Analysis of middle ear morphology for design of a transnasal endoscope

Vu, Minh, Banalagay, Rueben, Zhang, Dongqing, Rivas, Alejandro, Fichera, Loris, et al.

Minh Vu, Rueben A. Banalagay, Dongqing Zhang, Alejandro Rivas, Loris Fichera, Robert Webster, Robert F. Labadie, Jack H. Noble, "Analysis of middle ear morphology for design of a transnasal endoscope," Proc. SPIE 10951, Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling, 109512S (8 March 2019); doi: 10.1117/12.2513841

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

Analysis of middle ear morphology for design of a transnasal endoscope

Minh Vu[a], Rueben A. Banalagay[a], Dongqing Zhang[a], Alejandro Rivas[b], Loris Fichera[c], [d]Robert Webster, Robert F. Labadie[b], [a]Jack H. Noble

[a]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235, USA
[b]Department of Otolaryngology – Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[c]Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609, USA
[d]Department of Mechanical Engineering, Vanderbilt University, Nashville, TN 37235, USA

ABSTRACT

Cholesteatomas are benign lesions that form in the middle ear (ME). They can cause debilitating side effects including hearing loss, recurrent ear infection and drainage, and balance disruption. The current approach for positively identifying cholesteatomas requires intraoperative visualization either by lifting the ear drum or transmitting an endoscope through the ear canal and tympanic membrane – procedures which are typically done in and operating room with the patient under general anesthesia. We are developing a novel endoscope that can be inserted trans-nasally and could potentially be used in an outpatient setting allowing clinicians to easily detect and visualize cholesteatomas and other middle ear conditions. A crucial part of designing this device is determining the degrees of freedom necessary to visualize the regions of interest in the middle ear space. To permit virtual evaluation of scope design, in this work we propose to create a library of models of the most difficult to visualize region of the middle ear, the retrotympanum (RT), which is located deep and posterior to the tympanic membrane. We have designed a semi-automated atlas-based approach for segmentation of the RT. Our approach required 2-3 minutes of manual interaction for each of 20 cases tested. Each result was verified to be accurate by an experienced otologist. These results show the method is efficient and accurate enough to be applied to a large scale dataset. We also created a statistical shape model from the resulting segmentations that can be used to synthesize new plausible RT shapes for comprehensive virtual evaluation of endoscope designs and show that it can represent new RT shapes with average errors of 0.5 mm.

Keywords: Middle ear, retrotympanum, sinus tympani, principal component analysis, coherent point drift

INTRODUCTION

Cholesteatomas are benign lesions that form in the middle ear. They can cause debilitating side effects including hearing loss, recurrent ear infection and drainage, and balance disruption if not diagnosed and treated early. Surgical resection of these lesions is the standard of care treatment. But, after initial resection, cholesteatomas have a high propensity to recur. While clinicians desire direct visualization of the middle ear upon first suspicion of recurrence, taking a second look typically requires general anesthesia and access of the middle ear by either puncturing the ear drum and inserting an endoscope and/or lifting the ear drum and visualizing with a microscope. Follow-up surgery is performed in approximately 60% of cases within the first year, yet the recurrence rates are approximately 25% [1]. Our group is working to develop a trans-nasal middle ear endoscope which could reach the middle ear through the Eustachian tube [2]. Such a scope could be used in the clinic and permit a less invasive and less expensive approach for visualizing the middle ear. The scope is currently in the design stage and operating requirements need to be determined to ensure that it has the degrees of freedom necessary to visualize critical sections of the middle ear. The middle ear has a number of concavities that are difficult to visualize with the most difficult being the retrotympanum (RT) (shown in Figure 1a), which lies deep and posterior to the tympanic membrane and includes the sinus tympani (ST) and subtympanic sinus. In this work, our goal is to use a database of CT images to construct a point distribution model of the RT representing population variability in its shape. This model can then be used to determine the degrees of freedom necessary for an endoscope to visualize the RT for cholesteatoma surveillance.

Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling,
edited by Baowei Fei, Cristian A. Linte, Proc. of SPIE Vol. 10951, 109512S · © 2019 SPIE
CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2513841

The RT is subject to great variability in size and posterior extent [3]. The ST portion of the RT is classified according to three main types of morphology based on its depth (Figure 2) [4]. Type A is a small ST. In these cases, the ST does not extend medial or posterior to the facial nerve as it descends within the mastoid. Type B is a deep ST where

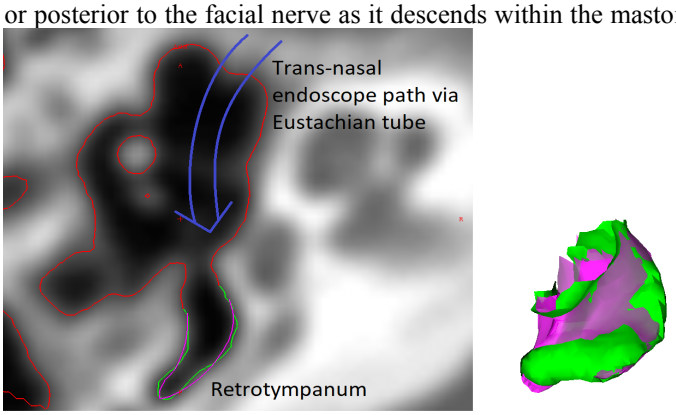

Figure 1: Axial CT slice (left) and 3D rendering (right) of a patient with a Type C ST. The red contour defines the borders of the middle ear and is created via thresholding the masked image. The green contour is the retrotympanum (RT) region. The magenta contour is the atlas RT non-rigidly registered to this subject. The trans-nasal endoscope will roughly follow the purple path from the Eustachian tube to visualize the RT cavity.

the deep boundary lies medially with respect to the facial nerve but does not extend posteriorly to the facial nerve. Type C is a deep ST with posterior extension, where the medial boundary of the ST lies medial and posterior to the facial nerve as it descends within the mastoid. It is necessary for our RT model to be representative of the population containing each ST type. In this work we propose an approach for segmenting the RT in CT scans to construct a standard statistical shape model of the RT representing population variance [5]. Our long-term aim is to localize the RT in hundreds of images from our clinical database for our model, and thus we need an approach that is semi-automated.

METHODS

While our long term goal is to produce a dataset of hundreds of RT exemplars for population analysis, in this preliminary study we utilized CT scans from a limited set of 20 patients from an IRB approved database to construct our model. In each of these target volumes, first, the middle ear region containing the RT was segmented. To do this, an initial middle ear segmentation was created using atlas-based techniques. Using mutual information-based techniques, an atlas was affinely and non-rigidly registered to the target using the process described in [6]. The middle ear, which has been manually delineated in the atlas, was projected through the resulting non-rigid registration onto the target image to serve as an initial segmentation. Errors in the registration are unavoidable due to topological differences in the pneumatization patterns of the temporal bone between individuals. Thus, this initial segmentation was imperfect. To obtain a better segmentation, we dilated the initial one, used the result to mask the CT volume, and thresholded the resulting masked CT volume. Because the middle ear border is a bone-air interface, thresholding was quite robust, and the mask provided by the atlas permits thresholding the image locally in the middle ear region. The mask dilation step was done using a spherical structuring element with the radius needed to compensate for the registration errors and ensure that the resulting mask includes the middle ear boundaries around the RT. The radius was between 3-5 mm depending on the case. An example result of this process is shown in Figure 1 in red. The segmented middle ear is then sorted into each of the three different types shown in Figure 2.

Once the middle ear was segmented, we defined the portion of it that corresponds to RT. To do this, VTK-based [7] 3D surface visualization software developed in house was used to select a point to represent the deepest part of the RT concavity. Then, a breadth-first traversal of the surface vertices through the surface triangle graph was used to label the vertices of the middle ear surface corresponding to the RT. The depth of the traversal was chosen unique to each case to ensure the resulting labels accurately represent the extent of the RT. For more complicated cases where the

RT was not radially symmetric about a single point, this process was used iteratively, combing several label sets, to result in an accurate overall labelling of the RT region of the middle ear. An RT surface can be defined by the patch of

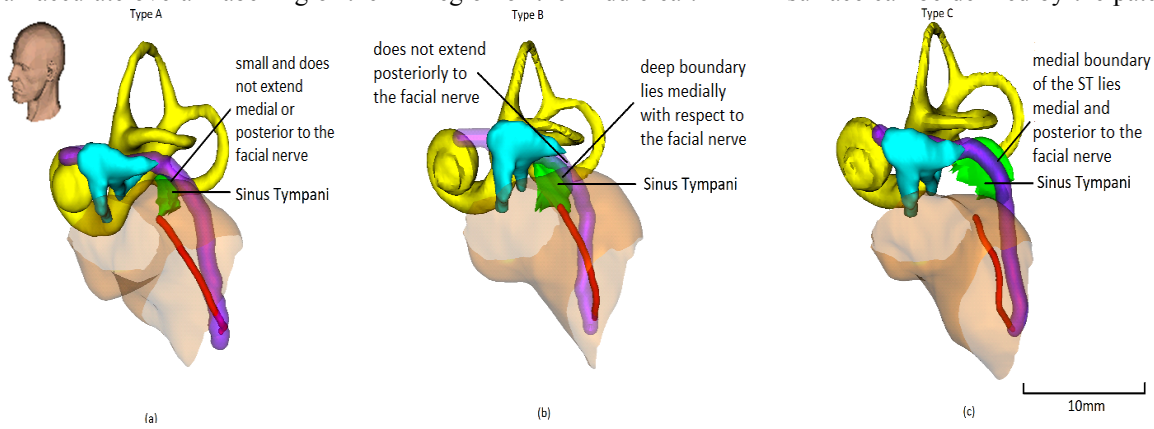

Figure 2: Three sinus tympani (ST) types. For each image, the ST is shown in green, facial nerve in magenta, cochlea in yellow, ossicles in cyan, chorda tympani in red, and external auditory canal in orange.

triangles in the middle ear surface that contain the vertices labelled as belonging to RT. An example result of this process is shown in Figure 1 in green.

After the RT was localized, we established a point correspondence across cases to construct a point distribution model. To do this, first, one RT surface was chosen randomly as the reference and was nonrigidly registered to the remaining RT surfaces using the coherent point drift algorithm [8]. Then, the closest point on each RT surface to each point in the registered reference surface was found to define a one-to-one point correspondence between the reference and remaining RT surfaces.

The statistical shape model was constructed using the standard approach proposed by Cootes [5]. First, the N pointsets $\{\overrightarrow{v_i}\}_{i=1}^{N}$ were rigidly registered [9] to the reference pointset $\overrightarrow{v_1}$ and a mean shape was estimated as

$$\bar{v}' = \frac{1}{N}\sum_{i=1}^{N} \Phi_i(\overrightarrow{v_i}) \,, \tag{1}$$

where Φ_i is the registration transformation between $\overrightarrow{v_i}$ and $\overrightarrow{v_1}$. Then, the pointsets are iteratively registered to the mean shape while re-estimating the mean shape at each iteration until the process converges to mean shape $\bar{v}$. Then, the coordinates of the shapes are concatenated into vectors $\{\overrightarrow{x_i}\}$, and we find the covariance matrix:

$$C = \frac{\sum_i^N (x_i - \bar{x})(x_i - \bar{x})^T}{N}. \tag{2}$$

The non-zero eigenvalues and corresponding eigenvectors of C are denoted $\{\lambda_i\}_{i=1}^{N-1}$ and $\{\overrightarrow{p_i}\}_{i=1}^{N-1}$, where $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{N-1}|$. Assuming the resulting point distribution model successfully captures population variance and that shape variability is well approximated by a normal distribution, a transnasal endoscope can be designed that can adequately visualize RTs with shape that falls within δ standard deviations to population average if it can be proven that the scope can successfully visualize RTs in the space $\bar{v} + P\vec{d} \;\; \forall \;\; \vec{d} \in D$, where P is the matrix of eigenvectors $\{\overrightarrow{p_i}\}_{i=1}^{N-1}$, and D is the space of all shape parameter vectors $\vec{d}$ that satisfy $\sum_{i=1}^{N-1} \frac{d_i^2}{|\lambda_i|} \leq \delta^2$. Using this framework, monte carlo simulations could be conducted in which $\vec{d}$ are randomly sampled from D and a given scope design can be validated virtually by ensuring the set of synthesized RTs are visualizable using, e.g., ray casting techniques such as we have done in prior work [10].

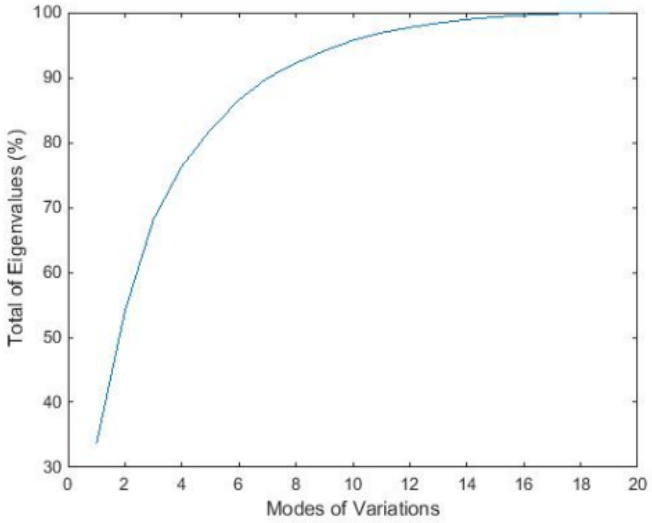

Figure 3: Total variation vs # of modes of variation.

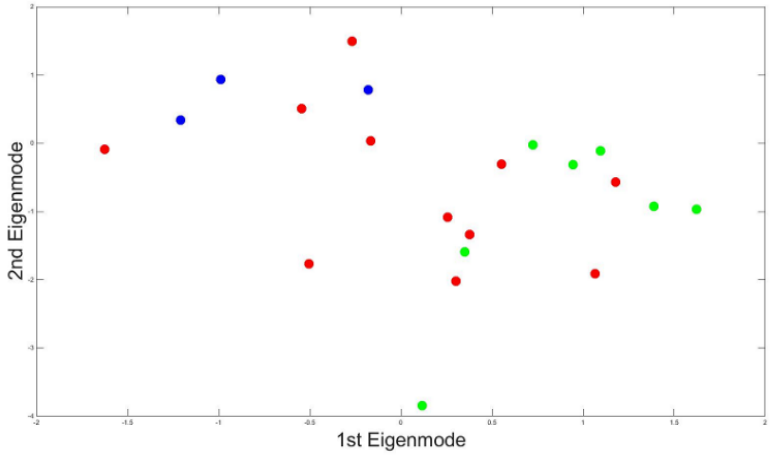

Figure 4: The first two shape space parameters for each case divided by the corresponding $\sqrt{\lambda_i}$. Type A, B, and C cases are shown in green, red, and blue.

RESULTS

The RT was successfully localized in our dataset of 20 volumes. Performing all manual interaction steps in each case required approximately 2-3 minutes. Approximately 10 minutes of total processing time was required for each case using a Windows server with an Intel Xeon X5570 2.9 GHz CPU. Each RT localization was reviewed by an experienced otologist and confirmed to be accurate. The rater also determined the ST classification for each case. All three classes of ST were represented in our set of models, with 7 Type A, 10 Type B, and 3 Type C cases. Figure 2 shows an example case from our dataset of each of the three classes.

Figure 3 shows the percent contribution of each eigenmode to the total variance captured by the point distribution model. We see that 95% of the total variance of the model is contained within the first 10 eigenmodes. This suggests that our dataset of 20 cases may reasonably well capture population variance in the RT. The first two

eigenmodes, which capture 54.05% of the variation, are visualized in Figure 5. The mean shape is shown in panel (a),

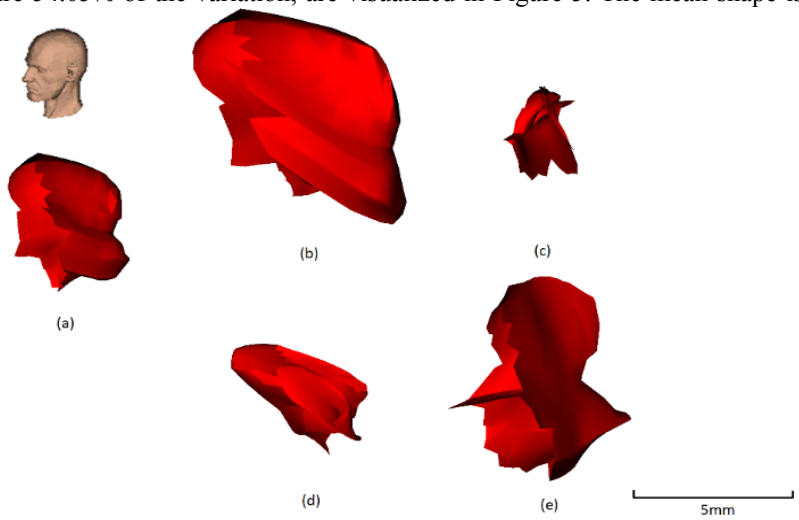

Figure 5: The effects of the first two modes of variation. The mean shape (a) plus (b) and minus (c) 3 standard deviations of the first eigenvector; and plus (d) and minus (e) 3 standard deviations of the second eigenvector.

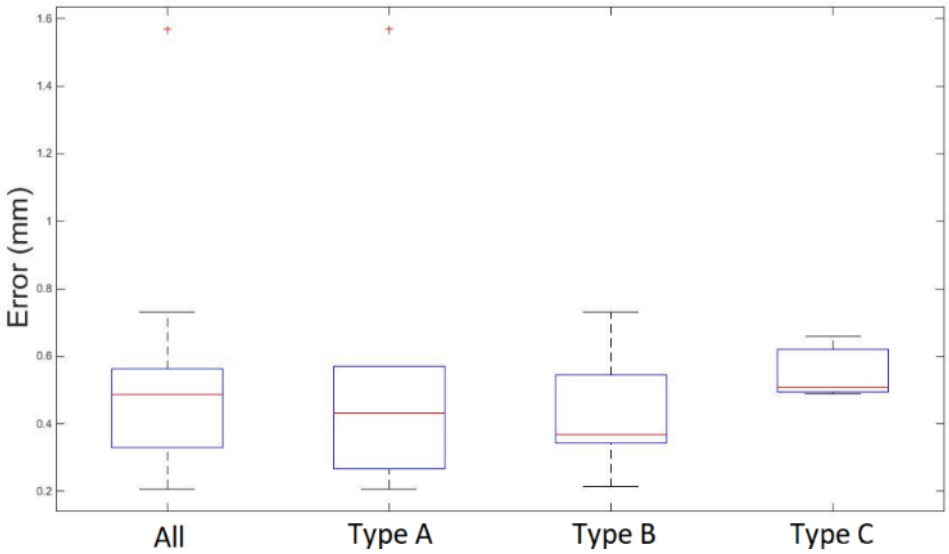

Figure 6: Boxplots of mean leave-one-out model fitting errors for all cases as well as by ST type.

and ± 3 standard deviations along eigenmode 1 are shown in (b) and (c) and eigenmode 2 in (d) and (e). The 2nd eigenmode appears to capture extent of the RT in the superior-inferior direction, while the first eigenmode appears to principally capturing overall scale of the RT. The shape space parameters for the first two eigenmodes of the cases in our dataset are visualized in a scatter plot in Figure 4. It can be seen that Type A cases are clustered to the lower right of the plot while type C cases are on the upper left. This distinction is likely due to how type A is usually very shallow while type C is much deeper, and the first two eigenmodes are principally capturing the overall size and depth of the RT. Type B cases, which can vary greatly in size and shape, do not appear separable by the first two eigenmodes but might be separable when considering more dimensions of the eigenspace.

We also computed leave-one-out model fitting errors in which the 10 most significant eigenmodes were fit to the left out exemplar, and point-to-corresponding-point errors were measured between the fitted model and the exemplar. 10 eigenmodes were used because we see that 95% of the total variance of the full model is contained within the first 10 eigenmodes (Figure 3). The resulting fitting errors are shown in Figure 6, which shows mean overall errors to be 0.5 mm, and with Type C errors being slightly higher at 0.6 mm. One outlier Type A case resulted in relatively high mean error of 1.6 mm. These results suggest that our model is generally effective at capturing population variance of the RT but could benefit from more training samples to improve robustness.

CONCLUSIONS

To the best of our knowledge, this is the first work proposed for semiautomatic localization of the RT and the first work in which a statistical shape model of the RT is created. These results are critical to aid the design and validation of a new trans-nasal endoscope for middle ear visualization. A statistical shape model of the RT will permit synthesizing new statistically plausible RT shapes and virtual evaluation of endoscope designs with Monte Carlo simulations. To construct the shape model, we have designed a semi-automatic procedure to efficiently localize the RT in CT volumes. The process takes around 2-3 minutes for each case and requires little input, which shows its applicability for larger scale studies. We are currently using this process on an expanded dataset to generate a larger database of RT exemplars to ensure that our RT statistical shape model properly represents population variance. Of particular importance is the Type C class of ST due to its difficulty in adequately visualizing. In our current set of 20 cases, we have only three cases with this morphology, and our current model results in slightly higher fitting errors for this type. Other future work will include evaluating the use of the statistical shape model for segmenting the RT and automatically classifying ST type by clustering the shape space parameters.

ACKNOWLEDGEMENTS

This research has been supported in part by a Vanderbilt University School of Engineering undergraduate summer research fellowship and grant R21DC16153 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] M. G. Crowson, V. H. Ramprasad, N. Chapurin, C. D. Cunningham, and D. M. Kaylie, “Cost analysis and outcomes of a second-look tympanoplasty-mastoidectomy strategy fr cholesteatoma,” Laryngoscope, vol. 126, no. 11, pp. 2574–2579, 2016.

[2] Fichera L, Dillon NP, Zhang D, Godage IS, Siebold MA, Hartley BI, Noble JH, Russell PT, Labadie RF, Webster RJ. Through the Eustachian Tube and Beyond: A New Miniature Robotic Endoscope to See Into the Middle Ear. IEEE Rob. and Autom. Letters. 2017 Jul;2(3):1488-94.

[3] Donaldson J.A., Anson B.J., Warpeha R.L., Rensink M.J., “The Surgical Anatomy of the Sinus Tympani,” Arch Otolaryng - Vol 91 (3), pp. 219-227, 1970.

[4] Marchioni D., Valerini S., Mattioli F., Alicandri-Ciufelli M., “Radiological assessment of the sinus tympani: temporal bone HRCT analyses and surgically related findings,” Surg Radiol Anat (2015) 37:385–392.

[5] T.F. Cootes, C.J.Taylor, D.H. Cooper, and J. Graham: Active Shape Models – Their Training and Application. Computer Vision and Image Understanding, Vol.61, No. 1, January, pp. 38-59, 1995

[6] Noble, J.H., Dawant, B.M., Warren, F.M., Labadie, R.F., "Automatic Identification and 3D Rendering of Temporal Bone Anatomy," Otol & Neurotol., 30(4):436-42, 2009.

[7] Schroeder, Will; Martin, Ken; Lorensen, Bill (2006), The Visualization Toolkit (4th ed.), Kitware, ISBN 978-1-930934-19-1

[8] Andriy Myronenko, Xubo Song: Point Set Registration: Coherent Point Drift. IEEE Transactions on pattern analysis and machine intelligence, vol. 32, NO. 12, December 2010

[9] K. S. Arun, T. S. Huang, S. D. Blostein, “Least-Squares Fitting of Two 3-D Point Sets,” IEEE Transactions on Pattern Analysis and Machine Intelligence . Volume 9 Issue 5, May 1987 Pages 698-700

[10] Bennett M.L., Zhang D., Labadie R.F., and Noble J.H., “Comparison of middle ear visualization with endoscopy and microscopy,” Otology & Neurotology, 37(4):362-6, 2016.