. In this work, we present a segmentation approach that relies on a weighted active shape model created using microCT scans of the cochlea acquired ex-vivo in which stimulation targets are visible. The model is fitted to the partial information available in the conventional CTs and used to estimate the position of structures not visible in these images. Quantitative evaluation of our method results in Dice scores averaging 0.77 and average surface errors of 0.15 mm. These results suggest that our approach can be used for position-dependent image-guided CI programming methods.

Keywords

Spiral ganglion; modiolus; cochlear implant; weighted active shape model segmentation

1 Introduction

Cochlear Implants (CIs) are considered standard of care treatment for severe-to-profound sensory-based hearing loss. CIs restore hearing by applying electric potential to neural stimulation sites in the cochlea with an implanted electrode array. An audiologist programs the CI by determining how it maps the spectrum of detected sound frequencies to the set of electrodes for each recipient. Programming is performed based solely upon patient response to judgments of perceived loudness for individual electrodes. This process can be difficult and time intensive, and the majority of potentially adjustable parameters are typically left at the default settings determined by the manufacturer. Because of this one-size-fits-all approach, standard programming methods may not result in optimal hearing restoration for all recipients.

In the natural hearing process, ear anatomy mechanically translates sound into vibrations of the basilar membrane, which separates the two principal cavities of the cochlea, the scala tympani and the scala vestibuli (see Figure 1). These vibrations stimulate nerve cells connected to the spiral ganglion (SG) and, eventually, the auditory nerve. Researchers have extracted the tonotopic mapping between the frequency of the sound and the SG nerve cells that are stimulated, i.e., higher frequencies lead to stimulation of more basal SG nerve cells, whereas, lower frequencies stimulate more apical SG nerve cells [1]. CIs exploit this natural

tonotopy by applying an electric field to more apical (basal) SG nerve cells to induce perceived lower (higher) frequency sounds. It is widely believed that programming schemes that use more explicit knowledge of electrode position hold great promise for improving hearing outcomes and could potentially be transformative to the field of CI audiology.

Since methods for identifying the position of implanted electrodes have already been established [2], the major obstacle for position-dependent programming is that there have been no techniques for accurately identifying the stimulation targets – the SG nerve cells. Identifying the SG *in vivo* is difficult because the nerve bundles have diameter on the order of microns and are too small to be visible in CT, which is the preferred modality for cochlear imaging due to its otherwise superior resolution (see Figure 1). However, the external walls of the cochlea are well contrasted in CT. Since the cochlea wraps around the SG, and, as shown in [3], external cochlear anatomy can be used to estimate the location of intra-cochlear anatomy using a statistical shape model (SSM); it is possible that a similar external anatomy driven SSM can be used to estimate the SG. In this work, we test such an approach for automatic segmentation and frequency mapping of the SG for computer assisted CI programming. The rest of this paper presents our approach.

2 Methods

The data set we have used in this study consists of images of six cadaveric cochlea specimens. For each specimen, we have acquired one μ CT image volume with a SCANCO μ CT. The voxel dimensions in these images are 36 μ m isotropic. We also have acquired CTs of the specimens with a Xoran xCAT[®] fpVCT scanner. In these volumes, voxels are 0.3 mm isotropic. In each of the μ CT volumes, the scala vestibuli, scala tympani, and SG were manually segmented. Figure 1 shows an example of a conventional CT image and its corresponding μ CT image.

2.1 Overview

Since the SG lacks any contrast in CT, we cannot segment it directly. The goal with our approach is to use the location of external cochlear features to predict the position of the SG. To do this, we have constructed a SSM of cochlear anatomy that includes the SG. Prior to constructing the SSM, we identify which points in the model will correspond to strong cochlear edges in CT. To those points we arbitrarily assign an importance weighting of 1. To all others we assign a lesser weighting of 0.01. These weights are used to construct a point distribution model (PDM) for weighted active shape model (wASM) segmentation [4]. The SSM is built as a standard PDM computed on the registered exemplar point sets. To segment a new image, the SSM is iteratively fitted in a weighted-least-squares sense to features in the target image. The edge points with their weight of 1 are fitted to strong edges in the CT. The non-edge points with low weight are fitted to the positions determined by non-rigid registration with an atlas image. With the weights that we have chosen, the non-edge points provide enough weak influence on the optimization to ensure that the wASM stays near the atlas-based initialization position, while the edge points strongly influence the whole wASM towards a local image gradient-based optimum for a highly accurate result. During model construction, the set of SG points in the model that interface with intra-cochlear anatomy were also identified. These points are referred to as the active region (AR) since they correspond to the region most likely to be stimulated by an implanted electrode. The tonotopic mapping of each point in the AR in the reference volume is computed based on angular depth using known equations [1]. Once a segmentation is completed, the tonotopic frequency labels from the model can be transferred to the target image. The following subsections detail these methods.

2.2 Model Creation

To model cochlear structures, we: (1) establish a point correspondence between the structures' surfaces, (2) use these points to register the surfaces to each other with a 7 degrees of freedom similarity transformation (rigid plus isotropic scaling), and (3) compute the eigenvectors of the registered points' covariance matrix. Point correspondence is determined using the approach described in [3]. Briefly, non-rigid registration is used to map each of the training volumes to a reference volume, and any errors seen in the results are manually corrected. Then, a correspondence is established between each point on the reference surface with the closest point in each of the registered training surfaces. Once correspondence is established, each of the training surfaces is point registered to the reference surface. Since the cochlear edge points will be the highest weighted points for the wASM segmentation, identical weights are used to register the training shapes in a weighted-least-squares sense using standard point registration techniques [5] prior to computation of the eigenspace so that the model will best capture the shape variations at these points.

To build the model, the principal modes of shape variation are extracted from the registered training shapes. This is computed according to the procedure described by Cootes et. al. [6]: First, the covariance matrix of the point sets' deviation from the mean shape is computed as

$$C = \frac{1}{N} \sum_{j=1}^N (\vec{v}_j - \bar{v})(\vec{v}_j - \bar{v})^T, \quad (1)$$

where the \vec{v}_j 's are the individual shape vectors and \bar{v} is the mean shape. The shape vectors are constructed by stacking the 3D coordinates of all the points composing each structure into a vector. The modes of variation in the training set are then computed as the eigenvectors $\{\vec{u}_j\}$ of the covariance matrix,

$$\{\lambda_j, \vec{u}_j\}_{j=1}^{N-1} : \lambda_j \vec{u}_j = C \vec{u}_j. \quad (2)$$

These modes of variation are extracted for the combined shape of the scala tympani, scala vestibuli, and SG for all the samples in the training set.

2.3 Weighted Active Shape Segmentation

The procedure we use for segmentation with a wASM follows the traditional approach, i.e., (1) the model is placed in the image to initialize the segmentation; (2) better solutions are found while deforming the shape using weighted-least-squares fitting; and (3) eventually, after iterative shape adjustments, the shape converges, and the segmentation is complete. Initialization is performed using the atlas-based methods proposed in [3].

Once initialized, the optimal solution is found using an iterative searching procedure. At each search iteration, an adjustment is found for each model point, and the model is fitted in a weighted-least-squares sense, as described below, to this set of candidate adjustment points. To find the candidate points, two approaches are used. For cochlear edge points, candidates are found using line searches to locate strong edges. At each external point \vec{y}_i , a search is performed along the vector normal to the surface at that point. The new candidate point is chosen to be the point with the largest intensity gradient over the range of -1 to 1 mm from \vec{y}_i along this vector. For all other points, it is impossible to determine the best adjustment using local image features alone because there are no contrasting features at these points in CT. Therefore, the original initialization positions for these points, which

were provided by atlas-based methods, are used as the candidate positions. The atlas-based result, as our results will show, is sufficiently accurate to provide this useful information to the segmentation process.

The next step is to fit the shape model to the candidate points. We do this in the conventional wASM manner. A standard 7 degree of freedom weighted point registration is performed, creating similarity transformation T , between the set of candidate points $\{\vec{y}'_i\}$ and the mean shape $\{\vec{v}_j\}$, where \vec{v}_j are the 3D coordinates of the j th point in the mean shape. Then, the residuals

$$\vec{d}_i = T(\vec{y}'_i) - \vec{v}_i \quad (3)$$

are computed. To obtain the weighted-least-squares fit coordinates in the SSM's eigenspace, we compute,

$$\vec{b} = (U^T W^T W U)^{-1} U^T W^T W \vec{d}, \quad (4)$$

where \vec{d} is composed of $\{\vec{d}_i\}$ stacked into a single vector, $U = [\vec{u}_1 \vec{u}_2 \dots \vec{u}_{N-1}]$ is the matrix of eigenvectors, and W is a diagonal matrix with the importance point weightings in the appropriate entries along the diagonal. This equation results in a vector \vec{b} that represents the coordinates in the SSM space corresponding to a weighted-least-squares fit of the model to the candidate points. The final approximation to the shape is computed by passing the sum of the scaled eigenvectors plus the mean shape through the inverse transformation, equivalently,

$$\vec{y}'_i = T^{-1}(\vec{w}_i + \sum_{j=1}^{N-1} b_j \vec{u}_{j,i}), \quad (5)$$

where $\vec{u}_{j,i}$ is the i th 3D coordinate of the j th eigenvector. As suggested by Cootes, the magnitude of the b_j 's are constrained such that

$$\sqrt{\sum_{j=1}^{N-1} \frac{b_j^2}{\lambda_j}} \leq 3, \quad (6)$$

which enforces the Mahalanobis distance between the fitted shape and the mean shape to be no greater than 3.

At each iteration, new candidate positions are found and the model is re-fitted to those candidates. The wASM converges when re-fitting the model results in no change to the surface. The tonotopic mapping of the SG points in the model, computed when the model was built, are directly transferred to the target image via the corresponding points in the converged solution. A result of this is shown in Figure 1f.

Validation—Segmentation was performed on CT's of the cochlea specimens using a leave-one-out approach, i.e., the volume being segmented is left out of the model. For one of the six specimens a CT was not available, and it was used as the model reference volume to simplify the leave-one-out validation study. Thus, the validation study measures segmentation error on the remaining five specimens when using PDMs with four modes of variation. Because these samples were excised specimens, rather than whole heads, the atlas-

based initialization process required manual intervention—however, in practice the approach is fully automatic. To validate the results, we again exploit the set of corresponding μ CT volumes. Each CT was rigidly registered to the corresponding μ CT of the same specimen. The automatic segmentations were then projected from the CT to μ CT space. Finally, Dice index of volume overlap [7] and surface errors were computed between the registered automatic segmentations and the manual segmentations to validate the accuracy of our results. The distributions of surface error we report include distances in the forward (automatic-to-manual) direction as well as the reverse (manual-to-automatic) direction. A smaller distance between the automatic and manually segmented surfaces indicates higher accuracy.

3 Results

Quantitative comparisons between manual and automatic segmentations of the SG are shown in Figure 2. The Dice index and bi-directional mean/max surface distances were computed between each pair of automatic and manual segmentations. Figure 2 shows the overall distributions of these recorded values. Surface errors were recorded between the whole SGs (WSG) and also between the active regions (AR). Dice indices were not computed for the AR because it is not a closed surface and does not represent a volumetric region. The green bars, red bars, blue rectangles, and black I-bars denote the median, mean, one standard deviation from the mean, and the overall range of the data set, respectively. As can be seen in the figure, the wASM achieves mean dice indices of approximately 0.77. For typical structures, a Dice index of 0.8 is considered good [8]. Here, we consistently achieve Dice indices close to 0.8 for segmentation of a structure that is atypically small and lacks any contrast in the image. Mean surface errors are approximately 0.15 mm for both the WSG and the AR, which is about a half a voxel's distance in the segmented CT. Maximum surface errors are above 1 mm for the WSG but are all sub-millimetric for the AR.

Segmentations for all 5 experiments are shown color encoded with surface error in Figure 3. It can be seen that the wASM results in mean surface errors under 0.15 mm for the majority of the SG with average maximum errors of about 0.7 mm (<3 voxels). As can be seen in the figure, errors in the AR above 0.5 mm are rare and highly localized. Shown in Figure 4 are contours of a representative automatic segmentation overlaid with the CT (the volume on which segmentation was performed) and the corresponding registered μ CT. It can be seen from the figure that the contours achieved by automatic segmentation of the CT are in excellent agreement with contours manually delineated in the high resolution μ CT, especially in the AR. Localization errors that are apparent in the μ CT are less than 2 voxels width in the CT.

4 Conclusions

In this work, we have presented a weighted active shape based approach for identifying the SG, which lacks any contrast in conventional CT. In this approach, we rely on high resolution images of cadaveric specimens to serve two functions. First, they provide information necessary to construct an SSM of the structure, permitting segmentation of the structure in conventional imaging. Second, the high resolution images are used to validate the results. This is performed by transferring the automatically segmented structures from the conventional images to the corresponding high resolution images and comparing those structures to manual segmentations.

This work has shown that it is possible to accurately identify the location of the SG in conventional CT because the position of the SG varies predictably with respect to the cochlea. The approach we present accurately locates the SG by attracting the exterior walls of the models of intra-cochlear anatomy towards the edges of the cochlea. This approach

achieves dice indices of approximately 0.77 and sub-millimetric maximum error distance in the active region, which represents the region of interest for CI stimulation. These results suggest that our segmentation approach is accurate enough to be used for position-dependent, image-guided CI programming methods.

Testing of computer assisted CI programming approaches using the presented methods has begun and has been completed on one patient thus far. The fully-automatically identified positions of the electrode and tonotopically mapped SG for this patient are shown in Figure 1f. For that patient, the settings that were originally considered optimal by the audiologist were modified using image-based information. Improvements in the patient's hearing were statistically significant using a binomial distribution statistic for the individual speech perception metrics tested [9–10]. The patient's monosyllabic word recognition scores (a quantitative measure of hearing performance) jumped from 33% to 84%, and sentence recognition performance in noise at +10 dB signal-to-noise ratio increased from 46% to 83%. Further, the patient reported significant improvement in hearing and overall sound quality. While very preliminary, these results indicate that image-based programming, enabled by the approach described herein, may significantly improve hearing restoration for CI users.

Acknowledgments

This research has been supported by NIH grants R01DC008408, R01DC009404, and R21DC012620. The content is solely the responsibility of the authors and does not necessarily represent the official views of this institute.

References

1. Stakhovskaya O, Spridhar D, Bonham BH, Leake PA. Frequency Map for the Human Cochlear Spiral Ganglion: Implications for Cochlear Implants. *J Assoc Res Otol.* 2007; 8:220–233.
2. Noble, JH.; Schuman, A.; Wright, CG.; Labadie, RF.; Dawant, BM. Automatic Identification of Cochlear Implant Electrode Arrays for Post-Operative Assessment. In: Dawant, BM.; Haynor, DR., editors. *Medical Imaging 2011: Image Processing; Proc. of the SPIE Conf. on Med. Imag.*; 2011. p. 796217
3. Noble JH, Labadie RF, Majdani O, Dawant BM. Automatic segmentation of intra-cochlear anatomy in conventional CT. *IEEE Trans on Biomedical Eng.* 2011; 58(9):2625–2632.
4. Rogers, M.; Graham, J. Robust Active Shape Model Search. *ECCV 2002, Part IV.* In: Heyden, A.; Sparr, G.; Nielsen, M.; Johansen, P., editors. *LNCS. Vol. 2353.* Springer; Heidelberg: 2002. p. 517-530.
5. Green BF. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika.* 1952; 17:429–440.
6. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models—Their Training and Application. *Comp Vis Image Unders.* 1995; 61(1):39–59.
7. Dice LR. Measures of the amount of ecologic association between species. *Ecology.* 1945; 26:297–302.
8. Zijdenbos AP, Dawant BM, Margolin R. Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. *IEEE Trans on Med Imag.* 1994; 13(4):716–724.
9. Thornton AR, Raffin MJ. Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res.* 1978; 21:507–518. [PubMed: 713519]
10. Spahr AJ, Dorman MF, Litvak LL, Van Wie S, Gifford RH, Loizou PC, Loiselle LM, Oakes T, Cook S. Development and Validation of the AzBio Sentence Lists. *Ear Hear.* 2012; 33:112–117. [PubMed: 21829134]

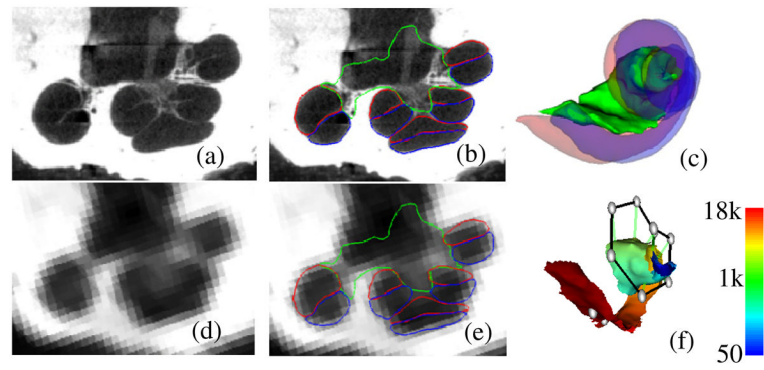


Fig. 1. Shown in (a) and (d) are a slice of a μ CT and a CT of a human cochlea. In (b) and (e), the scala tympani (red), scala vestibuli (blue), and bundle of nerve cells of the SG (green) are delineated in the same slice. These structures are shown similarly in 3D in (c). In (f), the active region of the SG is colormapped with its tonotopy (Hz), and the positions of implanted electrodes are shown (gray spheres).

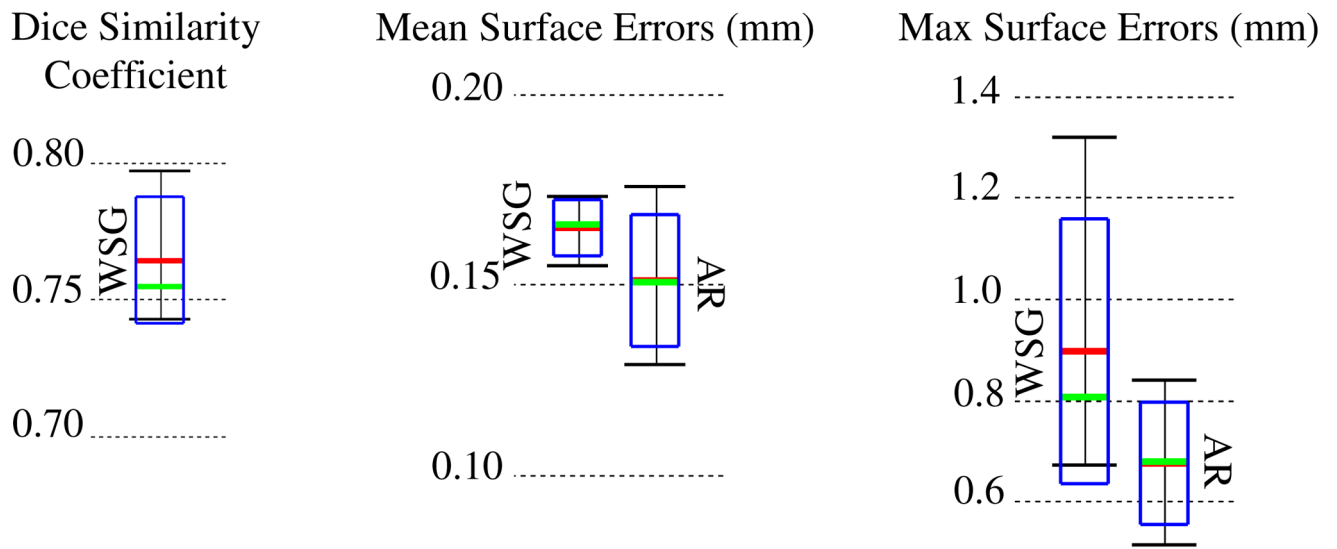


Fig. 2. Segmentation error distributions of Dice similarity scores for the whole SG (WSG) and mean and max symmetric surface error distributions for the WSG and in the active region (AR)

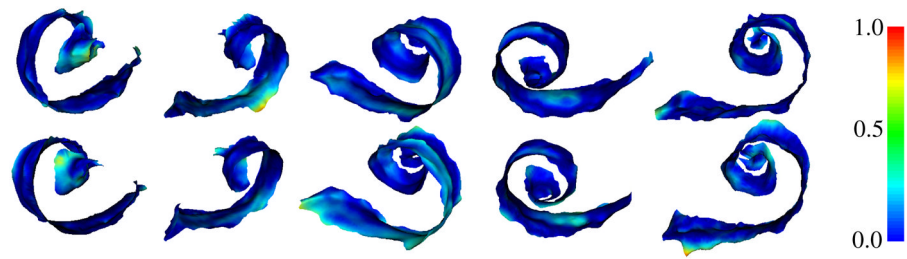


Fig. 3. Automatic (top row) and manual (bottom row) segmentations of the active region of the SG in the 5 test volumes (left-to-right) color encoded with error distance (mm)

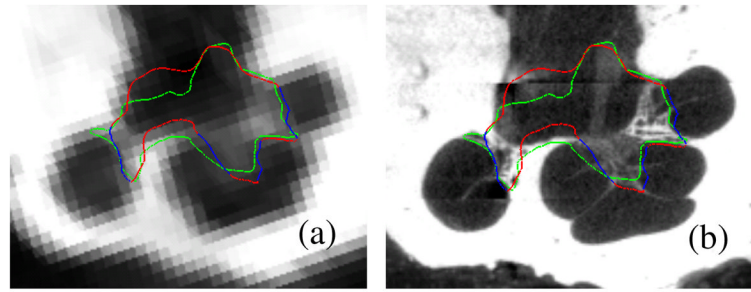


Fig. 4. Delineations of the automatic (red/blue) and manual (green) segmentation of the SG in the CT (a) and μ CT (b) slice from Figure 1. The active region is shown in blue