EECE593 Winter 2002 Project Report
Vocal Tract Visualisation

David Pritchard

April 9, 2002

1 Overview

For my project, I have created a 3D reconstruction of the human vocal tract. The intention of this project was to practice the volume image processing techniques learned in class, while working on a real problem.

Data for the project was obtained from a set of Magnetic Resonance Imaging scans published by Olov Engwall [EB99]. Like many linguistics researchers, he used this data to calculate the area of the vocal tract along its length. This area data can then be used to understand how the vocal tract forms different sounds. Acquiring these MRI scans is a slow process, typically requiring 40 seconds or more. Consequently, vowels are typically imaged instead of consonants. Engwall does describe the capture of consonants and fricatives, but I am uncertain of how this was achieved.

Engwall's data was separated into three sets: 18 slices in the transverse plane (through the pharynx), 18 tilted slices (through the velum), and 14 slices in the coronal plane (through the mouth and face). This data was captured in a single scan.

For my reconstruction, I registered these data sets in order to determine their relative positions. I then resampled the data and merged it into a single 3D volume. Finally, I had to segment the data in order to separate the vocal tract from the rest of the head.

2 Previous Work

The strange three-part data acquisition used by Engwall was apparently initiated by Demolin et al. [DMS96], but their paper gives no justification for this choice of scanning technique.

Cox [Cox96] provided a simple overview of image intensity registration techniques. Others have used feature-based registration, but this approach has the disadvantage of limited precision: subvoxel registration is not possible due to the error in locating features. [Cox96] described work by Woods [WCM92], Hajnal [HSS$^+$95] and Friston [FAF$^+$95]. Woods' error metric is based upon voxel ratios, but is said to be hard to program for multidimensional minimisation. Hajnal's error metric is based upon the L^2 mismatch and, while slow, uses standard techniques readily available in commerical optimisation packages. Friston's method also uses L^2 mismatch but sets up a linear system, and is said to be quite quick.

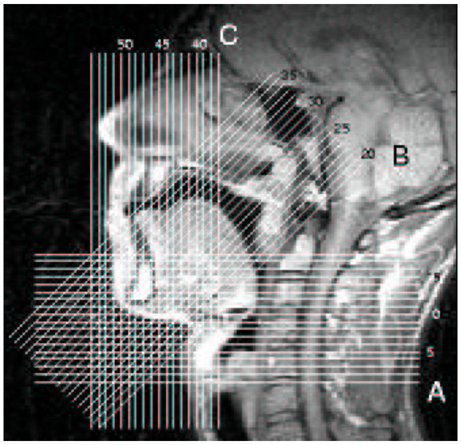


Figure 1: The 54 slices used as input to the data. A: pharynx, B: tilted, C: coronal.

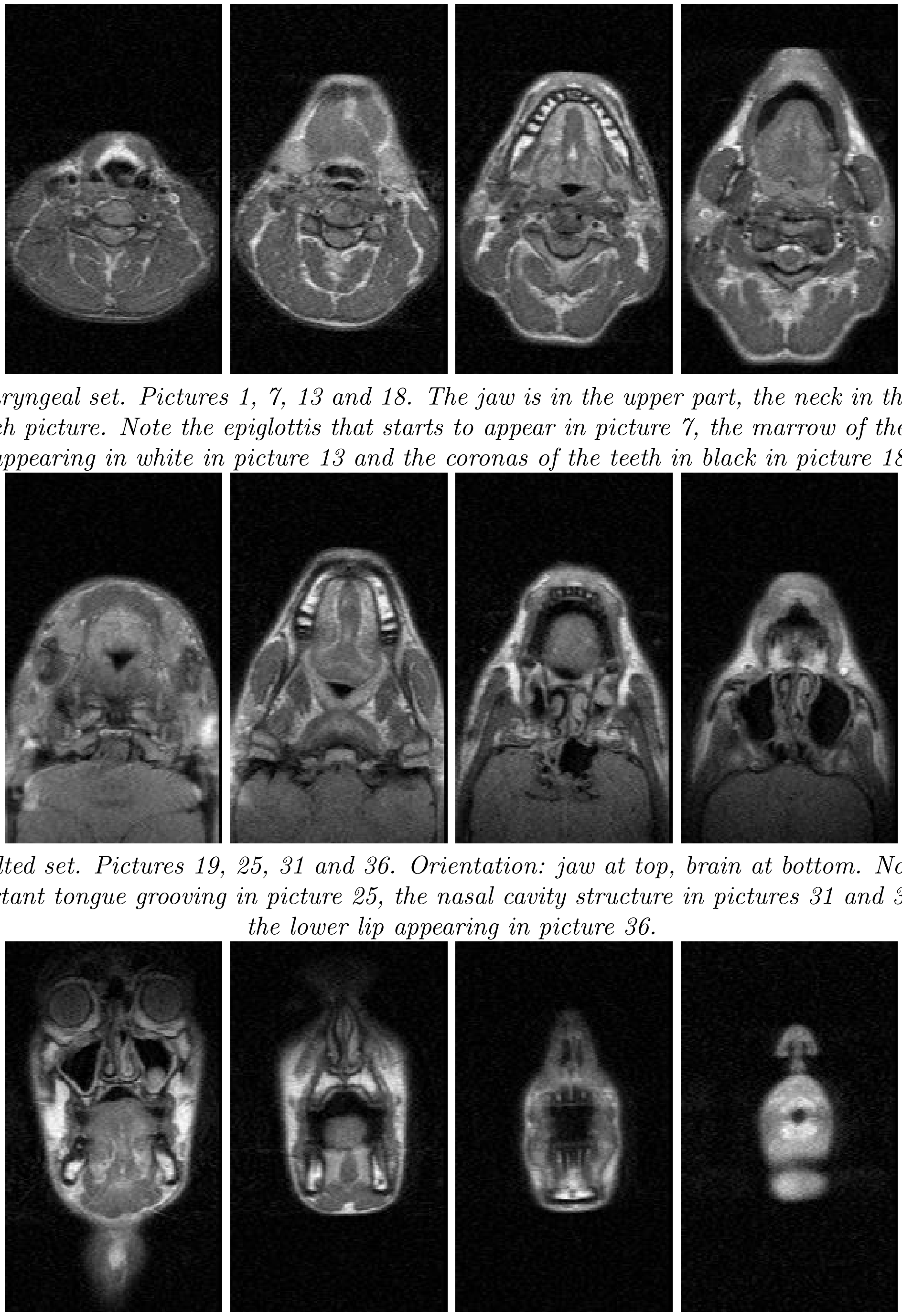
To date, most segmentation approaches have been fairly simple. Soquet et al. [SLM$^+$98] compared existing techniques: manual curve placement, thresholding, a threshold-zoom method, and an elastic curve method. They concluded that the thresholding techniques came closest to the manual curve approach. Demolin et al. [DLM$^+$98] used manual curve placement. Badin et al. [BBRS98] and Engwall [EB99] used thresholding. Mathematical morphology is not mentioned in the literature; most researchers seem to rely upon manual or semi-automatic segmentation. Tom et al. [TTHS99] and Story et al. [STH96] employed an automatic slice-by-slice region-growing technique with several manual steps with reasonable success, but they do not explain how they seed the region growing algorithm. They also used EBCT images, which do not have the same tooth-related problems as MRI images.

For reconstruction, [TTHS99] and [STH96] advocate Raya and Udupa's shape-based interpolation method [RU90]. However, the primary motivation for this choice seems to be processing difficulties, since hardware at that time had difficulty displaying volume data at interactive rates. [EB99] used a similar technique. Baer et al. [BGGN91] used voxel-based reconstruction, but lacked fast hardware for interactive visualisation.

3 Data

Engwall's data is shown in figures 1 and 2. These figures and captions were taken from his paper, although his paper describes the vowel /I*p*I/, while I used the data for the Swedish *sj* fricative of *sjutton*. As noted in Engwall's comments regarding the figures, calcified structures such as the teeth appear the same as air passages.

There were a few problems with the dataset. The pharynx set had one or two slices with substantially different intensities. This was corrected using histogram equalisation. The intensities of the different sets also varied, and an intensity-adjustment term had to be included in the registration process to deal with this.



A. Pharyngeal set. Pictures 1, 7, 13 and 18. The jaw is in the upper part, the neck in the lower in each picture. Note the epiglottis that starts to appear in picture 7, the marrow of the teeth appearing in white in picture 13 and the coronas of the teeth in black in picture 18.

B. Tilted set. Pictures 19, 25, 31 and 36. Orientation: jaw at top, brain at bottom. Note the important tongue grooving in picture 25, the nasal cavity structure in pictures 31 and 36 and the lower lip appearing in picture 36.

C. Coronal set. Pictures 37, 41, 45 and 48.

Figure 2: Excerpt from the 3D image set for Swedish *sj* of *sjutton*. Air passages and calcified structures appear in black, hydrogen rich parts, such as fat and marrow, in white.

4 Approach

4.1 Registration

After reviewing an overview paper [Cox96], I chose to perform registration according to the procedure outlined by [HSS+95]. In this approach, registration can be achieved at a sub-voxel level by defining an error function.

$$E = \sum_x (I_1(\mathbf{x}) - \beta \mathbf{I_2}(\mathbf{T}[\mathbf{x}]))^2$$

This error defines the error between volumes I_1 and I_2, where the second volume has been transformed by some transformation $\mathbf{T}$ and some intensity adjustment β. Using the Levenberg-Marquardt algorithm, the optimal value of $\mathbf{T}$ and β can be found, achieving the registration task. For my registration, I parameterised $\mathbf{T}$ using five parameters, three for translation and two Euler angles. For truly correct operation, the rotation should be parameterised using a quaternion with full freedom of movement. However, registering the rotational parameters was extremely slow, and for this dataset there appeared to be very little rotational registration required after the initial guess provided by Engwall.

4.2 Segmentation

Segmentation was performed in the space of the three original datasets, and the results were then merged with a union operation. Various techniques from mathematical morphology were used, including closing, opening, labelling and the watershed transform. The segmentation is almost entirely automatic, with only two marker points required as input: these are used to separate the nasal cavity from the vocal tract, and are currently hard-coded. Unfortunately, only one vocal tract shape was available, so the segmentation process is unlikely to be very general. More work is needed in this area. Sections of the segmentation process that are likely to be fragile are documented in the source code.

There are also some issues regarding the choice of segmentation approach. I chose to perform segmentation in the space of the three original datasets, with the segmented results then transformed into the pharynx co-ordinate system, where they were merged. The motivation for this approach was simple: the tilted dataset contains some very small features which could have been lost after rotating and resampling into the pharynx co-ordinate system. Without these fine details, segmentation would fail. However, this problem could have been avoided by performing all segmentation in the tilted co-ordinate system. The coarser-grained features in the pharynx and coronal datasets would not be lost, and the fine features in tilted dataset would not be resampled and would hence also be preserved. Unfortunately, there was insufficient time to attempt this approach. Consequently, there are visible boundaries between the datasets in the result, where a difference of segmentation approach can be seen.

5 Implementation

Matlab was used for the implementation. Current versions of the software are able to perform some volume transformations using the Imaging Toolbox, to perform Levenberg-Marquardt opti-

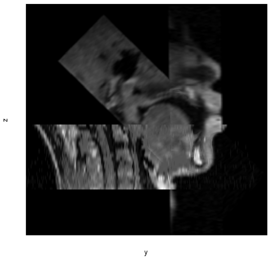


Figure 3: Mid-saggital view using a slice through the reconstructed registered volume data.

misation using the Optimisation Toolbox, and to construct and visualise isosurfaces from volume data using the marching cubes algorithm.

In addition, SDC Information Systems [Sys] publishes a Morphology Toolbox for Matlab that is capable of performing mathematical morphology operations on volume data. A 30-day trial license was used to perform this work. This toolbox is quite solid; I highly recommend it.

Matlab does have a number of shortcomings for volume visualisation. The imaging toolbox assumes that data is stored in matrix format, with low indices in the top-left of the image, and high indices in the bottom-right. For 3D data, this is inconvenient—it is more practical to use the Cartesian co-ordinate system. Consequently, I had to write special routines to display and transform the volume data.

Furthermore, for registration purposes, it is useful to be able to position volumes in space without lots of extra zero-padding. For this purpose, a special data structure was used, containing both the volume data and an origin representing the position of one corner of the data in 3D space. Finally, Matlab does not support translation of images/volumes by fractional pixel distances; I wrote a routine to do this (albeit only with linear interpolation).

6 Results

A mid-saggital view of the registered data is shown in Figure 3. Overlapping regions are displayed using the maximum operator for visualisation here. As can be seen, there are still some visible boundaries between the datasets, but the discontinuity is minimal. Most of these artifacts are due to intensity differences, not position differences, indicating that the registration was successful.

The final result includes the epiglottis (at the base of the pharynx), and the mouth cavity. Unfortunately, teeth are not registered in an MRI, and are hence indistinguishable from the vocal tract itself. Some researchers have made physical casts of the subjects' teeth in order to accurately remove this volume from the data. However, this was not possible for this project. Results after segmentation and reconstruction are shown in Figures 4 and 5.

In general, the results are reasonable. The shape of the vocal tract is definitely correct, although I have not evaluated the area function for comparison against other published data. There are visible artifacts in the reconstruction: some spherish blobs in the lower pharynx due to use of the greyscale morphological closing operator, and likewise in the coronal section. In the tilted section, where the most effort was expended on segmentation, the reconstruction looks

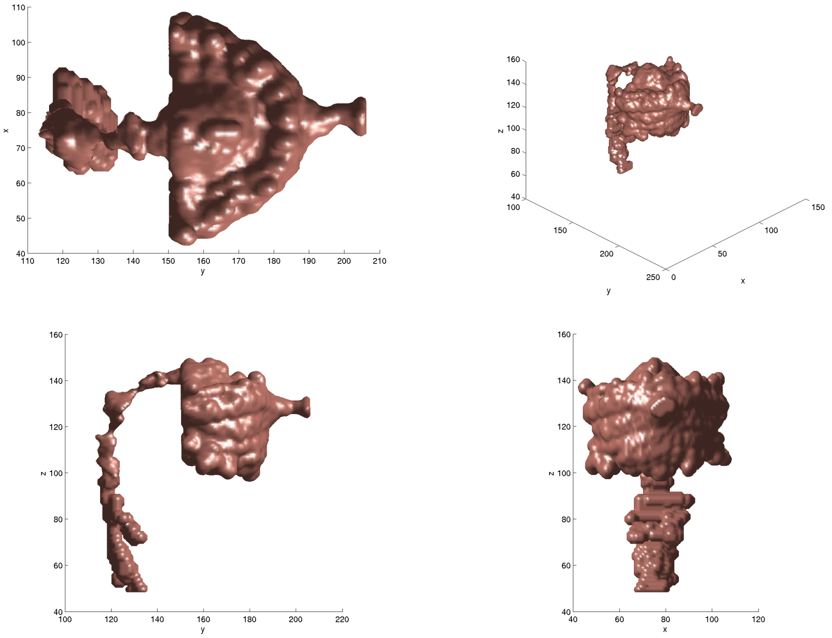


Figure 4: 3D reconstruction of the vocal tract.

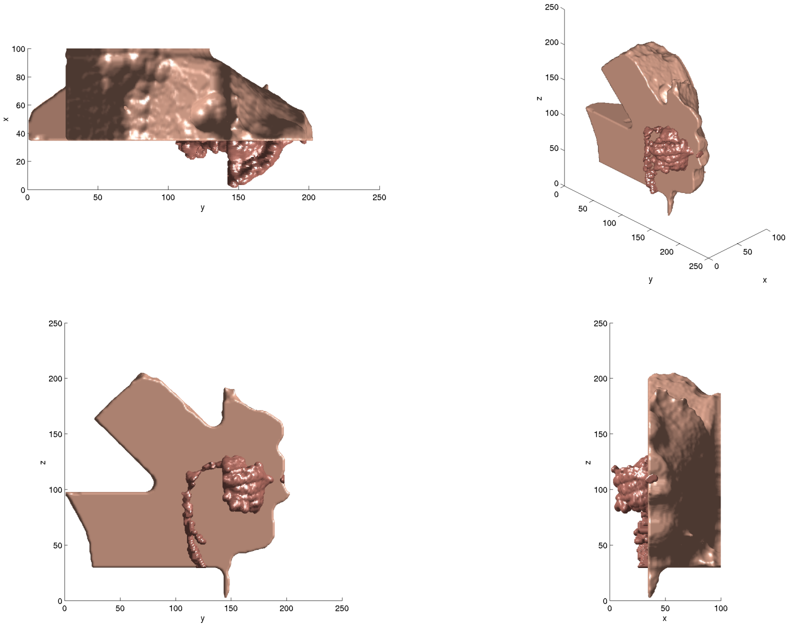


Figure 5: 3D reconstruction of the vocal tract, with a cutaway view of the head.

quite clean. For a real evaluation, however, area functions need to be computed and compared against published data.

References

[BBRS98] P. Badin, G. Bailly, M. Raybaudi, and C. Segebarth. A three-dimensional linear articulatory model based on MRI data. In *ESCA / COCOSDA International Workshop on Speech Synthesis*, pages 249–254, 1998.

[BGGN91] T. Baer, J. Gore, L. Gracco, and P. Nye. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *Journal of the Acoustical Society of America*, 1991.

[Cox96] R. Cox. Motion and functional MRI. In *Boston Workshop on Functional MRI*, 1996.

[DLM$^+$98] D. Demolin, V. Lecuit, T. Metens, B. Nazarian, and A. Soquet. Magnetic resonance measurements of the velum port opening. In *International Conference on Spoken Language Processing*, pages 425–429, 1998.

[DMS96] D. Demolin, T. Metens, and A. Soquet. Three-dimensional measurement of the vocal tract by MRI. In *International Conference on Spoken Language Processing*, 1996.

[EB99] O. Engwall and P. Badin. Collecting and analysing two- and three-dimensional MRI data for Swedish. *Kungl Tekniska Hogskalen Institutionen for Tal, Musick och Horsel Quarterly Progress and Status Report*, 1999.

[FAF$^+$95] K. Friston, J. Ashburner, C. Frith, J.-B. Poline, J. Heather, and R. Frackowiak. Spatial regis5tration and normalization of images. *Human Brain Mapping*, 3:165–189, 1995.

[HSS$^+$95] J. Hajnal, N. Saeed, E. Soar, A. Oatridge, I. Young, and G. Bydder. A registration and interpolation procedure for subvoxel matching of serially acquired MRI images. *Journal of Computer Assisted Tomography*, 19:289–296, 1995.

[RU90] S. Raya and J. Udupa. Shape-based interpolation of multidimensional objects. *IEEE Transactions on Medical Imaging*, 1990.

[SLM$^+$98] A. Soquet, V. Lecuit, T. Metens, B. Nazarian, and D. Demolin. Segmentation of the airway from the surrounding tissues on magnetic resonance images: A comparative study. In *International Conference on Spoken Language Processing*, pages 3083–3086, 1998.

[STH96] B. Story, I. Titze, and E. Hoffman. Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America*, 1996.

[Sys] SDC Information Systems. Internet, `http://www.mmorph.com`.

[TTHS99] K. Tom, I. Titze, E. Hoffman, and B. Story. 3-d vocal tract imaging and formant structure: Varying vocal register, pitch and loudness. *Status and Progress Report*, pages 101–113, 1999.

[WCM92] R. Woods, S. Cherry, and J. Mazziotta. Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, 16:620–633, 1992.