|  |
| --- |
| Summer Studentship: Vocal Tract Measurement |
| Refinement and proof of a measurement system for extracting length, area and volume data from MRI Images |
| By Helen Elizabeth Searle Supervisor: Dr. Catherine Watson Nov 2011-February 2012 |

Contents

[1. Introduction 2](#_Toc316398205)

[2. Preliminary Investigation 3](#_Toc316398206)

[2.1. Literature Review 3](#_Toc316398207)

[2.2. Approximations of Vocal Length from MRI images 3](#_Toc316398208)

[2.3. Brief formant analysis 3](#_Toc316398209)

[3. Refinement of MRI Measurement Methodology 4](#_Toc316398210)

[3.1. Vocal Tract Measurement 4](#_Toc316398211)

[3.2. Analysis of Results 4](#_Toc316398212)

[4. Results 5](#_Toc316398213)

[4.1. Purpose 5](#_Toc316398214)

[4.2. Method 5](#_Toc316398215)

[4.3. Results 5](#_Toc316398216)

[5. Conclusion/Future Work 21](#_Toc316398217)

[Appendix A: Literature Review of MRI Studies 22](#_Toc316398218)

[Appendix B: Preliminary Vocal Tract Length Approximation 30](#_Toc316398219)

[Appendix C: Formant Analysis 43](#_Toc316398220)

[Appendix D: Process for identifying oral and pharyngeal cavity markers. 44](#_Toc316398221)

[Appendix E: Node Placement Additional Information and Trouble-shooting. 50](#_Toc316398222)

[Appendix F: Matlab Code 56](#_Toc316398223)

[Appendix G: Speaker Data 57](#_Toc316398224)

[Appendix H: Additional Plots 58](#_Toc316398225)

# Introduction

The overall purpose of the study being conducted by Dr. Catherine Watson is to investigate the changes in the vocal tract; particularly the pharyngeal region, due to aging. The current focus is on analysis of this region during the sustained production of the 11 monophthongs vowels in New Zealand English.

Dr. Catherine Watson has investigated two different methods for measuring the vocal tract, firstly with an Acoustic Reflectometer (AR) and secondly, using Magnetic Resonance Imaging (MRI) and CMGui software. Previous summer student, Kalyan Chilukuri developed a method to calculate cross-sectional areas of the vocal tract from MRI images. The method was tested for accuracy; 6 subject data sets were processed and the results plotted. It was noted that the teeth and the lips were two areas which required further investigation as these had a significant impact on the accuracy of the area calculation. Another student, Chung Ting Justine Hui measured the vocal tract shape using acoustic reflectometry. One of the key issues identified by Justine was the difficulty in correctly identifying the velar-pharyngeal port.

The purpose of my summer studentship was to develop a reliable, repeatable and proven method for vocal tract measurement and analysis. This was achieved through three key stages. The first was a preliminary investigation of the existing methods and concepts used for vocal tract measurement and analysis. Secondly, the MRI measurement and analysis process was refined to make it more reliable and repeatable. Finally, the refined procedure was applied to eight sets of data from New Zealand English speakers to produce a set of results which is consistent with theory.

# Preliminary Investigation

The purpose of the preliminary investigation was to introduce key ideas, techniques and programs required to carry out this project. This included a literature review, approximations of vocal tract lengths and a brief formant investigation using existing AR data. By undertaking a preliminary investigation, the direction of the project which would provide the most valuable results was clarified.

### Literature Review

The motivation behind conducting the literature review was to gain a clear, comprehensive understanding of the methods used in previous studies which also employed MRI techniques for measurement of vocal tract dimensions. The literature review is provided in *Appendix A*. The knowledge gained from this literature review provided insight into other factors which may impact results and well as confirming that the MRI data to be used for this project was collected in a well-tested and proven manner.

(R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Preliminary Work\Literature Review of MRI Studies)

### Approximations of Vocal Length from MRI images

In this investigation rough measurements of the vocal tract length were taken from both AR and MRI data of the long vowels of four NZE speakers [*Appendix B*]. This investigation highlighted the need for a much more sophisticated measurement system. It also highlighted the fact that area and volume measurements are needed in order to compare between the oral and pharyngeal cavities.

(R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Preliminary Work\Vocal Tract Measurement.xlsx)

### Brief formant analysis

This investigation looked into how altering the pharyngeal volume affects the formants produced. This was achieved using the AR data in R 2.8.0 to get formant readings from different input volumes. The results of this investigation are provided in *Appendix C*.

(R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Preliminary Work\Formant Analysis.xlsx)

# Refinement of MRI Measurement Methodology

### Vocal Tract Measurement

The template for measuring vocal tracts was set up by Kalyan Chilukuri contains nine folders relating nine key steps (*figure 1*). Each of these folders contains “instruction” text files. A template of these folders can be found in R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Template. In order for these to produce consistent and reliable data sets, some refinements were needed.

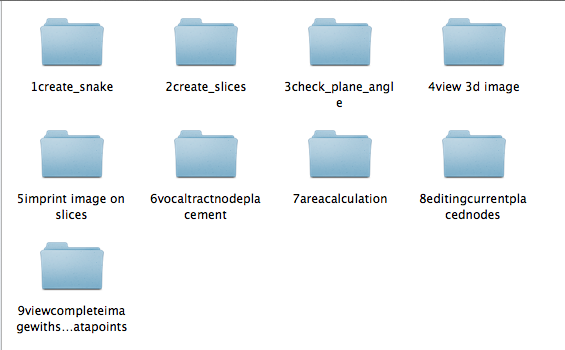


Figure : Folders in vocal tract measurement template

When first using CMGui it was important to read the “instructions” text files in each of these folders while using the full template to gain a clear understanding of the processes involved. Once this understanding was gained, only the four folders shown in *figure 2* are needed to produce sufficient data for vocal tract measurement and analysis.

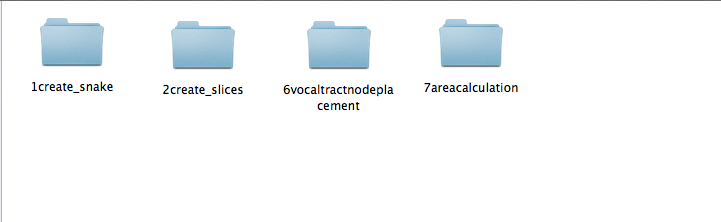


Figure : Important process folders for vocal tract measurement

The preliminary investigation indicated a need for a more sophisticated measurement system. The key points which were stressed were the need for clearly defined anatomical markers for the boundaries of the vocal tract. These are the start of the beginning and end of the oral and pharyngeal cavities; the lips, the velar-pharyngeal port, the glottis, and the boundaries of the cross-sections. These definitions are provided in *Appendix D* and *Appendix E.*

### Analysis of Results

The refinement of the method meant separate analysis of oral and pharyngeal cavities was required. This was performed using Matlab. The functions use distance and area functions developed by Kalyan Chilukuri. Explanation of the function of each of the .m files in each folder (e.g. VT01, VT02…) is provided in *Appendix F*.

# **Result**s

### Purpose

The aim of this experiment was to prove the method outlined in *section 3*.

### Method

The data used in this experiment was from eight different speakers of New Zealand English (NZE). During the MRI, speakers were supine and asked to produce each of the 11 monophthongs of NZE for a sustained period of time while the scan was taking place. 13 sagittal slices were taken for each speaker. The speakers were split into 3 different age categories, old (C), middle aged (M) and young (Y). The ages/ gender of the speakers are provided in *Appendix G.*

### Results

Shown in *figures 3 - 10* are plots of the distance vs. cross-sectional area with relative oral and pharyngeal volumes for speakers producing NZE Vowels. This is imported from Matlab figures (saved in each individual file e.g. R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Vocal Tracts from Varadha\VT01

Area data produced in step 7 of the vocal tract measurement of each vowel (e.g. R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Vocal Tracts from Varadha\VT01\Had\Oral\7areacalculation\area.txt) has been copied to a spreadsheet which can be found in R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Vocal Tracts from Varadha\Cross-sectional area.xlsx

The results of the volume calculation are provided in two forms. Firstly the comparison of the NZE vowels using an average ratio of oral and pharyngeal cavity volumes from all speaker data are found in *table 2* and *figure 12*. Secondly, the comparison between the ratio of oral and pharyngeal cavity volumes for vowel and each speaker (grouped in age brackets) are found in *table 3* and *figures 13-33.* The full data set for this information can be found in R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Volumes.xlsx.

Separate plots of the mid-sagittal images and comparative cavity volumes and lengths can be found in *Appendix H*.

#### Area vs. Distance Plot

#### 

Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT01 producing NZE Vowels

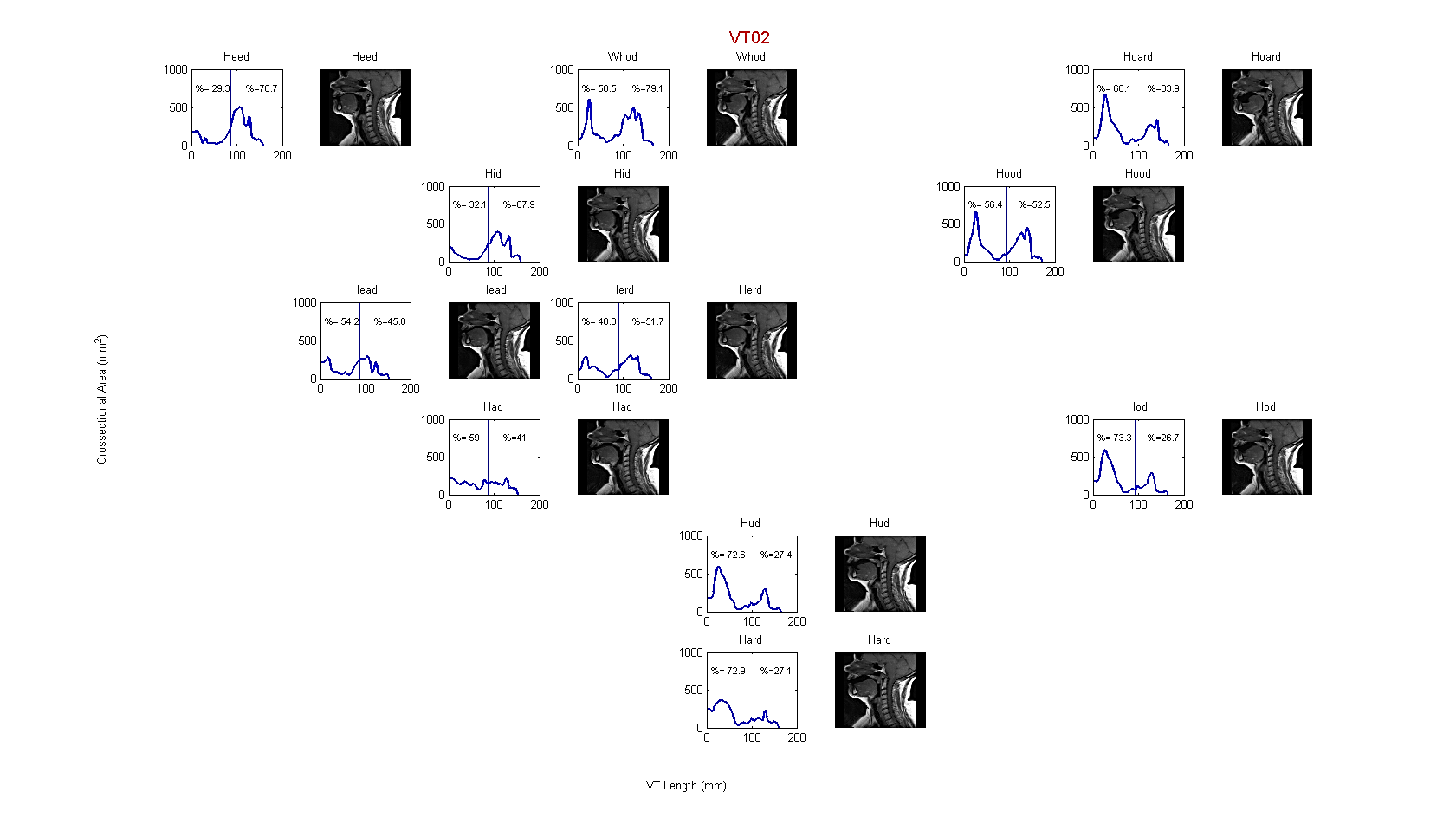


Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT02 producing NZE Vowels

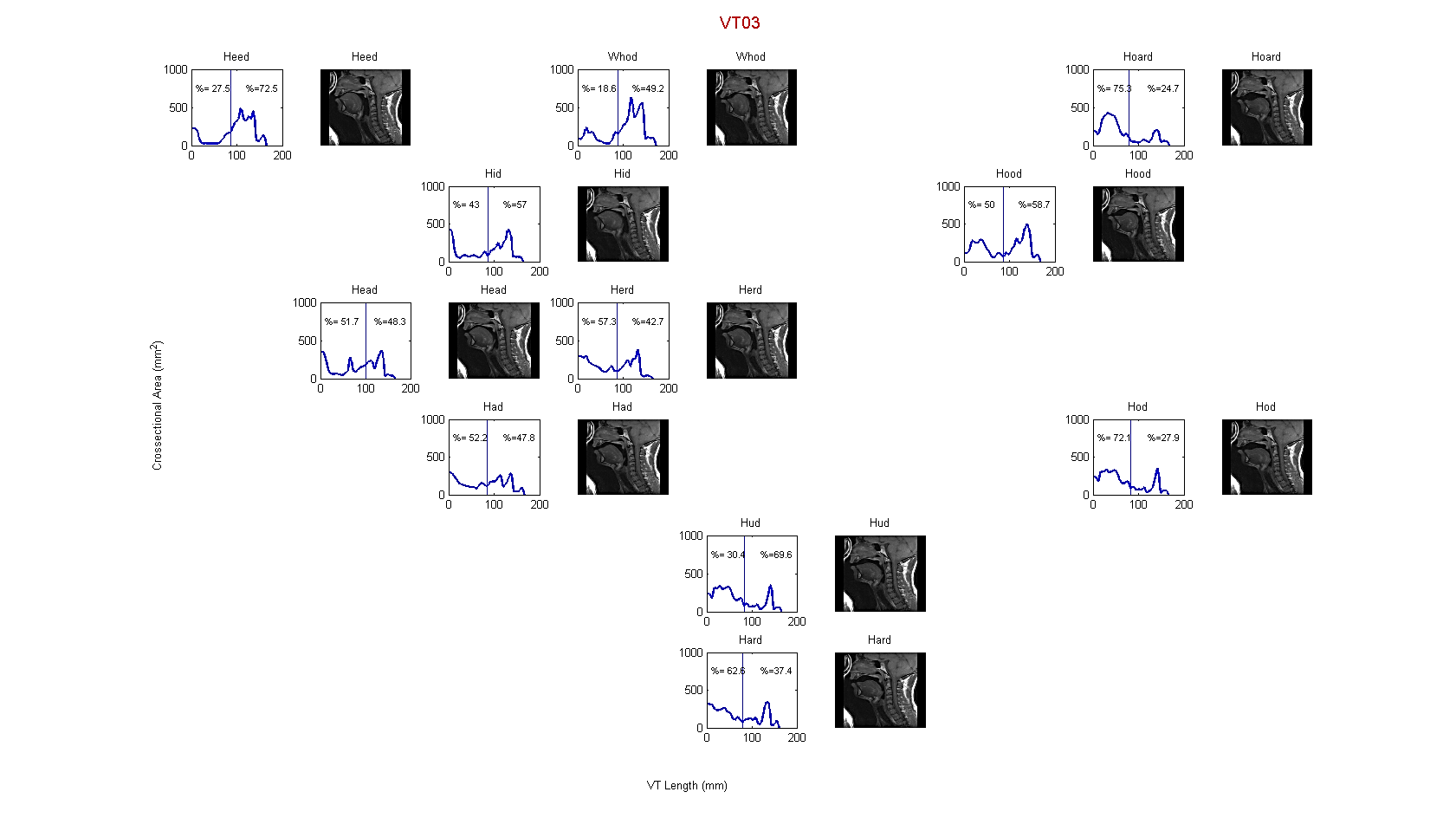


Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT03 producing NZE Vowels

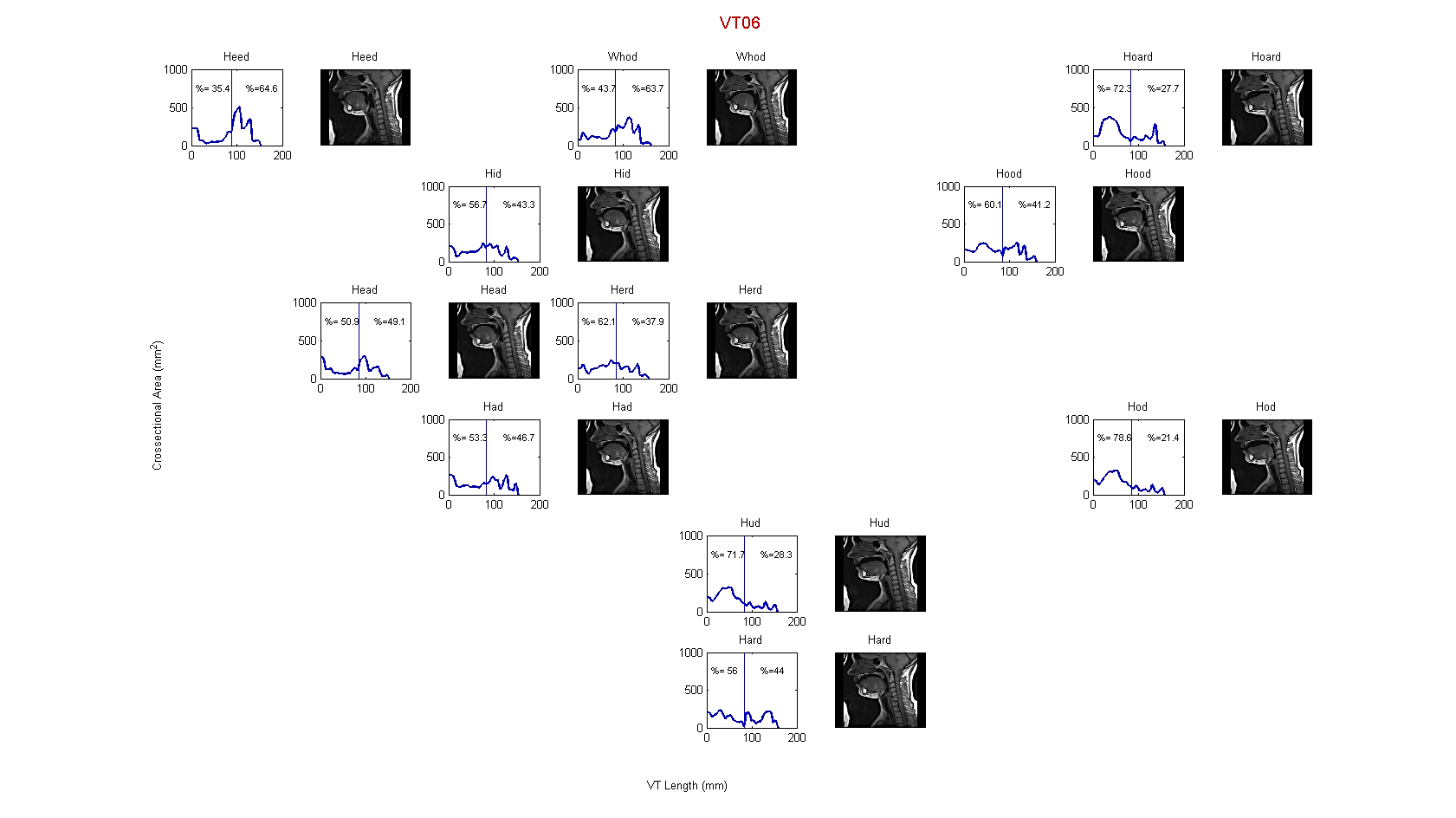


Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT06 producing NZE Vowels

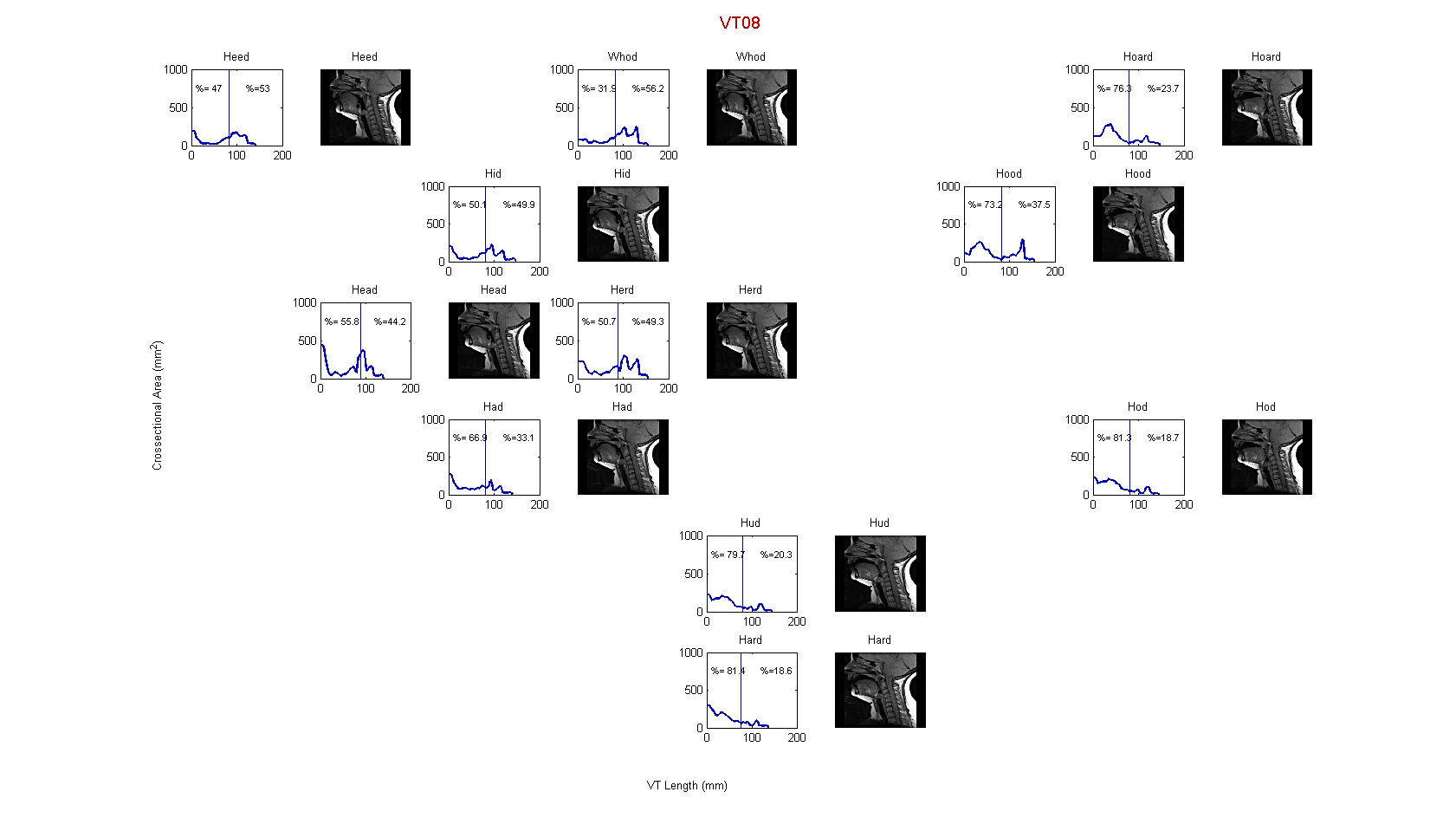


Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT08 producing NZE Vowels

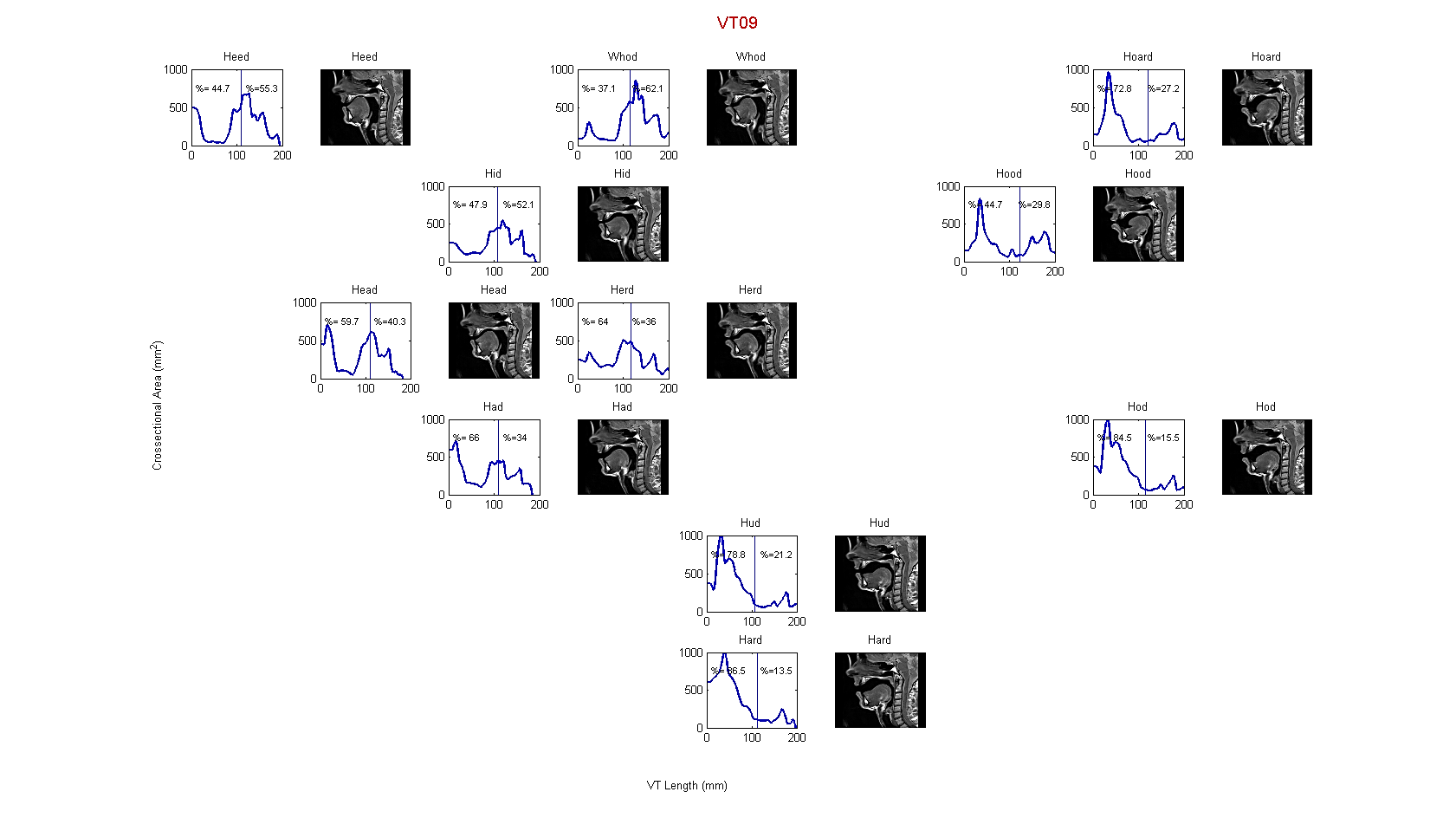


Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT09 producing NZE Vowels

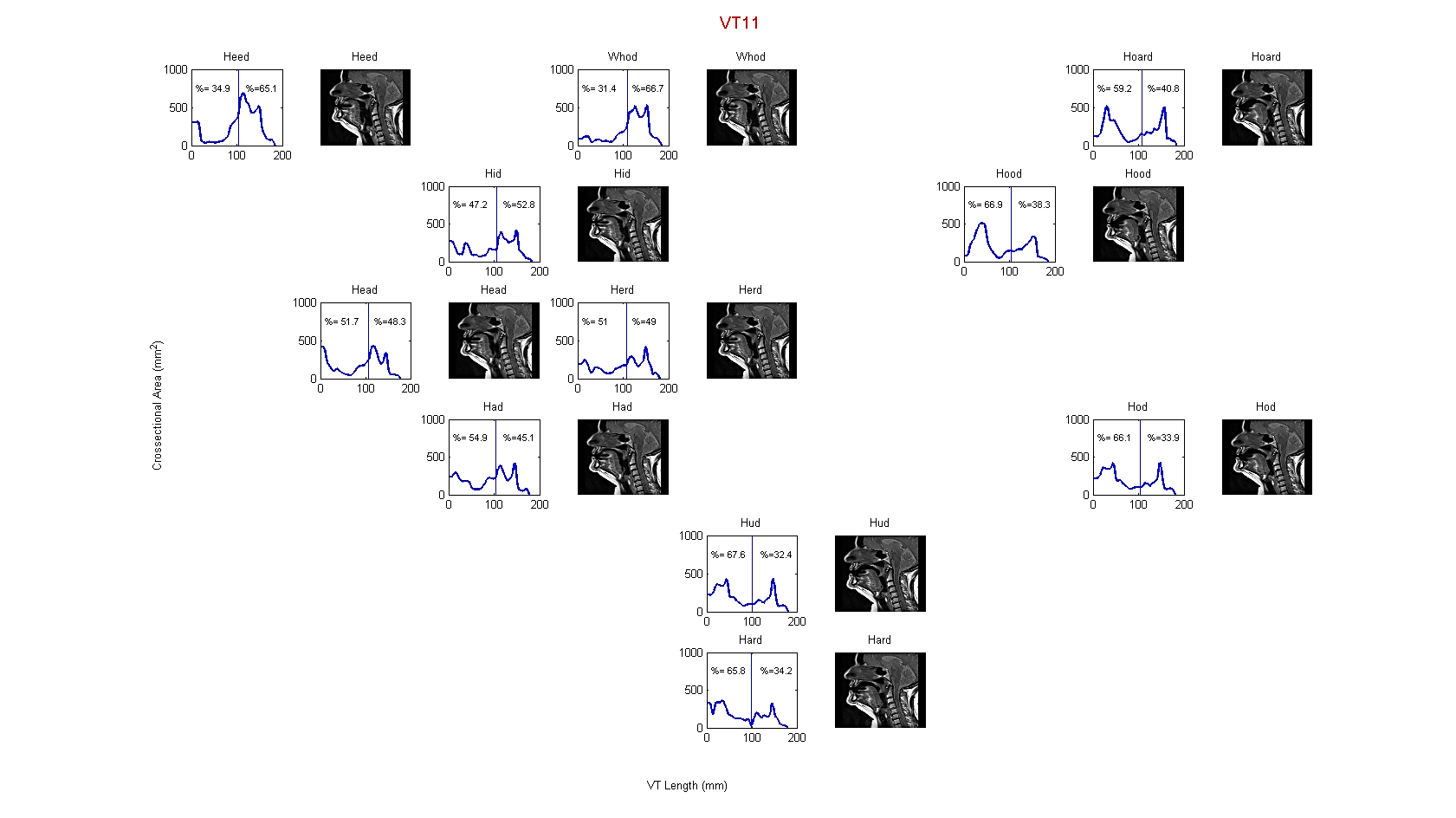


Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT11 producing NZE Vowels

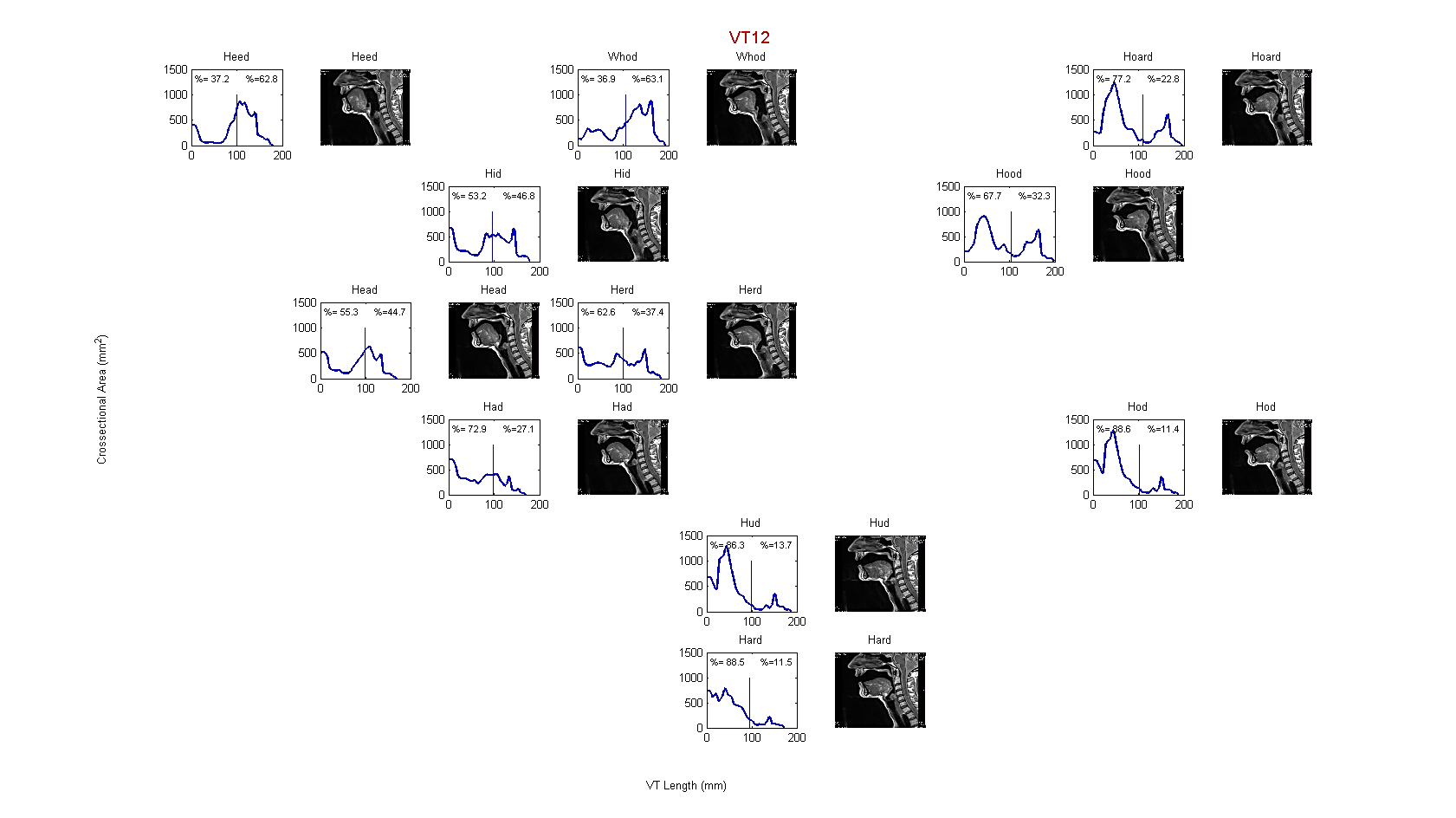


Figure : Distance vs. Cross-sectional Area Plots with Relative Oral and Pharyngeal Volumes for VT12 producing NZE Vowels

#### Understanding the Area-Distance Plot

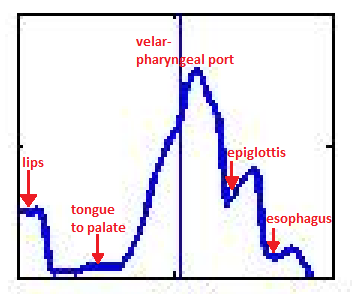


Figure : An example area-distance plot (heed)

Key differences to note between different vowels are shown in *table one*.

Table : Key features of distance-area plots

|  |  |  |
| --- | --- | --- |
|  |  | **Front vowel vs. back vowel**.  The front vowel should show large cross-sectional areas in the pharyngeal region, whereas the back vowel should show large cross-sectional areas in the oral region. |
|  |  | **Closed vs. open**  These vowels have the same level of “backness” therefore they should mainly appear different in oral region which is largely affected by jaw height. |
|  |  | **Rounded lip vs. unrounded lip**  The vowels which are considered rounded should have a much smaller initial ‘lip’ area than their respective unrounded vowel. |
|  |  | **Central vs. front/back**  The central vowel will have a much flatter area-distance plot. |

#### Oral/ Pharyngeal Volume Comparison

Figure : Average oral and pharyngeal volume ratios for all speakers producing NZE vowels

Table : Average volume ratio for all speakers

|  |  |  |
| --- | --- | --- |
|  | **Oral (%)** | **Pharyngeal (%)** |
| **heed** | 36.7 | 63.3 |
| **hid** | 46.3 | 53.7 |
| **head** | 54.7 | 45.3 |
| **had** | 61.9 | 38.1 |
| **hud** | 75.2 | 24.8 |
| **hard** | 74.4 | 25.6 |
| **hod** | 78.5 | 21.5 |
| **hood** | 59.8 | 40.2 |
| **hoard** | 72.9 | 27.1 |
| **who'd** | 35.8 | 64.2 |
| **herd** | 58.3 | 41.7 |
| **AVERAGE** | **59.5** | **40.5** |

Table : Comparative average volume ratios for C, M and Y speakers producing NZE Vowels

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Oral (%)** | | | **Pharyngeal (%)** | | |
|  | **Y** | **M** | **C** | **Y** | **M** | **C** |
| **heed** | 31.4 | 40.9 | 33.9 | 68.6 | 59.1 | 66.1 |
| **hid** | 49.8 | 50.5 | 39.9 | 50.2 | 49.5 | 60.1 |
| **head** | 51.3 | 57.5 | 54.8 | 48.7 | 42.5 | 45.2 |
| **had** | 52.7 | 69.5 | 61.2 | 47.3 | 30.5 | 38.8 |
| **hud** | 68.8 | 82.5 | 73.0 | 31.2 | 17.5 | 27.0 |
| **hard** | 59.3 | 87.5 | 73.4 | 40.7 | 12.5 | 26.6 |
| **hod** | 75.3 | 86.5 | 74.2 | 24.7 | 13.5 | 25.8 |
| **hood** | 52.7 | 63.8 | 59.7 | 47.3 | 36.2 | 40.3 |
| **hoard** | 73.8 | 75.0 | 69.8 | 26.2 | 25.0 | 30.2 |
| **who'd** | 34.1 | 37.1 | 35.9 | 65.9 | 62.9 | 64.1 |
| **herd** | 59.7 | 63.3 | 56.5 | 40.3 | 36.7 | 43.5 |
| **AVERAGE** | **55.4** | **64.9** | **57.5** | **44.6** | **35.1** | **42.5** |

These results do not appear to show the expected “aging” trend, that the relative pharyngeal volume increases with age. It should be noted that the “average” data shown in *table 3* aligns closely with the expected relative volumes based on the placement of the articulators. The volume ratios for each individual speaker are also provided in *figures 13 to 23*.

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "heed" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "hid" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "head" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "had" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "hud" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "hard" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "hod" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "hood" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "hoard" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "who'd" vowel

Figure : Relative Pharyngeal and Oral cavity volume for C, M and Y speakers producing "herd" vowel

# Conclusion/Future Work

At the completion of this summer studentship, I have presented Dr. Catherine Watson with a clear, proven methodology for the measurement of the vocal tract. The method allows the vocal tract to be expressed in terms of the length, area and volume. This can be achieved for the oral and pharyngeal cavities as separate entities or the vocal tract in its entirety.

As this project was focused on the method development, only preliminary analysis of the results has taken place. It is suggested that before this analysis to be taken further, a greater number of vocal tracts should be measured. Repeat measurements of the same speakers are available and should also be processed.

Some regions which have been identified as being of interest include:

1. Analysis into the comparative lengths of the oral and pharyngeal cavities for each vowel sound provided in ‘lengths1’ folders for each individual e.g. R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Vocal Tracts from Varadha\VT01\lengths1 and in the percentage-length plots in *Appendix H.* This is considered interesting as the length will have a significant effect on the volume of these cavities (since volume area length)
2. Analysis of the ‘trends’ shown between groups of speakers in *figures 12-22.* For example, do the younger speakers have smaller relative oral cavities for the front vowels than the older speakers? Do they have smaller relative pharyngeal cavities for the pharyngeal cavities? Are these results at all related to the training level of the speakers involved? Separate analysis of each speaker may also be of interest.
3. Application of the specific knowledge of the location of the velar-pharyngeal port to AR data to allow for better comparison between the two methods.
4. Further investigation into altering the R2.8.0 code (R:\ECE\Signal Processing\Catherine Watson\MRI studies\AR\09-02-12 Version I.r) to look at how MRI data can be imported and formants analysed.

# Appendix A: Literature Review of MRI Studies

Note: short summaries of other related (mostly AR) experiments in work book

**Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels (Baer, 1991)**

Objective: Use MRI data to obtain vocal tract area functions

Experiment One Set Up:

Subjects: two males

Vowels tested: /ɑ i æ u/ canonical target vowels (pre-recorded tape) were played to subject through an ear piece in an attempt to maintain the vowel quality over a long period of time.

Machine used: experimental MR (EMR) machine. This permitted the introduction of an apparatus for head stabilisation, sound recording and reproduction.

Subject voice recording: Simultaneous to the MRI

Subject positioning: Supine with head inside a saddle-shaped radio-frequency coil for receiving the resonance signal.. Cephalostat (consisting of a custom made head mold attached to a rigid base with a hinged plastic locating wand which was designed to make contact with the subject’s nose to give a tactile sensation as a reminder of the need to stay as still as possible) to ensure that head could be returned to same position for each test.

Time required: 3 to 4 hours for one vocal tract configuration. Producing each selected vowel in a sustained monotone for about 15s between brief inspirations, continuing to do so for the 3.4 minutes required for image acquisition. Up to 19 axial and 18 coronal images were taken.

Data collected: Vocal tract configurations of the two subjects from larynx to lips while they produced the four point vowels. The segment of the vocal tract examined was from the superior margin of the vertebral axis to the inferior margin of the sixth vertebra

Experiment Two Set Up:

Subjects: Two males

Vowels tested: four point vowels /ɑ i æ u/ and five other intermediate vowels /ɪ ɛ ɔ ʊ ʌ/

Machine used: SIGNA machine (1.5 Tesla). Demand for machine high so not possible to use external equipment such as cephalostat and sound recording. The majority of images were axial images. Much higher flux density, higher speed, less need for head stabilisation.

Subject position: Supine, with head in a padded universal cranial molding with a built in rf coil serving both transmitting and receiving functions. Laser reference beams were used to restore the head to a predetermined position.

Time required: Producing each selected vowel in a sustained monotone for about 15s between brief inspirations, continuing to do so for the 3.4 minutes required for image acquisition.

Data collected: Complete set of coronal and midsagittal images in addition to axial images obtained for /i/ vowel. Only axial images for all other vowels and only axial images were analysed.

Measurement of Cross-sectional Areas

1. Boundary Tracing- contour defined at 50% level of the density profile and traced by hand under X3 magnification to reduce tracing errors, using a trackball pen. For the SIGNA data, contours were traced by a computer algorithm which similarly traced the 50% density level. Near the air-teeth boundary, the density contour was unable to be detected and was therefore traced by hand.
2. Calibration: a wedge made from Plexiglas and filled with mineral oil and placed in the machine. The known cross-sectional area was compared to the measured cross-sectional area to give calibration factors for each of the three planes. This was only required for the EMR machine.

Determination of Area Functions

Sets of axial and coronal coordinates were arranged in a 3D matrix. The gaps between were filled using a process of duplication to thicken each ‘slice’. Areas were measured from a series of planes spaced about 0.5 cm along the approximate midline of the tract. The areas of the airway were calculated using Simpsons Rule. In order to remove error due to teeth, molds were made of each subjects palate and dentition. Images were hand-traced in digital form and then super-imposed on digital images.

Formant Analysis

The average formant frequencies were obtained using the 14-pole LPC analysis of three 5s long natural vowel samples drawn randomly from points close to the start, middle and end of many image acquisitions. Also, digital filters derived from area functions were used to resynthesize the vocal sounds to compare synthesised formants with the actual formants.

Perceptual analysis of the actual and synthesised vowels was also undertaken.

Key data: Area functions describing the individual tract shapes, formant comparison of actual and synthesised vowels, plots of cross-sectional area versus the midsagittal width of the tract at different locations in the pharynx for different vowel productions. Experiment one also allowed data relating midsagittal width to cross-sectional area in the oral cavity to be examined.

Other notes of interest: one must remember that measurements are derived from static vocal tract shapes and we can only guess how closely that matches dynamic production. It is suggested that subjects who have received formal phonetic training are used to ensure more consistent vowel quality. LPC assumes no zeros i.e. no nasalisation.

**An MRI study on the relationship between oral cavity shape and larynx position (K. Honda et al)**

Subjects: 12 Japanese (two female, ten male) and 12 English speakers(three female, nine male). Ages ranged from 23 to 48 years.

Vowels tested: resting (non-phonating) position

Machine used: Shimadzu clinical MRI scanner (1.0 Tesla)

Subject audio recording: N/A

Subject positioning: Supine position (noted that results may differ from larynx in natural speaking position)

Time required: not given

Data collected: Orofacial morphological factors that correlate with individual variation in larynx position in a non-phonating rest position. The objective of this experiment was to establish a method for predicting larynx position from the orofacial geometry based on midsagittal craniofacial MRI data collected from the same language groups. In addition to this the authors expected to identify certain regularities of morphological variation in the orofacial structure as anthropological evidence that bridges the evolution of speech organs and the origin of human speech.

Other notable techniques:

* Landmarks and measure were identified independently by the two authors and the results averaged
* This report uses some indicators of height values which could be used in our study to identify the cut off between the oral and pharyngeal cavity

**Morphology and development of the human vocal tract: A study using magnetic resonance imaging (W. Fitch, 1999)**

Objective: Changes in vocal tract anatomy with age, in particular on sex differences and on the relation between vocal tract length and body size. Note that it has been assumed since 1952 that vocal tract length is correlated with body size although it has never been empirically examined.

Subjects: 129 ‘normal’ children and adults, ages 2 to 25 years. Exclusion criteria for abnormal were congenital abnormalities, or history of speech delay or language impairment, in the subject or first degree relatives of the subject. Pubertal status was quantified using administered questionnaire yielding Tanner ratings of pubertal stage.

Vowels tested: N/A (asked to lie motionless and breathe quietly while being scanned) It should be noted that this mostly resulted in scans showing subjects in a nasal breathing posture.

Machine used: General Electric 1.5 Tesla scanner

Subject audio recording: N/A

Subject positioning: Supine

Time required: not provided (not important for this as no sustained vowel production)

Data collected: Height and weight of subjects. Single midsaggital slice MRI image. ‘Line segment’ technique developed from traditional methods(recorded to spreadsheet and saved for further processing) and the ‘curved line’ technique were both used to find vocal tract length. Shape and proportions of VT were found using computer graphic techniques.

Other notable techniques or information gained:

* Measurement of the vocal tract: Uses the ‘curved line’ technique, measuring from the glottis traveling up through the pharyngeal midline and between the tongue and palate and terminating at a plane touching the upper and lower external borders of the lips. This path was chosen as it closely approximates the path taken by the longitudinal pressure waves generated at the larynx and emanating at the lips. Note the level of the glottis was indicated by one/more of the following: vocal fold visible beneath the ventricular fold, a supraglottal notch, the caudal delimitation or the border between the cricoid and arytenoid cartilages.
* Used ANOVA to check significance of results
* It has been suggested that lying in a supine position may lead to a raising of the hyolaryngeal complex, therefore the overall length measurement of the VT measurement may be compromise but comparisons between different ages/ sexes should be consistent

Key Results: Clear correlation between VTL and body size. There is a clear difference in male and female VT morphology, overall VTL and the relative proportions of the oral and pharyngeal cavity. This is not evident in children.

**Vocal tract area function for vowels using a three-dimensional MRI: A preliminary study (P. Clément et al, 2006)**

Objective: To assess whether MRI allows the vocal tract area function to be determined for a normal male speaker.

Subjects: One male (40 years) native French speaker with no history of speech or voice disorders and who was phonetically untrained.

Vowels tested: French point vowels only / i/, /a, /u/

Machine used: General Electric Horizon 1.5 Tesla machine.

Subject audio recording: An attempt to emulate as closely as possible the conditions experienced in the MRI sessions. (Supine position, soundproof room, earplugs for similar acoustic feedback, same head position). Speech sounds were recorded three times from sustained vowels which the subject attempted to produce as close as possible to those acquired during the MRI session.

Subject positioning: The subject was positioned in a supine position with a cushion and foam to support the head and neck but to allow phonation. The subject was given earplugs to attenuate the sound of the MRI machine.

Time required: The subject sustained phonation until the end of the acquisition of the image. The amount of time to image one shape was 32 seconds allowing for short inspirations while keeping the vocal tract stable.

Data collected: Mid-sagittal (defined when the corpus callosum could be observed) image for each vowel. Cross-sectional slices were computed from a series of planes spaced at intervals of 1cm along the length of the VT and used as an input to an acoustic VT model the vowel sounds based on the area functions were able to be simulated. The formants taken from the acoustic recordings of natural speech were found using Linear Predictive Coding. These values of the first three formants were compared with the computed formants from the MRI measurements and VT model.. A perceptual identification test was also performed to assess the perceptual significance of the vowels recovered from the measured VT area functions, with the computed and simulated vowels used as stimuli.

Other notable techniques:

* Cross-sectional area determination: All cross-sectional areas were traced five times by two authors. Cross-sectional areas were calculated to be locally perpendicular to the airway centreline. Each area function was generated as a set of x-y co-ordinates that included the length co-ordinate and the corresponding cross-sectional area. The area function was assumed to begin at the glottal end of the vocal tract and terminate at the lips.

Key result: MRI gives the correct picture of the mid-sagittal cross-section but does not necessarily create an accurate 3D reconstruction.

**Magnetic Resonance Imaging procedures to study the concurrent anatomic development of vocal tract structures: preliminary results (H. Vorperian, 1999)**

Objective: To demonstrate the feasibility of using MRI to examine the growth processes of the vocal tract. The purpose of the report is to describe the MRI processes implemented to obtain quantitative measurements on the concurrent anatomic development of a number of vocal tract structures from two paediatric subjects.

Subjects: Two male paediatric (5 times from age 0 – 2, 3 times from age 2-3) subjects who had received serial/repeat MRIs for medical conditions that are not thought to affect the overall physical development.

Vowels tested: N/A

Machine used: GE scanner or Resonex

Subject audio recording: N/A

Subject positioning: Subjects underwent sedation, facial structures were placed centrally in the head coil using the laser lights of the scanner

Time required: Not provided

Data collected: Both axial and sagittal slices. Measurements were made between key landmarks. These were; head length, face height, hard palate length, soft palate length, maxillary arch width, maxillary arch length, mandibular width, mandibular length and depth, vocal tract length, tongue length, tongue area, hyoid bone level/ tongue level, laryngeal level, naso-oro-pharyngeal length, maxillary lip thickness, mandibular lip thickness, maxillary lip length mandibular lip length, mandibular lip area, maxillary lip area.

Key Results: MRI is an excellent tool to obtain quantitative results of anatomic development of various vocal tract structures. Synchrony of growth in soft and hard tissues which persists during growth spurts.

**Comparison of Magnetic Resonance Imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002 (Story, 2007)**

Objective: The aim of this report was to compare the new set of area functions obtained with MRI with the area functions obtained 8 years earlier for the same speaker. Specifically, this is to report the area functions in numerical form, compare them graphically and acoustically and to provide some explanation of how the differences observed between the two are function sets support a downward shift in F2.

Subjects: 37 year old male American

Vowels tested: 11 American English vowels /i ɪ e ɛ æ ʌ ɑ ɔ o ʊ u/

Machine used: General Electric Sigma 1.5 Tesla Scanner

Subject audio recording: On following day, production of all 11 vowel sounds was recorded. Subject wore earplugs and lay supine on a cushioned table. At least three repetitions of each vowel were produced with similar production to the previous day, speaker attempted to produce similar vowels in terms of both quality and loudness.

Subject positioning: with a flexible anterior neck coil, supine with earplugs, limited feedback conditions.

Time required: produced as if were to be spoken in hVd but sustained for 8s, the subject was allowed to breathe and this was repeated for about 30 times, each image set required 4min and 16s. Each image set about 10-15min. Goal was for “normal production” of each vowel.

Data collected: For each vocal tract shape the procedure included; segmentation of the airspace from the surrounding tissue, shape-based interpolation to generate a 3D reconstruction of the airspace and a cross-sectional area analysis of the air space.

Key Result: Tendency towards a relatively constricted pharyngeal region and expanded oral cavity, a clear downwards shift of F2.

**Two Short Studies in Vocal Tract Measurement (C. Watson & J. Hui, 200??)**

Objective: This paper presents the findings of two short studies that were completed as a consequence of the larger on-going study on modelling aging in speech production. The first study used AR. The second study, which will be focused on in this summary presents the first collection of MRI mid-sagittal vocal tract images of the New Zealand English monophthongs.

Subjects: Eight speakers (seven are NZE speakers, two of which are presented in the report)

Vowels tested: NZE monophthongs /i ɪ ɐ ɛ æ ɑ ɔ ɜ u/ sustained vowels using an hVd format

Machine used: 1.5 T Siemens Magnetom Avanto MRI scanner.

Subject audio recording: Part of the AR study, recorded in a sound both. For each speaker, citation forms of hVd words for nine monophthongs were collected. Five versions were collected which were presented to the speaker in a random manner. / ʌ/ and / ʊ/ were not recorded . Only differ from other monophthongs in length. (noted that this was discussed as a poor assumption)

Subject positioning: supine ... ?

Time required: sustained pronunciation of vowels, scanning time approximately 15 seconds (reduced from 24 seconds in the pilot study).

Data collected: Sagittal scans of the head only in the jaw area, audio recording/formant data, AR area functions( first section-not discussed)

**Aging influences on pharyngeal anatomy and physiology: The predisposition to pharyngeal collapse (A. Malhotra et al, 2006)**

Objective: By combining MRI techniques with pharyngeal physiological assessment, the author sought to determine the structural and functional basis for the increased propensity for airway collapse among older persons.

Subjects: Eighteen mean and 20 women across a range of ages. Women under 50 were premenopausal based on regular menstrual cycles, whereas women over age 50 were postmenopausal for at least 2 years.

Vowels tested: N/A

Machine used: Signa Advantage, General Electric 1.5 Tesla MR Scanner

Subject audio recording: On following day, production of all 11 vowel sounds was recorded. Subject wore earplugs and lay supine on a cushioned table. At least three repetitions of each vowel were produced with similar production to the previous day, speaker attempted to produce similar vowels in terms of both quality and loudness.

Subject positioning: supine with head secured in the neutral anatomic Frankfort position.

Time required: not provided.

Data collected: Soft palate area and length, pharyngeal length (defined as hard palate to epiglottis), tongue height, width and area- from sagittal anatomy. From the axial airway image, the cross-sectional area was examined, lateral wall thickness, thickness of the pharyngeal fat pads, skeletal anterioposterior (mandible to vertebrae) and intramandibular distances. Edge detection algorithms using simple thresholding were used to avoid subjective bias in measurement.

# Appendix B: Preliminary Vocal Tract Length Approximation

The purpose of this investigation was to approximately measure the lengths of the oral and pharyngeal cavities provided from both AR and MRI data of different speakers. This was done as a preliminary investigation of the ‘aging’ hypothesis; that the pharyngeal cavity increases with age, proposed by Dr. Catherine Watson.

1. **Method:**

**AR Data:**

The location of the velar-pharyngeal port from the AR data was identified using the rules presented by Justine. “The pharyngeal region is the region between the velar-pharyngeal port and the glottis. This point was located in the data by automatically searching for the end of the first ‘valley’ from the lips. If the vocal tract shape did not have this valley, the point where the slope is the flattest after the first “hill” is taken as the velar-pharyngeal port.”

Lengths were measured using the “average” data reading.

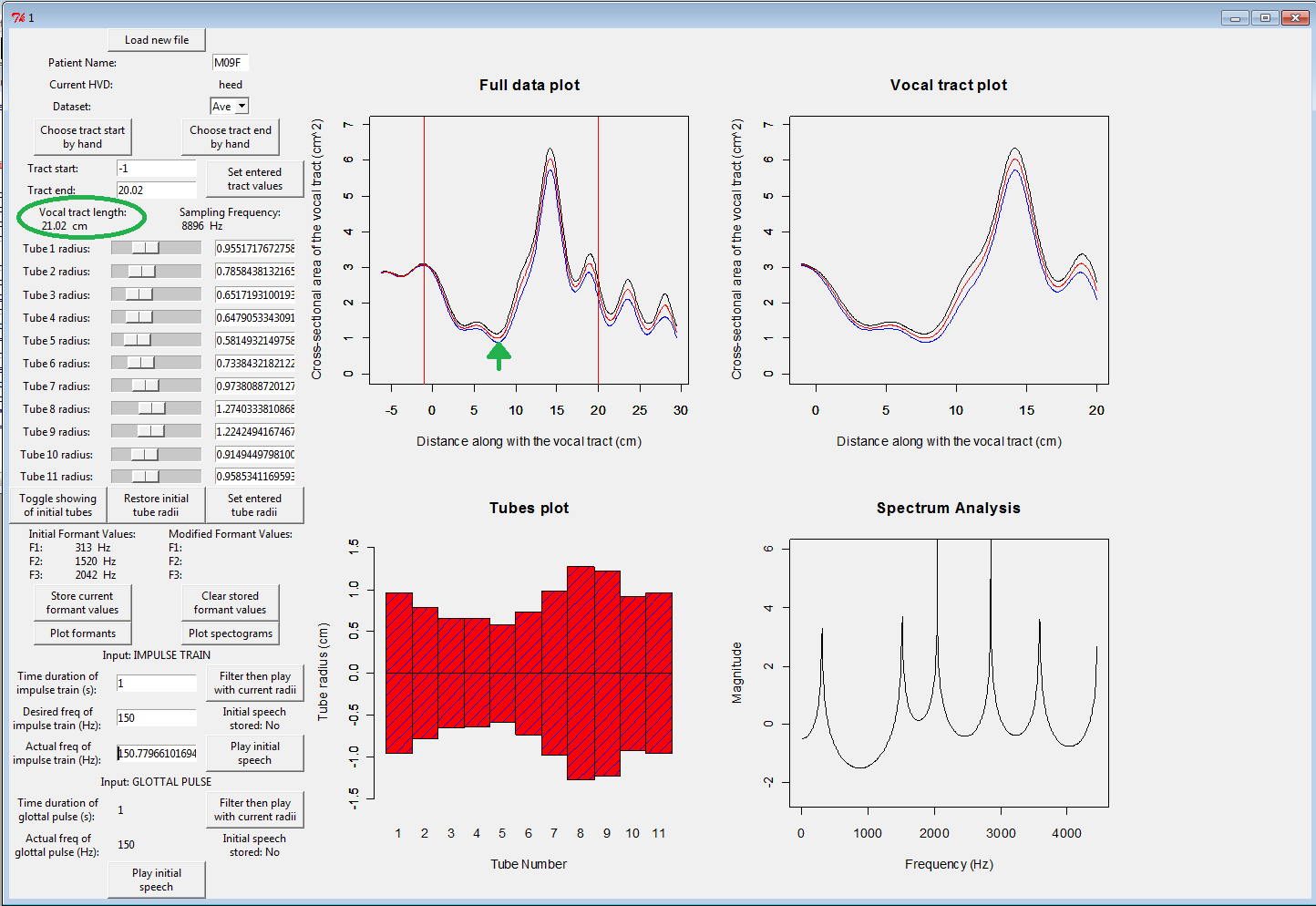


Figure : Example window from R2.8.0 showing where oral length and vocal tract length were taken for AR data

**MRI data:**   
Measurements were taking using the basic distance tool (the yellow ruler in *figure 25*) in the syngo FastView program, provided as an executable file with the MRI data. The start of the vocal tract was identified as the outmost point of the lips and the glottis was identified as the lowest “fuzzy” region of the MRI (approximately in line with the fourth vertebrae).

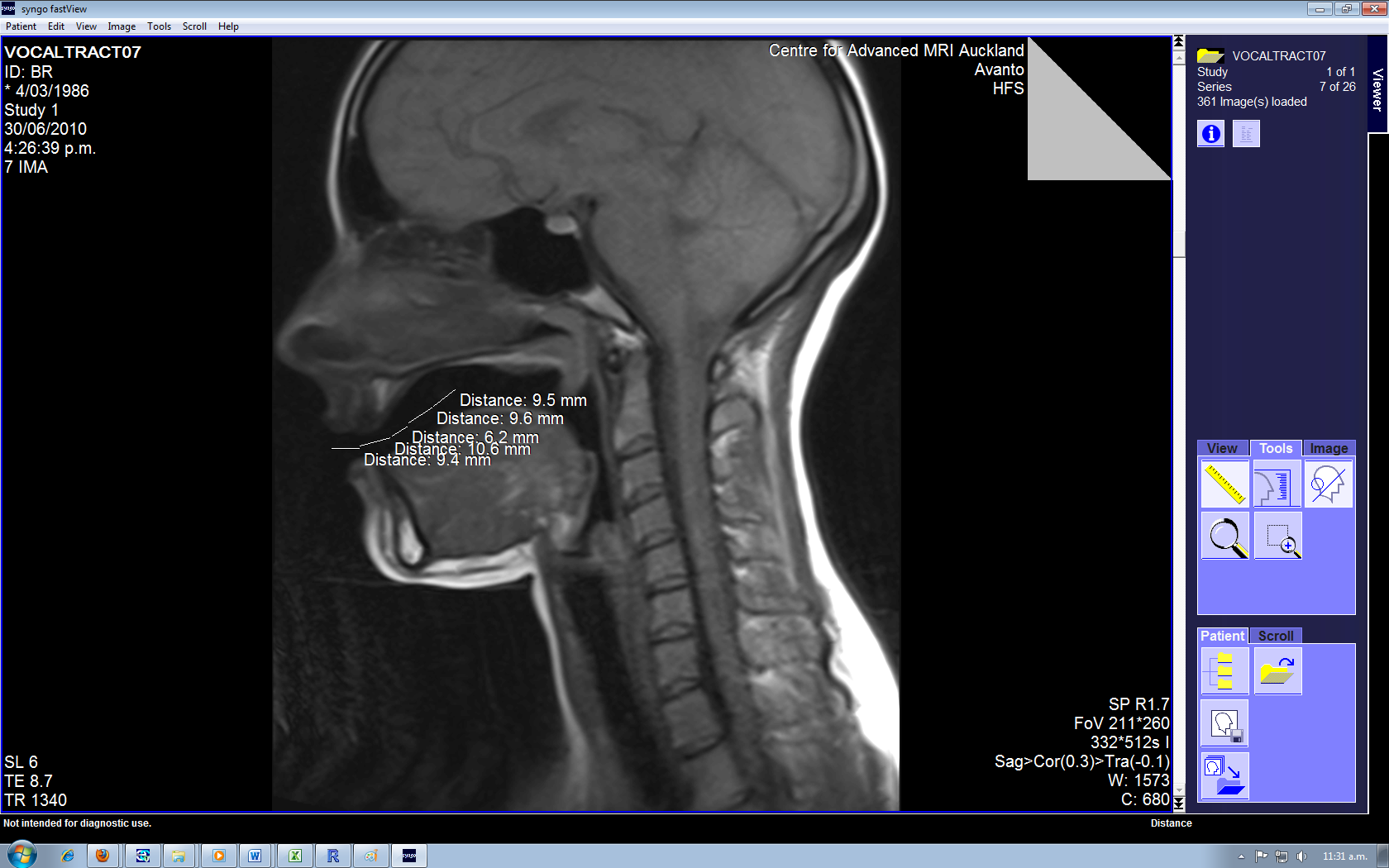


Figure : Example Window from syngoFastView showing distance measuring

1. **Results**

**Table 4: Comparison between AR and MRI measurements of oral (O), pharyngeal (P) and total (T) vocal tract length for a 45 year old male NZE speaker**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **HEED** | | | **HARD** | | | **HOARD** | | | **WHO’D** | | | **HERD** | | |
|  | **T** | **O** | **P** | **T** | **O** | **P** | **T** | **O** | **P** | **T** | **O** | **P** | **T** | **O** | **P** |
| MRI | 16.99 | 10.36 | 6.63 | 19.12 | 8.94 | 10.18 | 19.97 | 11.04 | 8.93 | 18.54 | 10.34 | 8.2 | 18.04 | 10.21 | 7.83 |
| AR | 19.74 | 5.7 | 14.04 | 19.74 | 6.2 | 13.54 | 19.31 | 6.8 | 12.51 | 21.02 | 6.4 | 14.62 | 19.31 | 5.15 | 14.16 |
| Comparison | 2.75 | -4.66 | 7.41 | 0.62 | -2.74 | 3.36 | -0.66 | -4.24 | 3.58 | 2.48 | -3.94 | 6.42 | 1.27 | -5.06 | 6.33 |

Table : Comparison between oral to pharyngeal length ratio derived from MRI and AR measurements for a 45 year old male NZE speaker

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **HEED** | **HARD** | **HOARD** | **WHO’D** | **HERD** |
| MRI | 1.56 | 0.88 | 1.24 | 1.26 | 1.30 |
| AR | 0.41 | 0.46 | 0.54 | 0.44 | 0.36 |

Table : Comparison between AR and MRI measurements of oral (O), pharyngeal (P) and total (T) vocal tract length for a 45 year old female NZE speaker

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **HEED** | | | **HARD** | | | **HOARD** | | | **WHO’D** | | | **HERD** | | |
|  | **T** | **O** | **P** | **T** | **O** | **P** | **T** | **O** | **P** | **T** | **O** | **P** | **T** | **O** | **P** |
| MRI | 14.59 | 8.73 | 5.86 | 14.05 | 7.92 | 6.13 | 15.65 | 8.14 | 7.51 | 15.62 | 8.57 | 7.05 | 15.54 | 8.53 | 7.01 |
| AR | 20.09 | 5 | 15.09 | 17.59 | 12.75 | 4.84 | 20.02 | 10.25 | 9.77 | 21.02 | 7.8 | 13.22 | 18.02 | 4.8 | 13.22 |
| Comparison | -5.5 | 3.73 | -9.23 | -3.54 | -4.83 | 1.29 | -4.37 | -2.11 | -2.26 | -5.4 | 0.77 | -6.17 | -2.48 | 3.73 | -6.21 |

Table : Comparison between oral to pharyngeal length ratio derived from MRI and AR measurements for a 45 year old female NZE speaker

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **HEED** | **HARD** | **HOARD** | **WHO’D** | **HERD** |
| MRI | 1.49 | 1.29 | 1.08 | 1.22 | 1.22 |
| AR | 0.33 | 2.63 | 1.05 | 0.59 | 0.36 |

Figure : Relative Oral- Pharyngeal Lengths based on MRI measurements for a 25 year old male NZE speaker producing NZE long vowels

Figure : Relative Oral -Pharyngeal Lengths based on MRI measurements for a 45 year old male NZE speaker producing NZE long vowels

Figure : Oral-Pharyngeal Lengths based on MRI measurements for a 25 year old male NZE speaker producing NZE long vowels

Figure : Oral-Pharyngeal Lengths based on MRI measurements for a 45 year old male NZE speaker producing NZE long vowels

Figure : Relative Oral Pharyngeal Lengths based on MRI measurements for a 45 year old female NZE speaker producing NZE long vowels

Figure : Relative Oral Pharyngeal Lengths based on MRI measurements for a 25 year old female NZE speaker producing NZE long vowels

Figure : Oral Pharyngeal Lengths based on MRI measurements for a 45 year old female NZE speaker producing NZE long vowels

Figure : Oral Pharyngeal Lengths based on MRI measurements for a 25 year old female NZE speaker producing NZE long vowels

1. **Discussion of Results**

*Figures 26-27, 30-31* show the relative length of the pharyngeal and oral cavities. Results from these figures do not support the hypothesis that aging results in an increase pharyngeal length relative to the length of the oral cavity. This does not however, rule out the hypothesis that the pharyngeal volume increases with aging. This requires integration of the cross-sectional area functions.

In *figures 28-29, 32-33* the absolute lengths of the pharyngeal and oral cavities; and hence, the overall length of the vocal tract are presented. The ‘front’ vowels /i/ and /a/ show slightly shorter overall lengths for the 45 year old female than for the 25 year old. However, for both vowels the oral cavity appears to be longer. This shows a much shorter pharyngeal region for the older speaker. The back vowels and schwa vowel show little difference between the two females. The back vowels for the 45 year old male show shorter overall lengths and oral cavities than the 25 year old male. For the hoard vowel, the pharyngeal length is relatively constant.

*Tables 4-7* show the comparison between vocal tract lengths using the AR data and the MRI. The large discrepancies in these values suggest that a measurement system with much greater accuracy needs to be developed.

# Appendix C: Formant Analysis

1. **Aim:**

To investigate the aging hypothesis and its effect on the formants produced.

1. **Method:**

The formants for each of the NZ long vowels were recorded based on the LPC analysis using R2.8.0 for four speakers; a male and female from each of the “young” and “middle age” categories.

The cross-sectional areas for the young speaker, provided from the AR data tubes, were then increased by 1mm2 at each slice in the pharyngeal region. The formants of the young speakers were recorded with this increased pharyngeal cavity and compared to the formants of the middle age speaker.

1. **Results:**  
   Data tables and charts can be found in R:\ECE\Signal Processing\Catherine Watson\MRI studies\Helen Summer Work\Summer\Preliminary Work\Formant Analysis.xlsx
2. **Conclusion:**

Results were not as expected, with each vowel providing similar formants and vowel space thus making it difficult to determine whether and increased pharyngeal region was giving the young person more “middle” age vowels.   
  
It was thought that the formant calculator set up in R may not have functioned properly and this investigation was discarded from further work.

# **Appendix D: Process for identifying oral and pharyngeal cavity markers.**

1. **Background information**

Tortora’s **Principles of Anatomy and Physiology** states that “the larynx lies anterior to the esophagus in the midline of the neck and anterior to the fourth through sixth cervical vertebrae.” As the cervical vertebrae are immobile, while supine, and the larynx can move up and down; we need a better anatomical reference point for identifying the glottis.

In **Morphology and development of the human vocal tract: A study using magnetic resonance imaging (W. Fitch, 1999)** the level of the glottis was indicated by one/more of the following: vocal fold visible beneath the ventricular fold, a supraglottal notch marking the vestibule and anterior commissure, the caudal delimitation of the pre-epiglottic fat, or the border between the cricoid and arytenoid cartilages.

Houri K. Vorperian defined the vocal tract length as: the curvi-linear distance along the midline of the tract, starting at the superior edge of the thyroid cartilage to the intersection with a line drawn tangentially at the lips.

We need to use this information to clearly define the borders of the vocal tract.

1. **Process- Creating a snake for the oral cavity**
2. Place 15 nodes approximately between the lips and velar-pharyngeal port. These are the ‘snake’ nodes as shown in *figure 34.*



Figure : Plotting the 15 nodes

1. Align the point at the lips by placing a new data point on the outermost point of both the maxillary and mandibular lip and create a straight line from data points as in *figure 35*. The first of the 15 ‘snake’ nodes should be in the centre of this straight line.

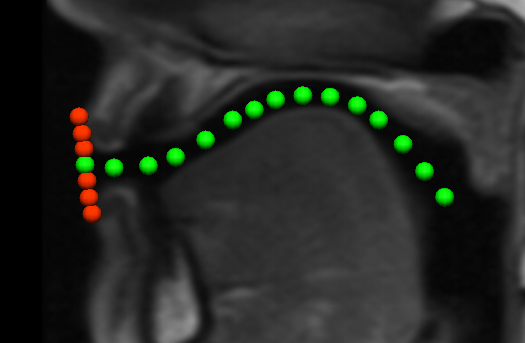


Figure : Aligning "Node 1" at the lips

1. Select all unnecessary points (shown in red in *figure 35*) and select “destroy selected nodes.”
2. Re-align the central 13 nodes, making sure that they are in the centre of the vocal tract.
3. Align the point at the velar-pharyngeal port by creating a line of new data points from the tip of the uvula to tongue which is perpendicular to the ‘snake’ curve. The 15th node of the ‘snake’ should be the central point of this line. If it is difficult to determine a perpendicular angle, add an extra node amongst the ‘snake’ nodes for clarity (as shown in *figure 36* )

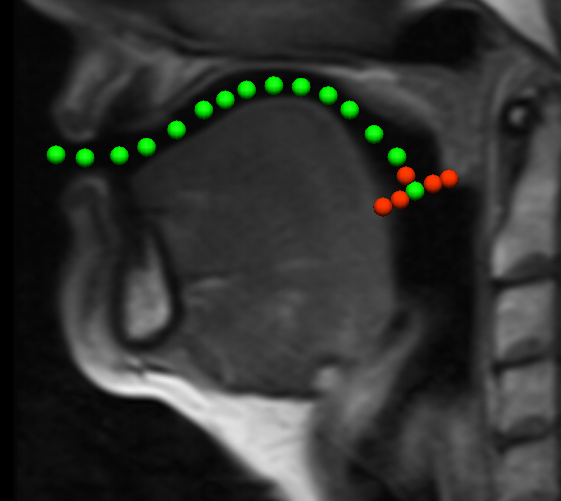


Figure : Aligning "Node 15" at the Velar-Pharyngeal Port

1. Select all unnecessary points (shown in red in *figure 36*) and select “destroy selected nodes.”
2. **Process- creating a snake for the pharyngeal cavity**
3. Place 15 nodes approximately between velar-pharyngeal port the glottis as in *figure 37*. These are the ‘snake’ nodes.

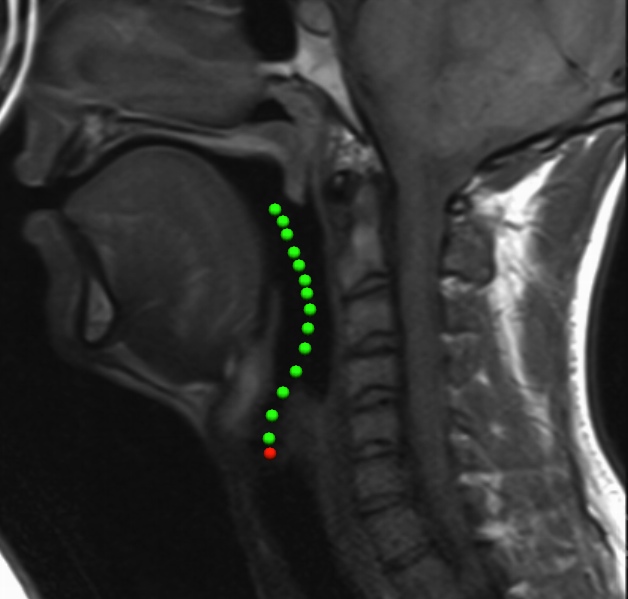


Figure : Placing 15 nodes approximately

1. As for the oral cavity, align the point at the velar-pharyngeal port by creating a line of new data points from the tip of the uvula to tongue which is perpendicular to the ‘snake’ curve. The 1st node of the ‘snake’ should be the central point of this line.

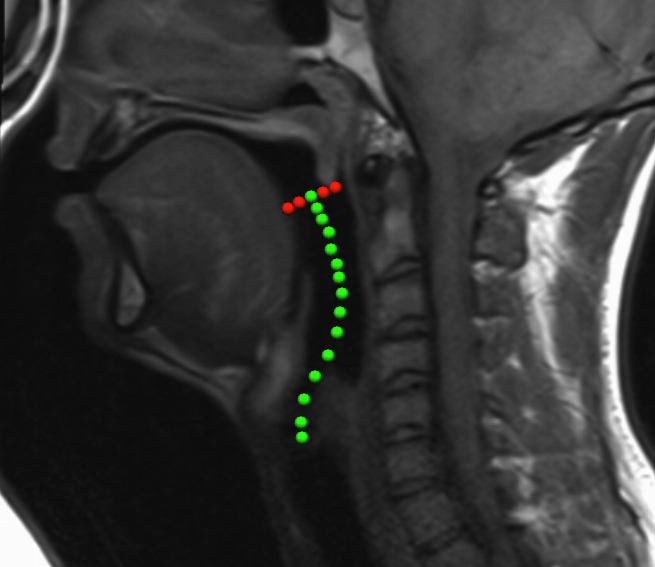


Figure : Aligning "Node 1" at the Velar-Pharyngeal Port

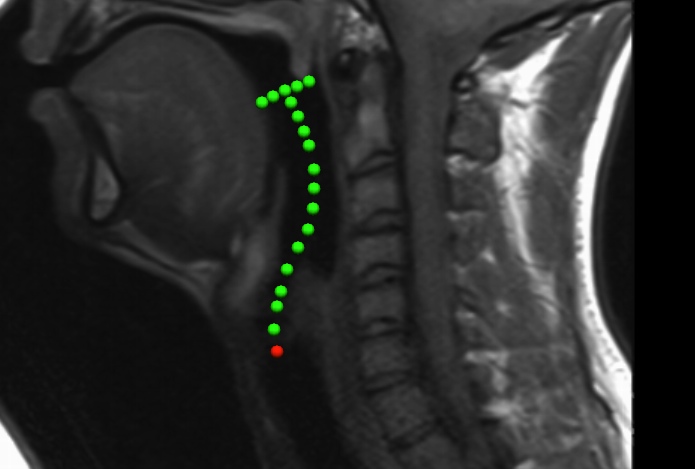
1. Select all unnecessary points (shown in red in *figure 38*) and select “destroy selected nodes.”
2. Re-align the central 13 nodes, making sure that they are in the centre of the vocal tract.
3. While phonating, the vocal folds are constantly moving, creating blurring in the MRI image. This means that glottis is much more difficult to define than the other anatomical landmarks of the vocal tract.. In *figures 39 - 41* are vocal tracts from different people, with the glottis identified. It is important to look at as many different vocal tracts as you can so that you are clear about where the vocal tract finishes

Figure : Identifying the glottis

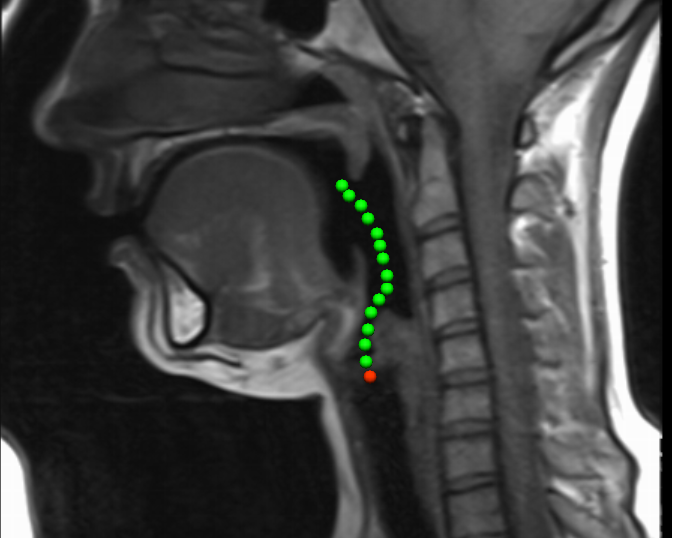


Figure : Identifying the glottis (example 2)

Some images will be much clearer than others so it may help to use a clear image as a reference tool for the more ‘fuzzy’ images. The vocal tract shown in *figure 41* shows a very easily identifiable glottis.

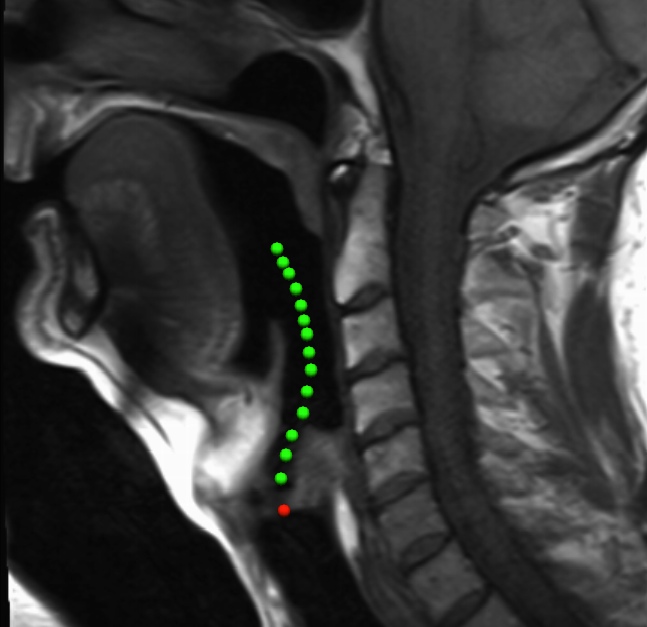


Figure : Identifying the glottis (clear example to use for reference)

The key anatomical features are indicated on this vocal tract in *figure 42* and in the diagram, *figure 43.*

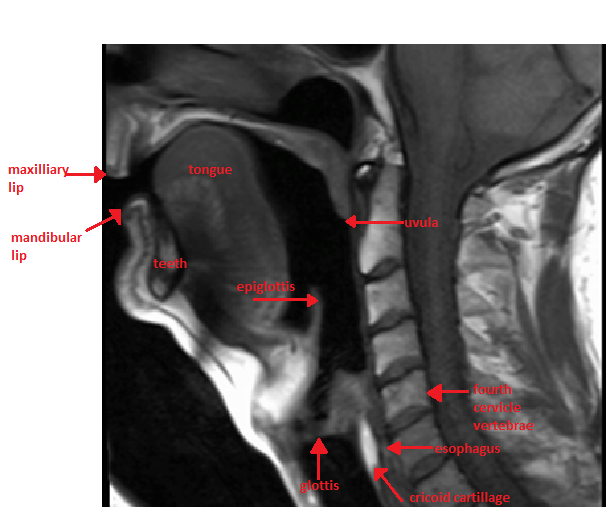


Figure : Anatomical Markers on MRI

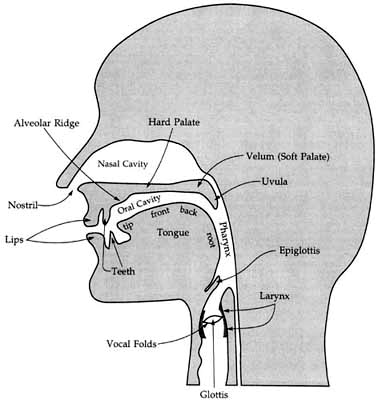


Figure : Key Anatomical Markers

# Appendix E: Node Placement Additional Information and Trouble-shooting.

The instructions provided in the template folder “6vocaltractnodeplacement” give provide details on how to run the file but require some extra tips and suggestions as to how to navigate through CMGui and place the nodes accurately. Provided below is a basic tutorial highlighting some of the commonly encountered problems.

1. Copy in the slices folder and curve.exnode
2. Run the data\_point\_placement file

**Troubleshooting**: Why are the images not printing onto the slices? If the Graphics window appears like *figure 44*, the image input code does not match the images in that folder.

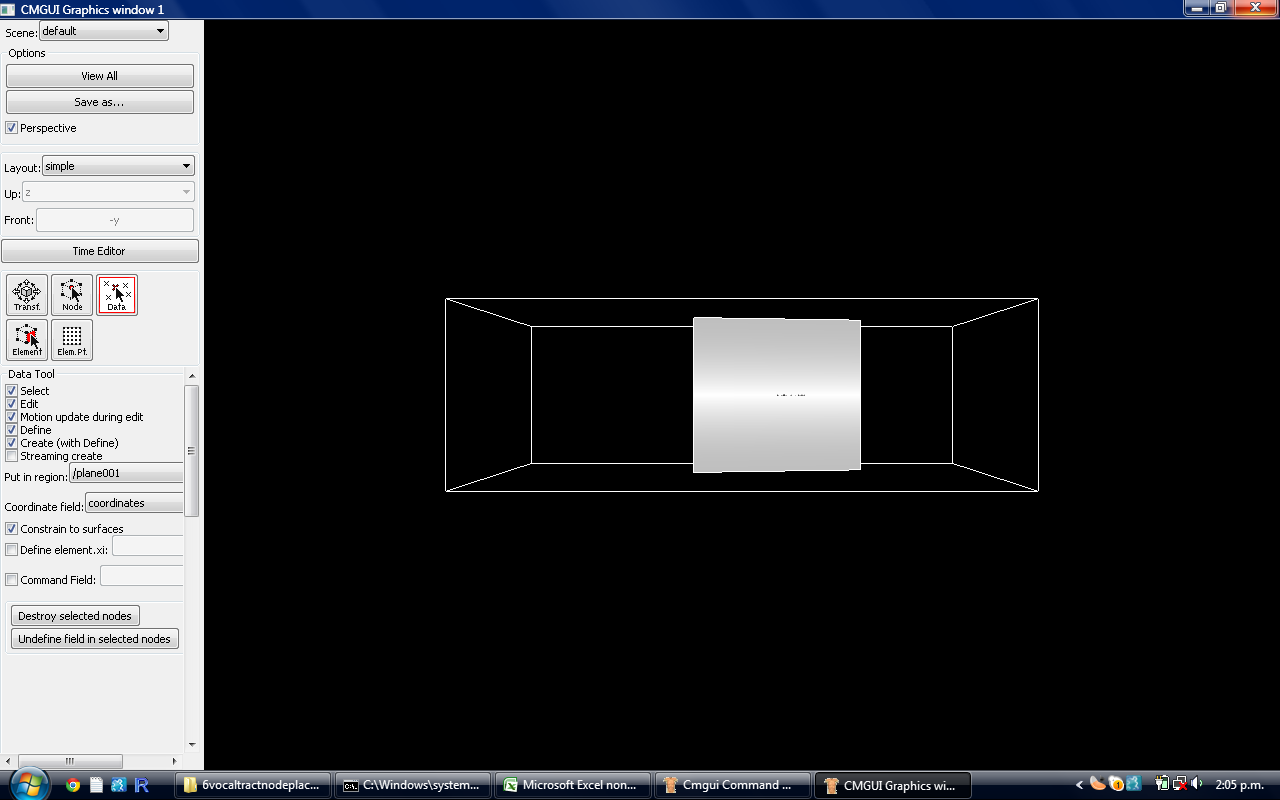


Figure : Code error

1. Edit the data\_point\_placement file in Notepad ++to fit the input images (highlighted text in *figure 45*)

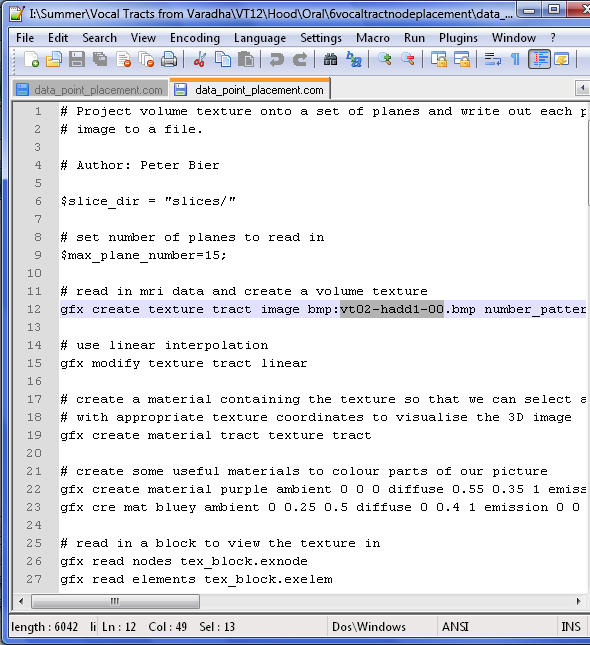


Figure : Editing the code

1. When images print correctly, set screen up as in *figure 46*, with a large graphics window and small scene editor.

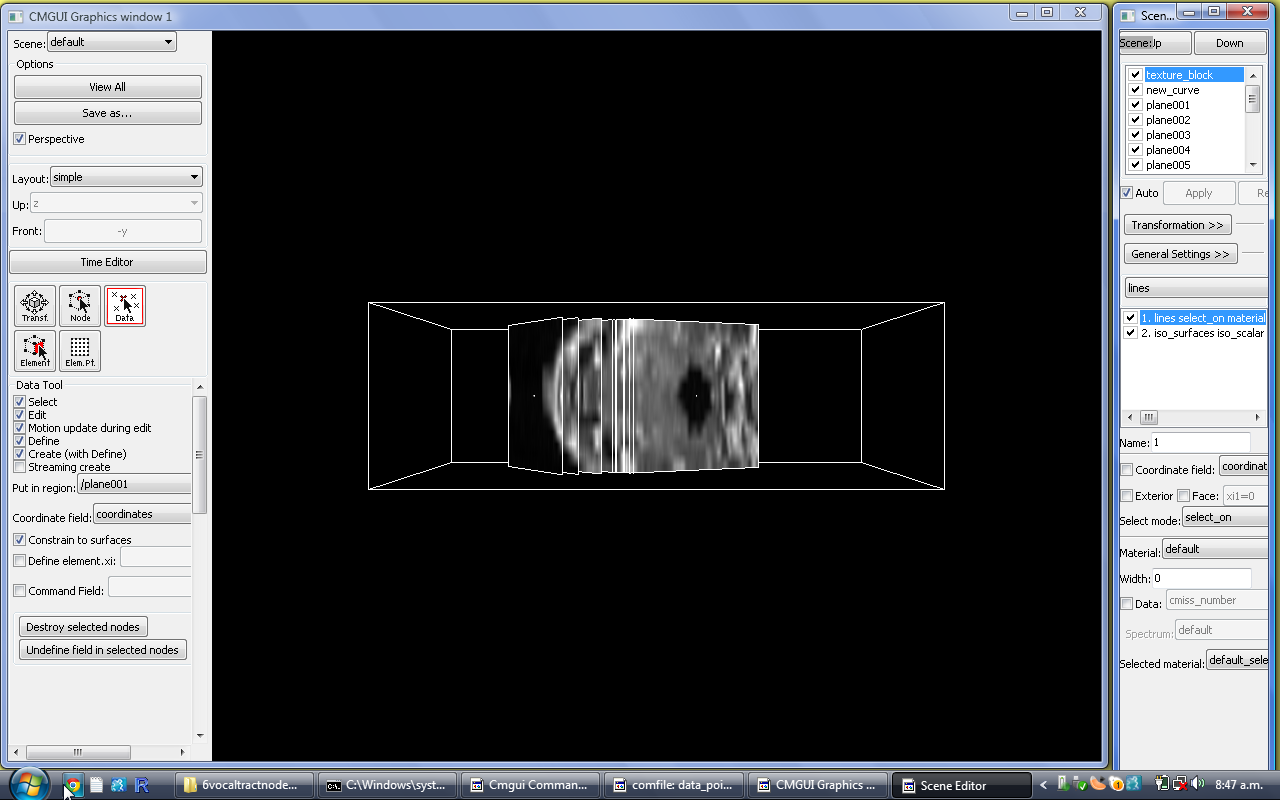


Figure : Recommended Screen Set-up

1. Deselect the “perspective” tick box. To enable rotation (left mouse button), translation (central) and zoom (right) of the image, select the “Transf.” button as shown in *figure 47*.

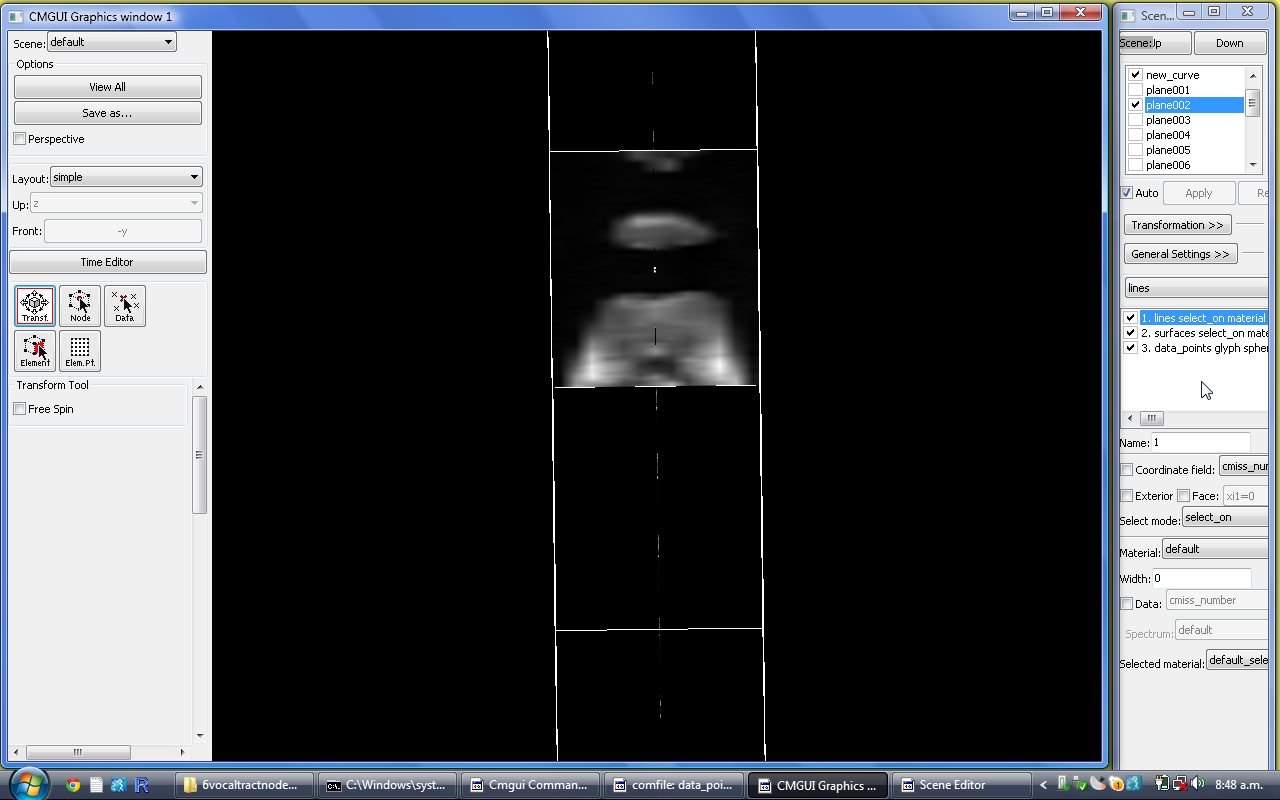


Figure : “Transf” to rotate, translate and zoom.

1. Select the “Data” button (*figure 48*) and choose the plane you want add data points to. The plane selected for editing in the Scene Editor window should match this, otherwise data will be written to the wrong plane.

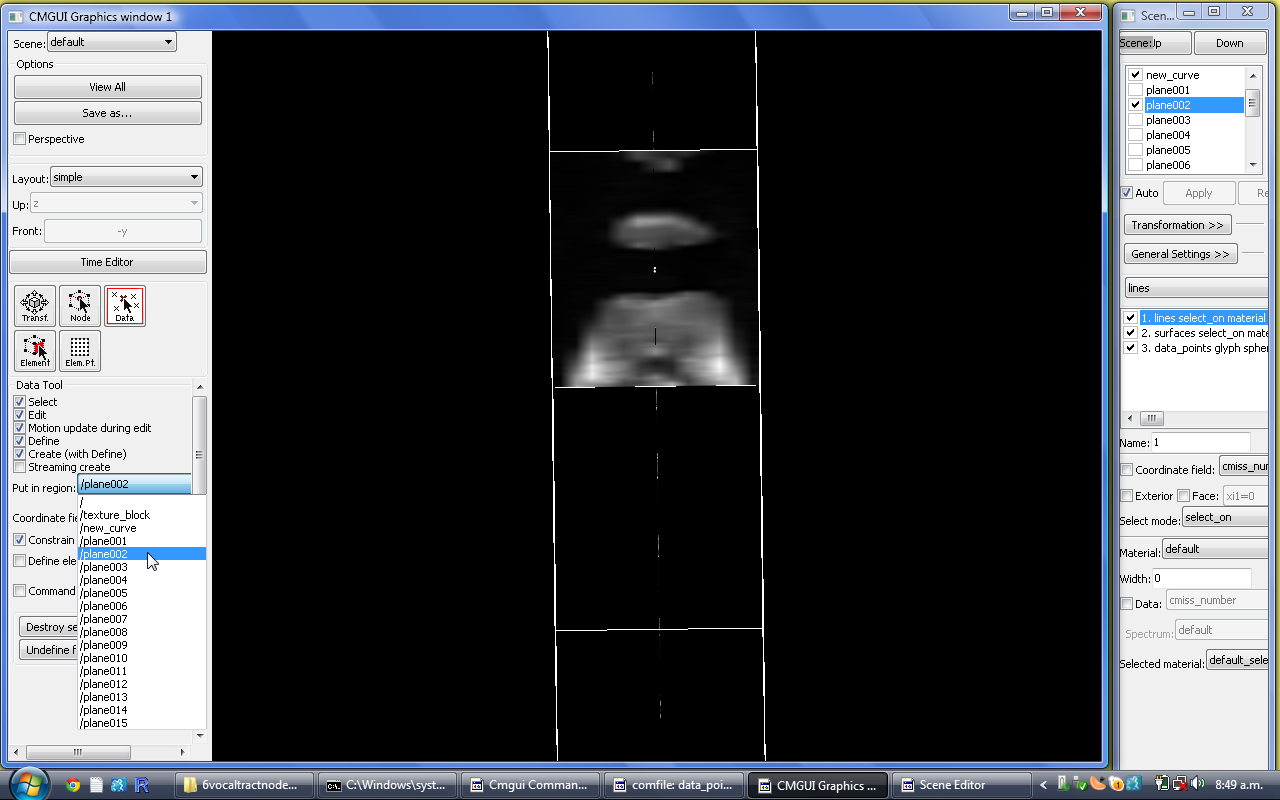


Figure : Adding data points to the correct plane

1. Place the data points around the vocal tract (do it sequentially i.e. only clockwise or anticlockwise). For the oral cavity start at plane002 (the first plane should only touch the tips of the lips)

**Troubleshooting**: The image does not allow for an enclosed shape (common for plane 2/3).   
Use the shape of the inner curve to approximate an enclosed shape (*figure 49*).

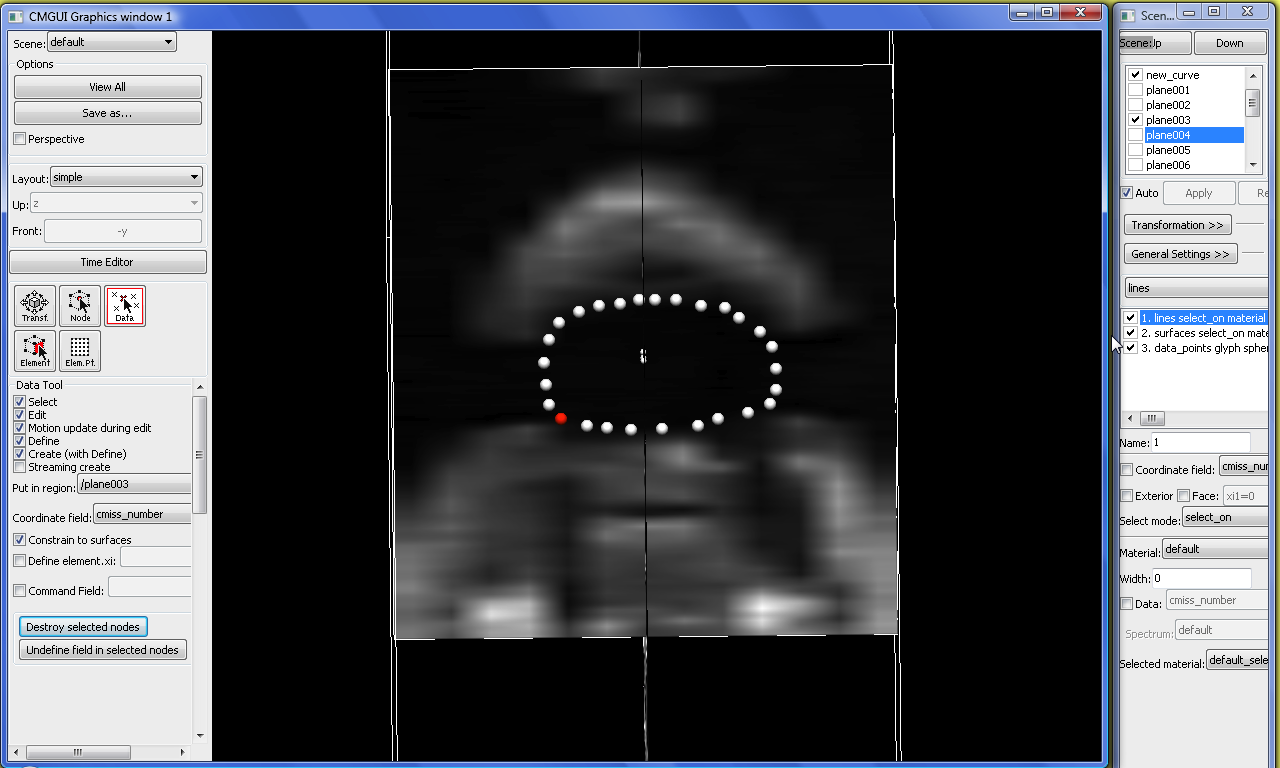


Figure : Enclosing the vocal tract

**Troubleshooting**: What if I make a mistake?   
You can select a number of nodes by holding down the Ctrl button and delete them by pressing “Destroy Selected Nodes”, highlighted in *figure 49.*

**Trouble shooting**: The image contains strange shapes (oral cavity)

It is likely that these shapes are created by the teeth and/or the placement of the tongue. Treat teeth problems as shown in *figure 50*.

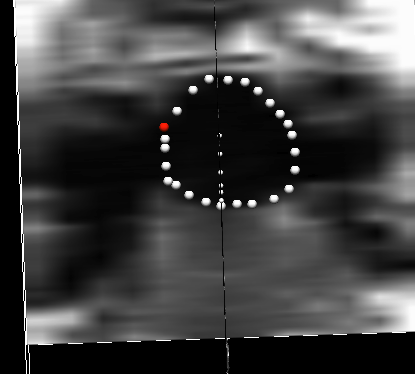
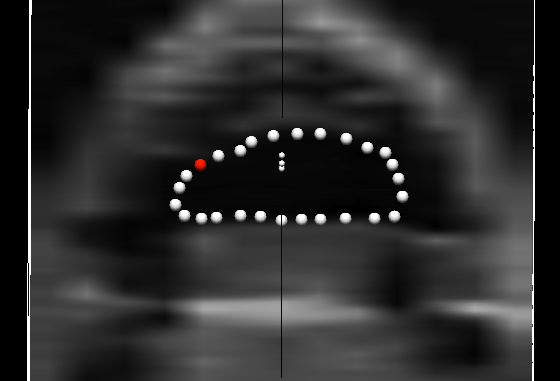


Figure : Three examples of avoiding measuring teeth

The gap created by pulling the tongue back/teeth should be ignored using vertical symmetry with the roof of the mouth about the central node *(figure 51)*

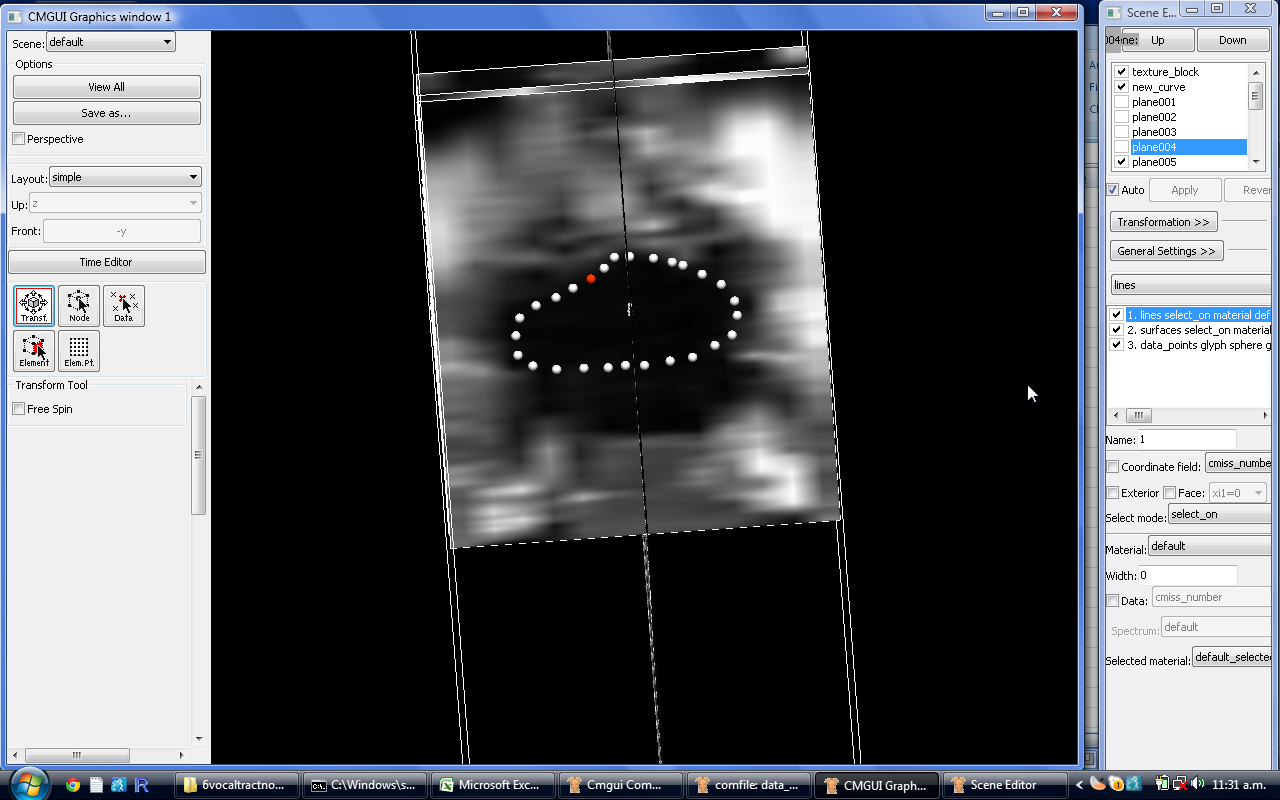


Figure : Vertical symmetry

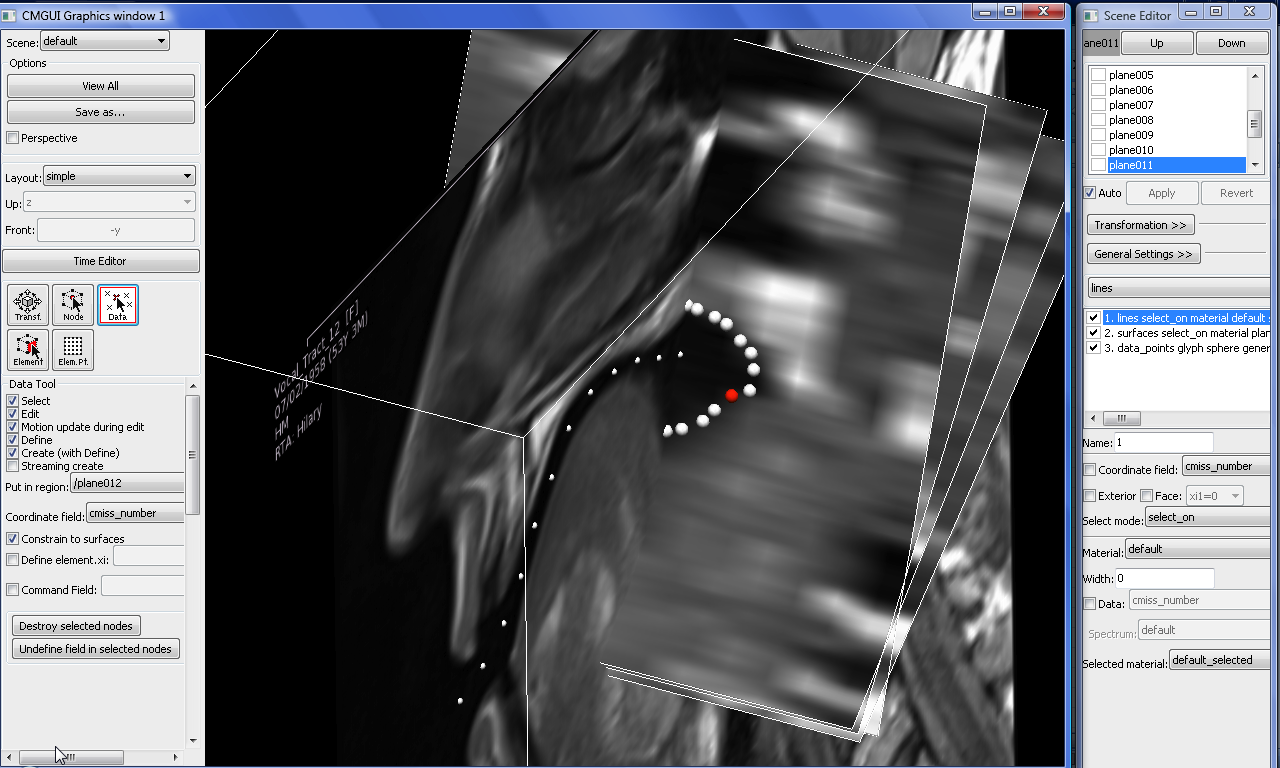
If required, rotate (remember to select “Transf.” button) the image and check that the data point placement aligns with the mid-sagittal plane as shown in *figure 52.* 

Figure : Rotating image to check data point

**Troubleshooting**: Extra ‘vocal tract’ shown in the pharyngeal cavity

The black space as shown in *figure 53* is due to the position of the epiglottis. Data points should only include the main tract.

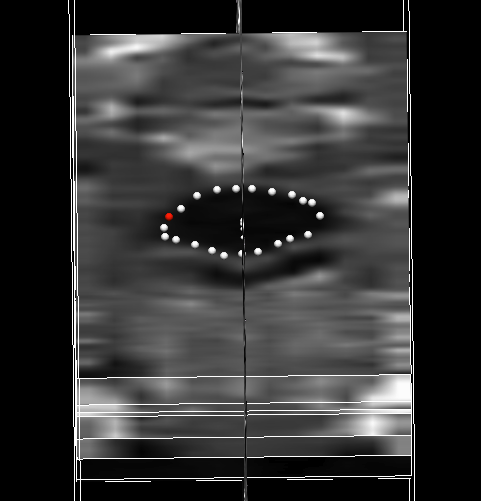


Figure : Appearance of epiglottis in cross-section

**Troubleshooting**: Part of the image is blurry.

Use a rule of symmetry to create an enclosed space from the blurry image as shown in *figures 54.* It may also help to rotate the image and compare the data point placement with the mid-sagittal image.

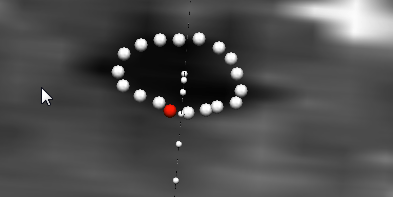
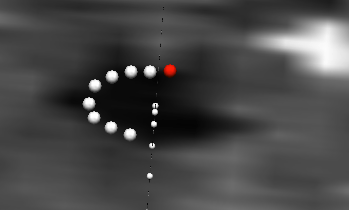


Figure : Using symmetry for blurry regions

Very blurry regions (especially near glottis) may require use of the mid-sagittal image to approximate the shape.

1. Leave Plane 15 in the pharyngeal cavity empty. This should have a zero value in order for the vocal tract to be approximated as a closed ended tube.

Note: the best way to get the data point placement right is to practice on different vocal tracts and vowels to gain a clear understanding of the expected shapes and anatomy.

1. Finally type in write\_to\_file in the cmgui command line to save the placed data points.
2. The data points will be saved as .exnode in the folder ‘exnodefordisplay’ and .exdata in the folder ‘exdataforareacalculation’

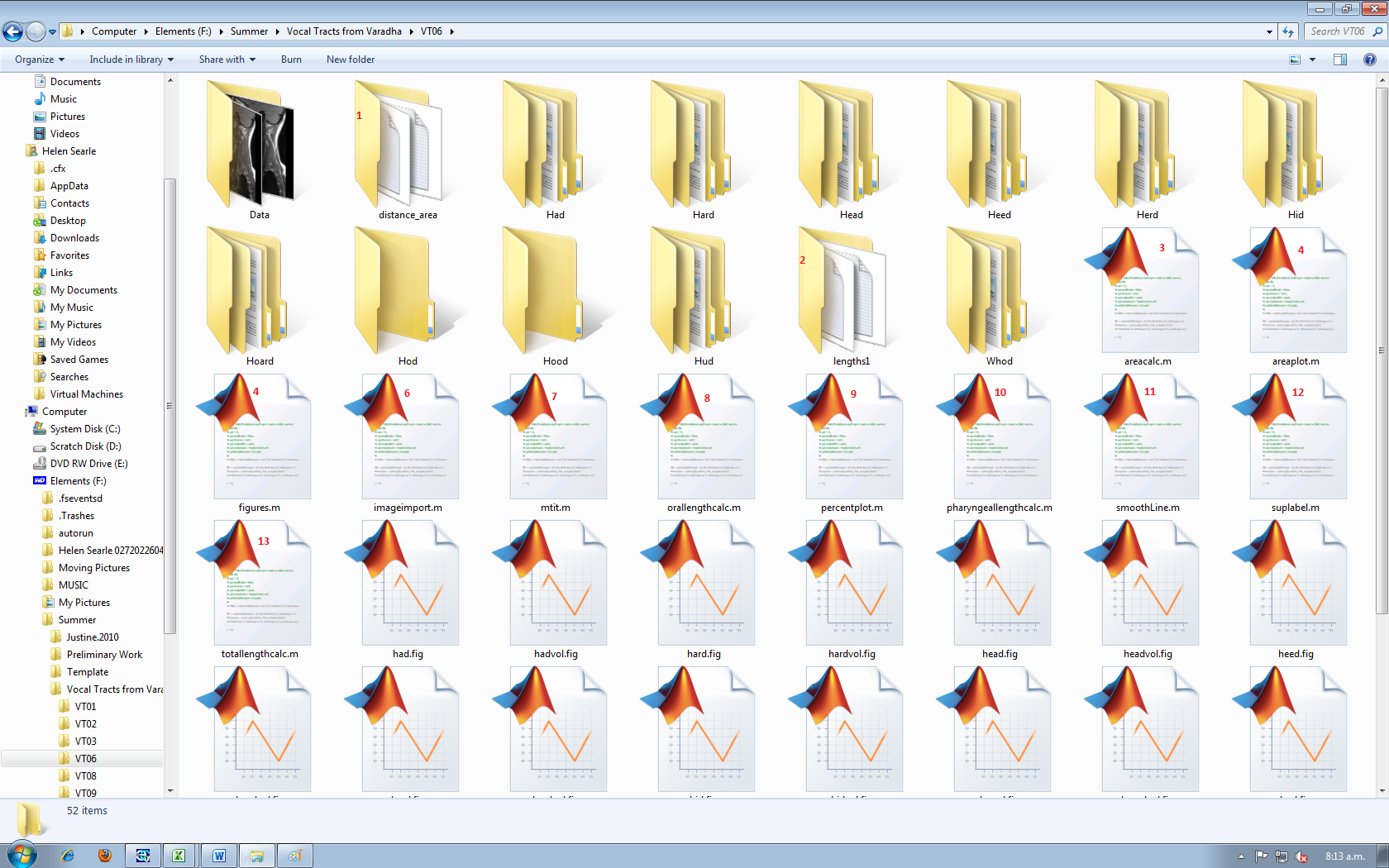
Appendix F: Matlab Code

Figure : Example folder for individual speaker

Table : Explanation of Matlab files

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Function** | **Run this directly?** | **Parent file** |
| 1 | Stores text files with distance vs. area data note: open in Notepad ++ for full formatting | - | - |
| 2 | Stores text files with length data | - | - |
| 3 | Takes CMGui data from the area.txt file (step 7-vocal tract measurement). Checks for discrepancies, averages the final node of the oral file and the first node of the pharyngeal file and makes the final node zero (closed tube approximation). | No | areaplot.m |
| 4 | Collates all length and area data and produces a plot of distance vs. area for each vowel in the vowel space. If mid-sagittal images also required, first run imageimport.m. | Yes | - |
| 5 | Produces a plot with all the figures. Must be run once imageimport.m has been run | Yes | - |
| 6 | Imports images from “Data” folder and creates a plot of the vowel space. | Yes | - |
| 7 | Creates an overall plot title (imported function) | No | areaplot.m |
| 8 | Inputs data from curve.exnode file (step one-vocal tract measurement), uses Pythagoras to calculate the oral length and outputs it to a text file in “lengths1” | No | areaplot.m |
| 9 | Creates a plot of the vowel space with oral and pharyngeal percentage volume ratios. | Yes | - |
| 10 | Inputs data from curve.exnode file (step one-vocal tract measurement), uses Pythagoras to calculate the pharyngeal length and outputs it to a text file in “lengths1” | No | areaplot.m |
| 11 | Existing function, smooths area vs. distance line and creates a matrix of data which is plotted. | No | areaplot.m |
| 12 | Creates axis titles for overall plot (imported function) | No | areaplot.m |
| 13 | Combines the oral length and pharyngeal length files and outputs the total length to a text file contained in “lengths1” | No | areaplot.m |

# Appendix G: Speaker Data

Table : Speaker Data

|  |  |  |  |
| --- | --- | --- | --- |
| Code | DOB/Age | Age Bracket | Gender |
| Vocal Tract 1 | 1945/65 | C | Male |
| Vocal Tract 2 | 1945/65 | C | Female |
| Vocal Tract 3 | 1985/25 | Y | Male |
| Vocal Tract 6 | 1985/25 | Y | Female |
| Vocal Tract 8\* | 1964/ 45 | M | Female |
| Vocal Tract 9 | 1964/45 | M | Male |
| Vocal tract 11 | 1938/67 | C | Female |
| Vocal tract 12 | /45 | M | Female |

\*This data was not included in the volume comparison but area-distance plots are provided. The area-distance plot for this speaker displayed the expected shapes; however they were much less defined. This was assumed to be due to the difference in data collection (VT08 was the pilot study with 18 slices rather than 13 and a much longer sustained vowel required) and data was removed from the study.

Appendix H: Additional Plots 