Comparison of vocal tract shape modelling methods: MRI vs. AR

Created by Hsiao-Tien Fan

Supervised by Dr. Catherine Watson

A thesis submitted in fulfilment of the requirements for the degree of Masters of Engineering, The University of Auckland, 2013

Abstract

Investigation was carried out on the vocal tract structural data obtained with the magnetic resonance imaging and the acoustic reflectometery techniques during the vocalization of vowels. This was carried out as a determination of the merits of the data acquisition techniques in question. This investigation included data for both techniques from 5 speakers.

Vocal tract structural data was captured by the magnetic resonance method as a series of sagittal images which is converted into cross-sectional area functions through manipulation of the images in the open source modelling tool of CMGUI [http://www.cmiss.org/cmgui]. The data collected by the acoustic reflectometer is in the form of cross-sectional areas along the vocal tract. Voice recordings for each target vowel were also collected.

The cross-sectional area functions obtained by the two methods are processed, and with the application of the lossless tube model and the linear predictive coding method, the spectrum and resonances of a give vocal tract structure was obtained. The length and resonance values of each vowel was recorded and analysed.

Vocal tract shapes obtained from the two methods were compared and discussed, and it was found that both methods were successful in capturing the expected vocal tract geometry of the target vowels, though a few compromises to the shapes from the acoustic reflectometery method was observed due to the compromising nature of the measurement mouthpiece.

The resonances deduced from the vocal tract shapes were compared to the formants extracted from recorded speech. It was found that the magnetic resonance method yielded a more accurate estimate of the formant values (9 out of 11 monophthong with reasonable estimations) while the

acoustic reflectometery method was much less accurate (3 out of 9 monophthongs with reasonable estimation)

It was concluded that while the acoustic reflectometery method was able to capture the general shape of the vocal tract, only a number of vowels were able to be modelled accurately enough for the calculated resonances to be comparable to the formants of real speech. However, future research may allow for methods with counter acts the compromising effects of the mouth piece, and yield more promising results.

Acknowledgements

First of all, I would like express my deepest appreciation to Dr. Catherine Watson. I could not have asked for a more enthusiastic and caring supervisor. Her enthusiasm in the topic and dedication in providing me with guidance has made my experience through the duration of my program both enjoyable and rewarding.

I would also like to thank my parents, who have supported me both emotionally and financially throughout my degree. It is because of them that I have the opportunity to continue to better myself with further study.

Finally, I would like to say thank you my sister, who assisted me in the tedious task of proofreading the final thesis. It cannot be easy to read nearly a hundred pages on a topic she has no knowledge of nor interest in, and for that, I am grateful.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

An area of interest in the study of the vocal tract and speech production is the effect of aging on the various aspects of speech. From a perception point of view it is known that as an individual ages, audible changes in voice characteristic and quality occurs. These differences can be caused many different factors, including the psychological state of the speakers and the physical changes to the vocal tract structure as the speaker ages. This has been confirmed in studies, which have shown that the characteristics and quality of speech vary in speakers of different ages (Gregory, Chandran, Lurie, & Sataloff, 2012) (Dehqan, Scherer, Dashti, Ansari-Moghaddam, & Fanaie, 2013) (Sataloff, Caputo Rosen, Hawkshaw, & Spiegel, 1997).

The vocal tract, which is responsible for speech production, is an acoustic chamber which determines the nature of the sound it produces with its structure. Therefore, any audible change in speech characteristics is likely to be accompanied with a structural change in the vocal tract. To verify and further study how this occurs, geometric data of the vocal tract shape must be collected. With this data, 3D representations of the vocal tract can be reconstructed and compared, from which differences can be identified. Through analysis, it may be possible to attribute these differences to the discrepancies between the speech qualities of speakers of different ages. With this objective in mind, this study sets out to investigate modern imaging techniques suitable for obtaining 3D structural vocal tract data.

With the invention of modern non-invasive imaging techniques, it has become a popular practice for studies modelling vocal tract structure to first take 2D images and then converting them

to 3D data. The most common imaging techniques include the X-ray, the CT scan, the MRI and the ultrasound. While in the past X-ray and CT scanning have been widely used for vocal tract studies, their usage has been significantly decreased due to radiation risks(Baer, Gore, Gracco, & Nye, 1991). The MRI and ultrasound on the other hand have no known side effects. The ultrasound method has the advantage that the result is yielded instantly and is dynamic in nature. Both the MRI and ultrasound are now used frequently in studies for collecting structural vocal tract data.

While the MRI yields accurate and detailed data, its drawback is the time consuming nature of its scanning process. This causes discomfort for the participants, especially for the elderly. Not only this, the MRI technique requires the participants to be lying on their backs, which may compromise the structural integrity of the vocal tract. Having this in mind, in order to research a field that may includes participants of a higher age group, it is important to explore different measurement techniques to minimise discomfort for the participants.

The acoustic reflectometery (AR) technique was chosen as a comparison to the MRI method. The acoustic reflectometer is a non-invasive acoustic device capable of detecting and recording the structural data of a cavity (A. Xue, 1999) (Steve An Xue & Hao, 2003). Though not extensively used in acquiring vocal tract structural data, the acoustic reflectometer has features such as having a fast data acquisition time and the participants are seated upright, which contrasts the features of the MRI.

Acoustic reflectometery, with the ease and speed at which the data is collected, is potentially ideal in studies requiring a large amount of data. The question is how well the vocal tract cross-sectional profiles compares to those obtained from analysis of images of the actual vocal tract physiology (i.e. the MRI). The aim of this thesis is to compare vocal tract cross-sectional areas obtained from the same speakers as well as the resonances deduced from these area functions via the two different methods and determine the merits of the AR process.

To understand the relevance of the comparison it is important to gain an understanding of the theories behind speech production, as well as the system structure. In the following sections, these features will be outlined.

1.2 Chapter overview

This chapter begins with the description of the vocal tract anatomy, outlining the structures present in the vocal tract and the position of such features. The principle of how the vocal tract structure produces sounds is described, and the methods of displaying the properties of these sounds in the form of spectral plots are presented. Following this, an overview of vowels, which are commonly used in studies investigating vocal tract structure using imaging techniques, is given. A brief description of the vocal tact modelling methods applied to this study is presented, along with the assumptions that have been made for the model to be applicable to the data acquisition methods of interest.

1.3 Vocal Tract Anatomy

The vocal tract and the larynx are both part of the structure responsible for the production of voice. For humans the vocal tract is comprised of several elements, some more important than others, which are situated around the front part of the head extending back towards the neck. The vocal tract is defined to be between the glottis and the lips, with the main aspects being the nasal cavity, the oral cavity, the pharynx. The larynx, which is the voice box, is joined at the end of the pharynx. An illustration of the vocal tract features is shown in Figure 1.1.

The oral cavity consists of the lips, teeth, and tongue. It starts at the lips and extends towards the back of the mouth. The nasal cavity is located directly above the oral cavity, which are separated by the hard palate and the soft. The main function of the nasal cavity is to prepare the air entering the respiratory system. This includes warming, filtering and humidifying the air, which is achieved by the large surface areas within the cavity. For humans, the nasal septum separates the nasal cavity into the right and left airways (Gray, 1918).

Both the nasal and oral cavities connect to the pharynx, which is the passage way that allows the passing of food or air from the upper cavities into the trachea and oesophagus. The opening

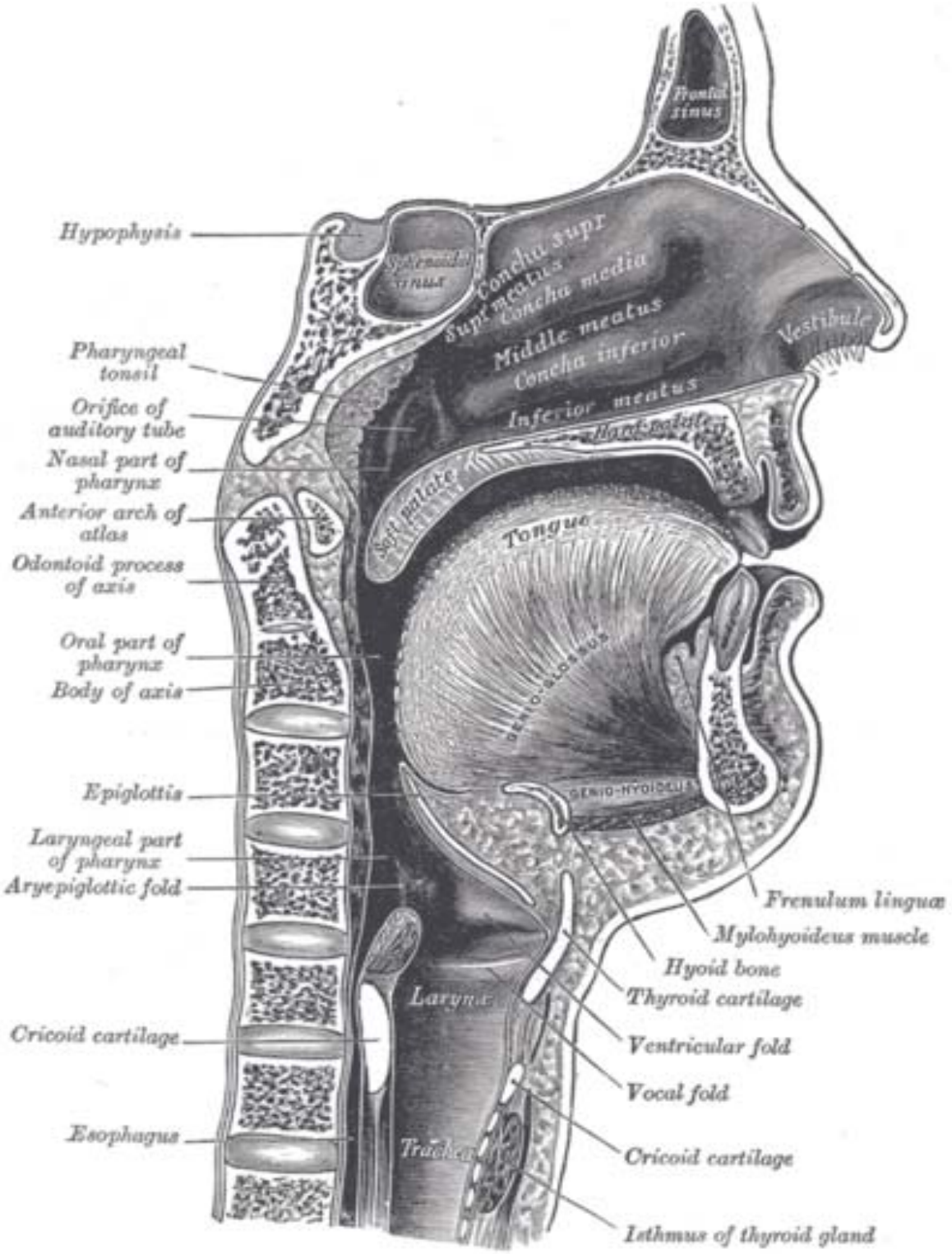


Figure 1.1: Illustration of the upper airway features. Reproduced from (Gray, 1918)

from the pharynx to the nasal cavity can be closed by moving the soft palate. The pharynx extends down the throat until it connects to the larynx and oesophagus. The opening for the oesophagus, which leads to the stomach, is closed until the event of swallowing.

The cartilaginous structure called the larynx, or more commonly known as the voice box, is located in the neck between the height of the third and sixth cervical vertebrae and is responsible for sound production. The larynx encloses the vocal folds, which are two flaps of membrane stretched across the interior of the larynx. This structure can be observed in Figure 1.2. The gap between the vocal folds, known as the glottis, can be varied in size by moving the muscles controlling the vocal folds. This affects the amount of air which passes through the larynx, which subsequently affects the pitch and loudness of the sound produced.

Aside from the described features, there are several cavities such as the space behind the epiglottis, the ventricular appendix and the piriform sinuses which make up the local structures of the vocal tract. While these features do contribute to the chamber resonance of the vocal tract, their influence is not significant, and can be ignored when modelling the vocal tract as a curved tube with varying cross-sectional areas. (Gray, 1918) (B.H. Story, 2002) (Bier, 2003)

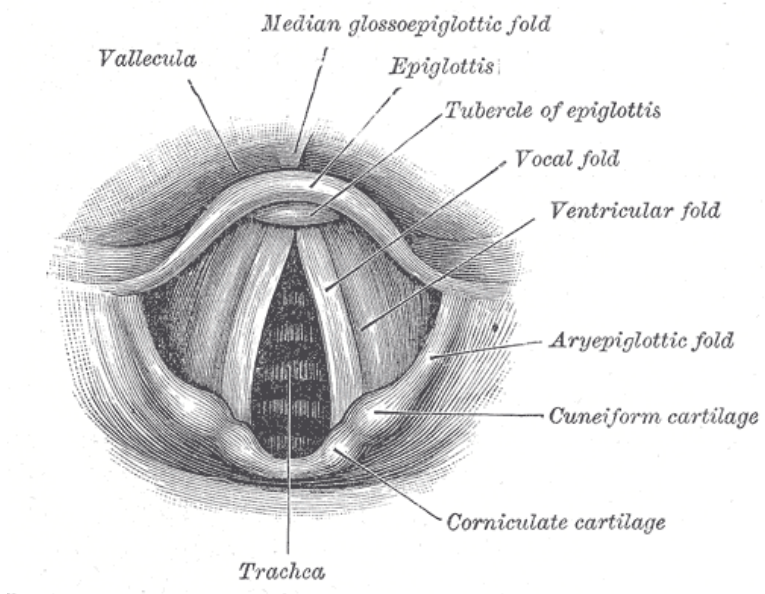


Figure 1.2: View of interior of larynx through a laryngoscope. Reproduced from (Gray, 1918)

1.4 Speech production

In the process of speech production, sounds are produced by the movement of air in the cavities within the vocal tract. Various types of sounds may be achieved by the manipulation of articulators to change the shape and structure of the tract. A collection of these sounds are produced in succession to form recognisable speech.

Speech sounds are commonly divided into two types: vowels and consonants:

Vowels

- When vowels are produced, the vocal tract is unobstructed, so there is no restriction of airflow. In general, vowels are voiced speech sounds, meaning that they result from the air pulses originating from the glottis being resonated in the vocal tract. For the voiced speech sounds, the vocal tract remains relatively unchanged through the duration of the vocalisation, meaning that the structure is time invariant. For this reason vowels are suitable for investigation within the scope of this study, as the time invariant nature of the vocal tract allows data acquisition techniques such as the MRI and acoustic reflectometery to be used for the capture of its structural data.

Consonants

- Consonants are harder to define, as they include semi vowels and laminar flow, but they are mostly produced with constrictions in the vocal tract which causes an obstruction to the air flow. These are not described in detail as they are not of interest to this study. (Please refer to 'Elements of Acoustic Phonetics' by Peter Ladefoged for a full description)

The components involved in a voice production system include the lungs and the upper respiratory tract, which includes the nasal cavity, the pharynx and larynx. The lungs provide the source of airflow for the duration of the voice production. For the production if a vowel sound, air

exits the lungs, through the trachea, and passes through the vocal folds in the larynx causing them to vibrate. This creates a sonic pulse composed of various frequencies to be sent into the pharynx and into the oral cavity. Depending on the geometry of the vocal tract, the natural frequencies of vibration within the chamber varies, and thus amplifies and attenuates different components of the sonic pulse. This directly reflects the resulting sonic vibrations which reaches and radiates from the lips (Bier, 2003).

To produce a different sound, the geometric structure of the vocal tract needs to be changed to provide different natural frequencies of vibration. This can be achieved with the movement of the articulators such as the lips, the jaw, the tongue and the velum. The velum is capable of closing the passage from the pharynx to the nasal cavity, thus preventing sounds of a nasal nature. By combining different orientations of the articulators, it is possible to produce the required deformations in the vocal tract to produce the required the phonemes used in speech. (Ladefoged, 1995)

1.5 Vowels - Overview

Vowels are suitable for vocal tract studies involving static imaging techniques, due to their time invariant nature during the vocalisation of the target sound. This is particularly true for monophthongs, which are pure vowel sounds, meaning that their vowel quality remains relatively fixed from the beginning to the end of their articulation. For the scope of this study, the 11 monophthongs of New Zealand English were the primary vowels of interest.

In studies requiring the vocalisation of vowels, it is common for words representative of the vowels of interest to be chosen instead of the target vowel being articulated in a standalone fashion. One of the contexts in which the monophthongs are presented is within the /HvD/ bracket, where the vowel is placed in between the consonants 'h' and 'd' to form single syllable words. This configuration was chosen based on the fact that the effect on vowel quality of this frame was minimal (Catherine I. Watson, Harrington, & Evans, 1998). In Table 1.1, the 11 monophthongs of NZ English and their corresponding HvD words are presented.

Phonetic Symbol	æ	a	ɛ	i	ɜ	I	ɔ	ɒ	ʊ	ʌ	u
HvD	HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	HOOD	HUD	WHO'D

Table 1.1: Phonetic symbols of the 11 monophthongs of NZ English and their corresponding words in HvD frame.

1.5.1 Phonetic vowel space

The phonetic vowel space is a visual representation of the auditory quality of different vowels and their relation to the positioning of the articulators specific to each vowel. Within this space, the vowels are placed in relative positions of how they are acoustically perceived compared to each other. For instance, [i], as in 'HEED', sounds higher than [ɛ] (as in 'HEAD'), and is therefore defined as a 'higher' vowel. The phonetic vowel space is defined by the corner vowels, the high front [i], high back[u], low front [æ], and the low back [a] (Ladefoged, 1995). In figure 1.3 the general shape defining the vowel space using these four vowels is displayed.

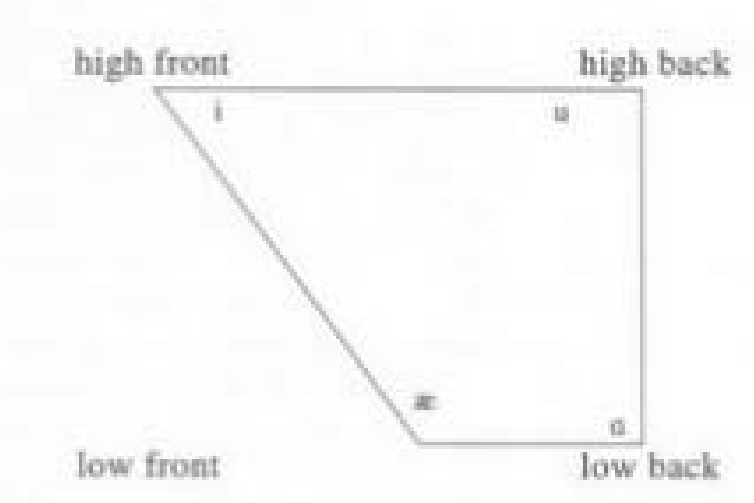


Figure 1.3: Vowel space defined by the four corner vowels (reproduced from (Ladefoged, 1995))

The positioning of the vowels within the vowel space also has a direct correlation to the articulator position associated with each vowel. For instances, the front vowels are articulated with the tongue positioned towards the front of the mouth. A summary for vowel types at extremities of the vowel space is presented below. The phonetic vowel space in its entirety is presented in Figure 1.4

Front vowels

- Front vowels are articulated with the tongue positioned towards the front of the mouth. The decreases the amount of space in the oral region of the tract and increases the volume of the pharyngeal region.

Back vowels

- Opposite to the front vowel, the tongue is positioned towards the back of the mouth when back vowels are articulated. This causes a constriction to occur in the pharyngeal region of the vocal tract and increases the volume within the oral cavity.

High vowels

- High vowels, also known as closed vowels, are defined by the amount of jaw opening during the articulation and how close the tongue is to the roof of the oral cavity. For high vowels, the jaw opening is small, and the tongue is positioned near the roof of the oral cavity, causing a constriction in that region.

Low vowels

- Low vowels, also known as open vowels, are articulated with a large jaw opening, and the tongue is positioned towards the bottom of the oral cavity. This generally creates a large oral cavity with constrictions in the pharyngeal region.

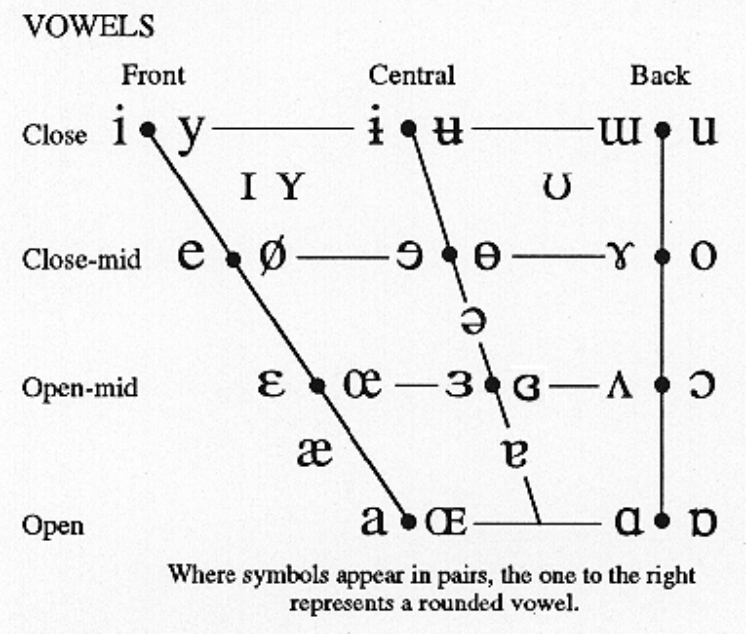


Figure 1.4: Vowels placed in the phonetic vowel space (reproduced from (Ladefoged, 1995))

1.6 Acoustic analysis of speech

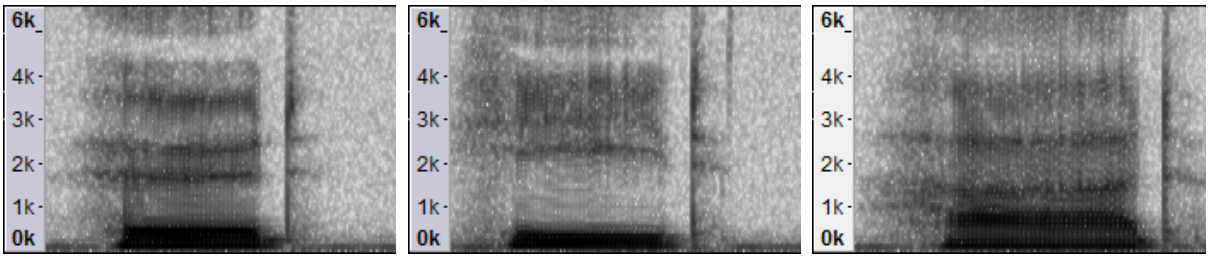
The acoustic analysis of speech involves the investigation of the physical properties of speech signals. Through this analysis, key elements in speech signals such as the pitch, loudness, formant and spectral information can be extracted. These acoustic characteristics are related to the configuration of the speech production process, such as the shape of the vocal tract for a given speech sound (Harrington & Cassidy, 1999). It is therefore possible, through acoustic analysis and the application of an appropriate mathematical model, to deduce the geometric structure of the acoustic chamber from the acoustic characteristics obtained from a given speech sound. This can also be performed in reverse to obtain the speech characteristics from an acoustic chamber of a given geometry. This is further discussed in the following section.

1.6.1 Spectrogram

Spectral analysis can be performed on any speech signal. It is of particular interest to the aims of this study, as it indicates the spectral density of the signal over a period of time, from which the dominant frequencies can be identified from the regions of the highest spectral density. This allows the frequencies of two different signals to be easily compared in a visual manner. In Figure 1.3 the spectrogram of three vowel sounds have been shown. It can be seen that each of the three spectrograms have dark bands spanning across the time axis. These bands mark the areas of high spectral density, indicating that the frequencies within these areas are of high amplitude.

For the vowel sound in 'HARD' (Figure 1.5(a)), dark horizontal bands can be seen at around 100 Hz, 500 Hz, 1800 Hz, 2500 Hz, and 3500 Hz. The lowest band is the fundamental frequency representing the pitch, and the higher ones are the formants. These formants are the defining characteristics of vowels which make them audibly distinguishable. In Figure 1.5(b), the dark bands can't be seen as clearly, but the fundamental at 100 Hz along with the first formant at around 300 Hz can still be recognised, while the second formant is situated at around 2200Hz. It can be seen that for all three vowels the dark bands are clearly situated in different places.

By comparing Figure 1.5(a) and (b) visually, it clearly shows that 'HERD' has lower fundamental frequencies than that of 'HEED's. In Figure 1.5(c), it can be seen that the bands are further apart than that observed in Figures 1.5(a) and 1.5(b), indicating more spread out formant frequencies. It is clear that the location of these dominant frequencies vary from one vowel to another.



(a) (b) (c)

Figure 1.5: Spectrograms of HERD (a) HEED (b) and HARD (c)

1.6.2 Formants and resonances

Formants are defined as the dominant frequencies present within a speech signal. For this study, this applies to the frequencies found within the vowel sounds chosen for investigation. As described previously, these formants can be obtained from performing spectral analysis on the speech signal being investigated (Titze, 1994) (Fant, 1970).

Resonances, which are often compared to the formants, are the calculated values which describe the behaviours of an acoustic chamber. It indicates the resonance frequencies which will be amplified by the given acoustic chamber when a sonic pulse is passed through it.

For the scope of this study, both the formants and the resonances are relevant. In order to compare two different scanning methods for acquiring the vocal tract geometry, we must first ensure that the model using the geometry yields reliable acoustic results. In order to verify the accuracy of this model, the resonance calculated from the vocal tract acoustic chamber geometry needs to be compared to the formants extracted analytically from an actual speech signal. Similar results in the formant and resonance values would indicate a model which closely reflects the acoustics of the vocal tract structure collected. In the next section the source filter model of speech, which is the chosen model for this study, will be outlined.

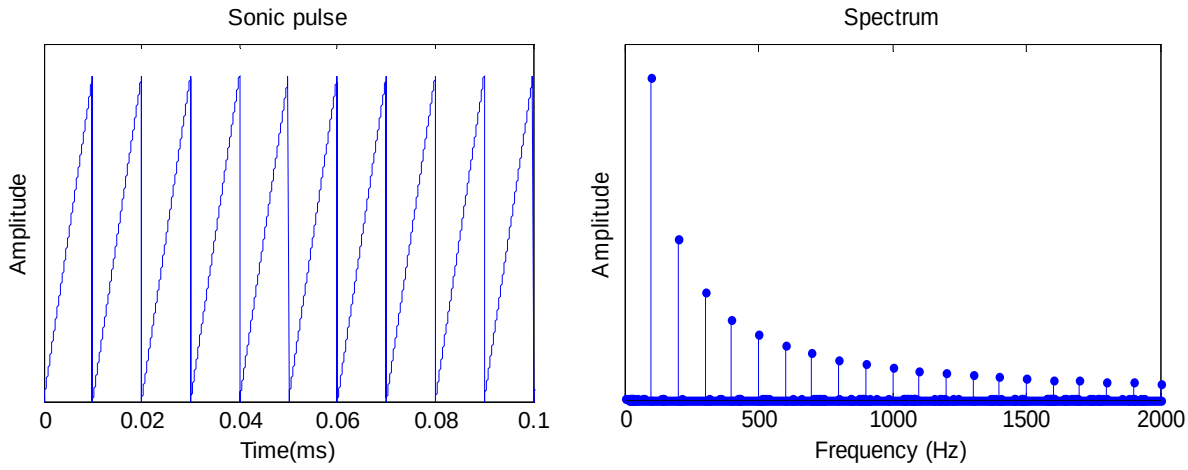
1.6.3 Source-filter model of speech

The source-filter model describes a time-invariant system where a defined source variable is modified by a filter function and an output is obtained. When this is applied to the concept of speech production, the sound being produced at the glottis can be seen as the source variable, with it being filtered by the vocal tract, and the final sound being emitted at the lips is the output (Rosen & Howell, 2010).

The source filter model can be represented in the form of an equation, where the output signal 'S' is the result of an operation between 'F' (a transfer function representing the vocal tract) and 'G' (the sonic pulses from the glottis)

$$G \cdot F = S$$

To model each of the aspects in speech production the different components must be realised is the appropriate forms. The signal from the source, or the input signal, is the sonic pulses generated by the vibrating glottis. The nature of this vibration allows the acoustic signal produced to be approximated as a sawtooth waveform, as shown in Figure 1.6(a). This information can also be represented in its frequency domain as an amplitude spectrum, as displayed in Figure 1.6(b) (Rosen & Howell, 2010) (Harrington & Cassidy, 1999).

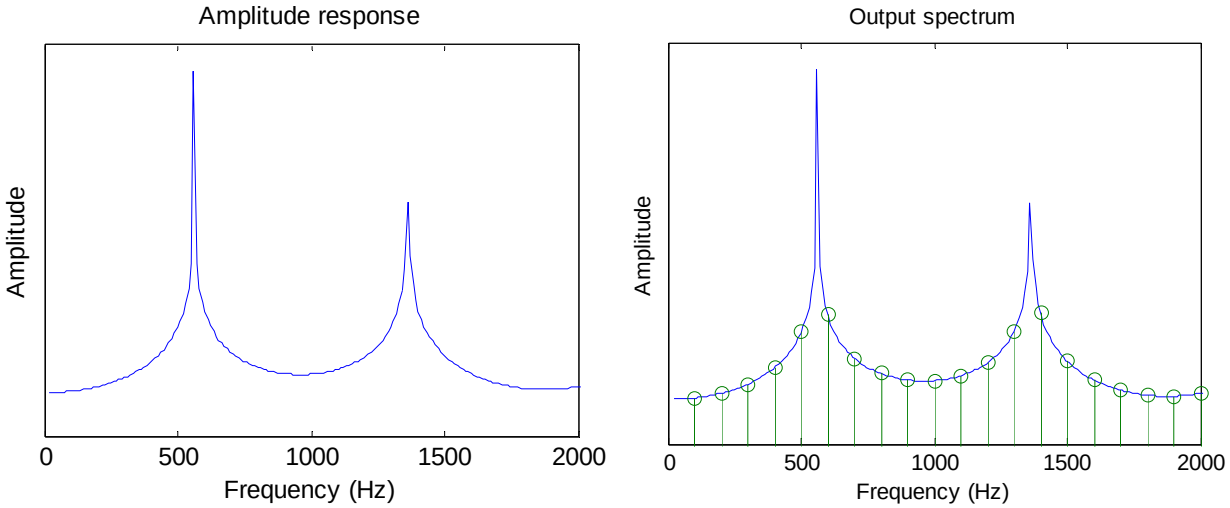


(a) (b)

Figure 1.6: Idealised representation of the sonic pulse in the time and frequency domain

For the vocal tract part of the model the tract structure determines the transfer function of the filter. The geometry of the vocal tract determines the natural resonance frequencies of its chambers, allowing the amplification of certain frequencies and the attenuation of others. The effect of this filtering process can be presented in the form of an amplitude response as shown in Figure 1.7(a).

Once the spectrum of the input signal (Figure 1.6(b)) and amplitude response of the vocal tract (Figure 1.7(a)) has been realised, it is possible to use the two to obtain the output spectrum, which describes the sound radiation from the lips (Figure 1.7(b)). The resonant frequency for this sound can then be easily identified as they can be visually located on the graphs as the peaks, which can then be compared to the formants extracted from the actual vowel sound.



(a) (b)

Figure 1.7: Amplitude response of the transfer function (a) and the spectrum of the output signal

This process can be summarised in Figure 1.8 where it is seen that the input signal is modified by the vocal tract transfer function and an output spectrum is obtained. The real challenge of using this model, however, is determining how to obtain the transfer function from the vocal tract geometry. A model which accurately describes the acoustic nature of the vocal tract shape needs to be used in order to obtain a reliable transfer function for the model. In the following section, models suitable for the portrayal of the vocal tract structure for the purposes of this study are described.

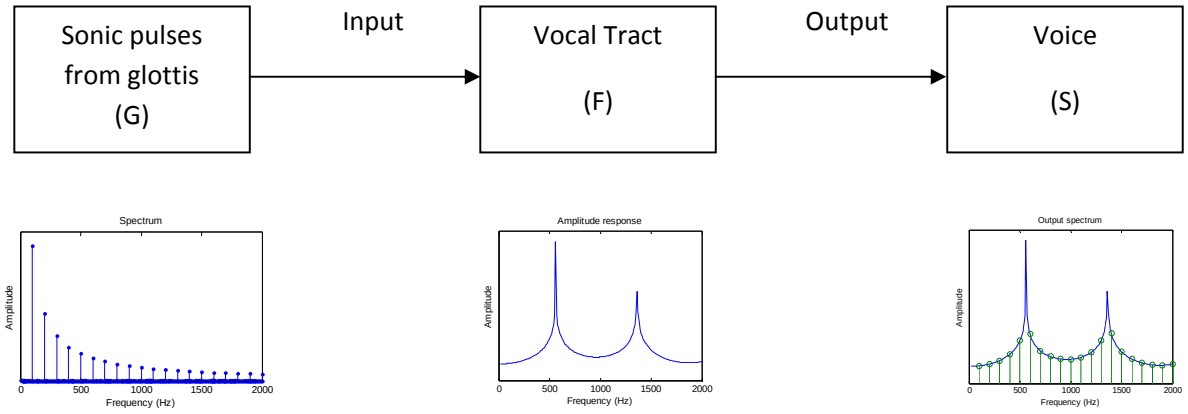


Figure 1.8: Application of a source signal to a transfer function in the frequency domain.

1.7 The vocal tract as an acoustic chamber

All acoustic chambers have fundamental frequencies that are specific to their geometric structure. These frequencies are determined by a number of factors, such as how sound reflects off the walls of the chamber, or how much resistance is met by the propagating sound waves. These factors contribute in the amplification and dampening of specific frequencies within the acoustic chamber (Rosen & Howell, 2010).

A simplified form of the vocal tract can be modelled as a tube of uniform cross-sectional area, open at one end and closed at the other and approximated to be 17cm in length for an adult male. In such a tube, sound waves of specific frequencies will resonate by being reflected at the end of the tube back towards the source. These frequencies can be calculated by Equation 1, where F_n is the instance of the formant, 'c' is the speed of sound in meters per second and 'l' is the length in meters

$$F_n = \frac{(2n-1)c}{4l}$$

However, a uniform tube model obviously does not take into account the effects of the varying levels of constriction along an actual vocal tract. To account for this, the lossless acoustic tube model is used (Harrington & Cassidy, 1999).

1.7.1 Lossless acoustic tube model

The lossless acoustic tube model is a widely used method for modelling speech production (Mullen, Howard, & Murphy, 2006). In Figure 1.9(a), an idealised representation of the vocal tract is presented, indicating the varying cross-sectional areas $A_{(X)}$ at various positions of the vocal tract. In Figure 1.9(b), a concatenated lossless acoustic tubes presentation of the vocal tract is shown. In this figure, it is seen that each tube has its own cross-sectional area A_n.

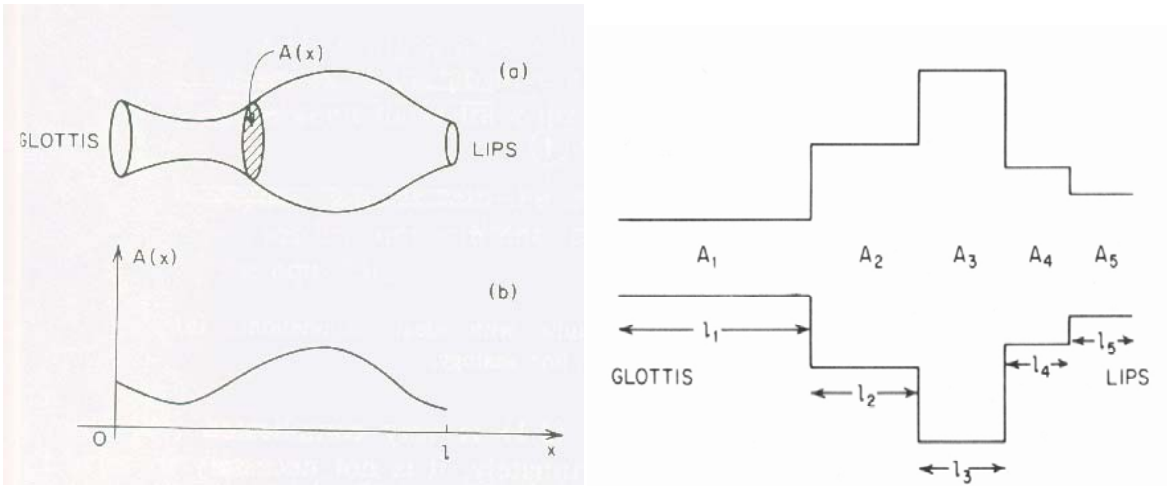
The cross-sectional area A_n for each tube is chosen to approximate the varying area function $A_{(X)}$ of the vocal tract. This is a similar approach to that of the uniform cross-section tube model, except that more tubes of varying cross-sectional areas are now used to capture a more realistic acoustic representation of the tract. It is obvious that, the higher number the tubes used, the more accurate the geometric representation of the actual vocal tract profile.

By applying established theorems, the cross-sectional areas of the tubes can be calculated into filter functions resembling the vocal tract (This will be discussed in Chapter 2). However, it is rare to use very high numbers of tubes in modelling exercises, because as the number of tubes increases, the more likely it is for the filter function obtained from the tubes to be unstable.

The vocal tract can be represented as a series of concatenated lossless acoustic tubes if certain assumptions are made.

- The vocal tract shape is time-invariant.
- There are no losses in the energy of the sound waves caused by friction, heat conduction or wall vibrations.
- Structural features such as the piriform sinuses and the ventricular appendix do not contribute to the resonance of the vocal tract chamber.

By using this model to represent vocal tract structural information it is possible, by the use of linear predictive coding, to obtain the transfer function representing the tract. This can then be applied back to the source filter model, from which an output spectrum can be identified for further analysis. In order for this to occur, cross-sectional areas of the vocal tract need to be obtained using the scanning methods designated to this study. In the next chapter, the process of obtaining these areas and how the process of analysis is carried out will be described.



(a) (b)

Figure 1.9: Cross-sectional area function of vocal tract (a) and lossless tube model of vocal tract (b) (Reproduced from (Rabiner & Schafer, 1978))

1.7.2 All-pole Filter model

With vowels being the speech signals of interest in this study and by modelling the vocal tract shape as a series of lossless acoustic tube, it is assumed that the qualities of these signals are largely determined by the resonances of the vocal tract shape. Having made this assumption, the all-pole filter model of speech can be applied to estimate these speech signals. The all-pole filter models the current signal as the sum of scaled past values and the input value, which can be expressed as

$$y[n] = -a[1]y[n-1] - a[2]y[n-2] - \cdots - a[k]y[n-k] + x[n]$$

In this equation, the coefficients a[k] represent how the vocal tract filters the speech signal. These coefficients can be related to the cross-sectional areas of the acoustic tubes in the lossless tube model with the technique of Liner Prediction Coding (LPC). Using this method, it is possible to calculate these coefficients, and apply it back into the filter model. This process will be described in Chapter 2. (Harrington & Cassidy, 1999)

2 Modelling the Vocal Tract

2.1 Chapter overview

This chapter describes the processes behind the modelling of the vocal tract for this study. An overview of the data collection techniques commonly used for collecting vocal tract structural data including the X-ray, CT scan, ultrasound, acoustic reflectometery and magnetic resonance imaging are described. The process behind extracting the vocal tract area function from the MR images using the open source package CMGUI is outlined. This is followed by a description of using the LPC method to obtain the formants and spectrum from the raw cross-sectional area data.

2.2 Modelling components

For the scope of this study, the vocal tract was modelled in the form of lossless acoustic tubes of various cross-section areas. This model was applied with the aim of obtaining the first three resonances from the vocal tract shape of different vowel sounds of interest. Two methods for collecting the 3D structural data of the vocal tract - the acoustic reflectometery and the magnetic resonance imaging method - provided cross-sectional area profiles along the length of the vocal tract. These cross-sectional area profiles are then processed into the form of acoustic tubes. The LPC (linear predictive coding) method was subsequently used to obtain calculated the resonance values for a given vocal tract shape, which can be used to compare the two different scanning methods. The modelling process used in this study can be broken down into a number of components. It is shown in Figure 2.1.

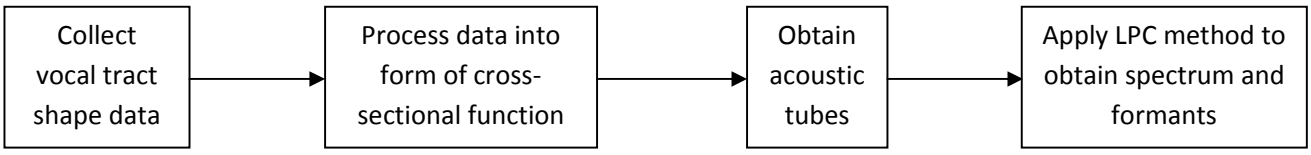


Figure 2.1: Flow diagram of the process involved in modelling the vocal tract with the LPC method.

2.3 Data acquisition techniques

Many studies have been conducted using imaging techniques as a non-invasive means to obtaining structural data on vocal tract shape. These include the X-ray, the CT scan, the MRI and the ultrasound. The technique of acoustic reflectometery has also been used. How these techniques have been used will be outlined in the following section.

2.3.1 X-ray

The vocal tract was modelled as a series of acoustic resonators of varying area by Fant (1970) using X-ray images. The features of the vocal tract were extracted manually from the X-ray scans. The images were not well defined at air-tissue boundaries, and the identification of the feature contours required much guesswork. X-rays are no longer used in recent studies, as repeated exposure to the radiation is a health risk.

2.3.2 CT scan

The CT scan provides high resolution images with a more defined air-tissue boundary than that of the MRI. It is capable of capturing teeth and has a faster scanning time than that of the MRI. As it is relatively quiet, there is less chance for a vocalisation of a target sound to be compromised by external noise.

In the study by Perrier, Boe, and Sock (1992), the CT scan was used to collect cross-sectional data on three vowels for the investigation of vocal tract area functions. B. H. Story, Titze, and Hoffman (1998) used CT images from two vowels to complement the MRI images used in their study, as they provided more structural detail of the piriform sinuses and provided data on the teeth. It also acted as a verification of the accuracy between the two different imaging methods. However the drawback of this method, like the X-ray, is its radiation risk.

2.3.3 Ultrasound

The ultrasound method has become increasingly popular within studies of the vocal tract structure. It has been used in many studies involving the tracking and mapping of the jaw and tongue position. Its dynamic nature allows the changes in the structural features to be tracked in real time. In these studies, it was possible to reconstruct the surface contours of articulatory features within the vocal tract such as the tongue. This method, however, has the restriction of having a limited scanning area. This means that it is

unlikely for more than one feature to be within the scanning area at the same time. (Stone & Davis, 1995) (Stone & Lundberg, 1996)

2.3.4 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique used in obtaining the structural data of a human body. Unlike the X-ray and CT scan, the MRI technique has no know ill effects towards human health, making it suitable for repeated use (Baer et al., 1991). The subject being scanned lies in the machines while it uses strong magnetic fields and electromagnetic radiation to distinguish different types of human tissue within a target plane and presents the results in a 2D image. The high resolution of the MRI technique allows detailed and accurate vocal tract structural data to be collected.

Baer et al. (1991) collected axial, coronal, or midsagittal MR images of the vocal tract during the vocalisation of four point vowels. Voice recordings were made of these vocalisations. Area functions of each vocal tract shape were extracted from the images, from which digital filters were derived. This was used to resynthesize the vowel sounds, which were compared both acoustically and perceptually to voice recordings.

Many MRI studies have a small number of participants and target vowels of interest, due to its time consuming nature. This presents a certain difficulty in the comparison of vocal tract features between subjects of different backgrounds, such as age and gender. In Yang and Kasuya (1994), MRI data of the vocal tract was collected from child, male and female subjects, from which the data gave an insight of the differences and similarities of the various regions of the vocal tract between the different participant groups.

Story et al. began a series of MRI related vocal tract structure studies in the mid 90's (B. H. Story, Titze, & Hoffman, 1996), gathering MR images and extracting structural features for comparison. The earlier studies were complemented with C-T scans as a means to compare and contrast the results to the MRI data. (B. H. Story et al., 1996) (B. H. Story et al., 1998). It was common for these studies to have only one scan of each target vowel. This means that it is difficult to test the validity and consistency of the method across different scans of the same vowel. It was shown in (B. H. Story, Titze, & Hoffman, 2001) and (Takemoto, Honda, Masaki, Shimada, & Fujimoto, 2006) that the speaker can produce different vocal tract shapes for the same target vowel produced. To address the issue of repeatability of

the MRI technique B. H. Story (2008) revisited data from two previous studies where the data was collected from the same speaker. Consistency between the two datasets was shown.

The MRI technique, however, is not without its drawbacks. The MRI technique has a relatively high scanning time which means that only sustained vocalisations can be studied. Not only this, the quality of the air-tissue boundary is largely dependent on the movement of the individual during the scanning period and the teeth are badly imaged. This causes difficulties in obtained a defined structure in the areas where the teeth are located. To identify the air-tooth and air-tissue boundaries, a thorough understanding of the vocal tract structure is required.

2.3.5 Acoustic reflectometery

The acoustic reflectometer is a non-invasive method for collecting structural vocal tract data. The reflectometer sends a series of sonic pulses down the vocal tract and the reflected response is recorded and calculated into the cross-sectional areas along the vocal tract. Using this method it is possible to quickly make repetitions on the measurements of each vowel without causing excess discomfort to the speaker, as each measurement takes only two to three seconds. The advantage of this method is that the data collected is presented directly in cross-sectional area, which makes it easy to access for the analysis

In contrast to the MRI, acoustic reflectometery has a much shorter data acquisition time, and since the results are readily presented in a cross-sectional area form, it requires no post collection processing before the analysis can begin. Due to its convenience, acoustic reflectometery is well suited for studies requiring large number of participants. In A. Xue (1999), data was collected from two groups of female speakers, 10 speakers within the ages of 33-48 and 12 between 50-66. It was found in this study that the older age group had a significantly larger total vocal tract volume than that of the younger group, while the length of the vocal tracts did not have a significant difference between the groups.

A similar study was carried out in (Steve An Xue & Hao, 2003) where 19 young and 19 elderly speakers were recruited from each gender. Similar results were found in this study, where the vocal tract lengths of the different ages were similar and the elderly had a higher vocal tract volume than that of the young. It was found that the elderly of both genders displayed similar difference in speech acoustics from their younger counterparts.

Acoustic reflectometery was also used by Steve An Xue and Hao (2006) where 120 subjects, evenly spread between white Americans, African Americans, and Chinese, were recruited. In this study,

the subjects were controlled for age, gender, height, and weight. It was found that the age and ethnicity of the subjects had an effect on the dimensions of certain aspects of the vocal tract. The aim of the study was adding to the anatomical database of the vocal tract for interested parties.

In the study by Tameem and Mehta (2004), the structural geometry of the vocal tract was collected using acoustic reflectometery and MR imaging techniques. The MRI data was used to reconstruct a 3D model of the vocal tract, which was subsequently compared to the model derived from the acoustic reflectometery data. It was found that the data recorded using the acoustic reflectometery technique was suitable for creating an estimate of the vocal tract shape.

In the study by C.I. Watson, Thorpe, and Lu (2009) acoustic reflectometery data was compared to the area function extracted from MRI data for four vowels for a single speaker. The cross-sectional area functions from both methods were processed acoustically to obtain the first three formants for each vowel. It was found upon analysis that the two methods yielded vocal tract of similar contours and the formant values derived from the area functions were of reasonable similarity. However, as there was only a single speaker and small number of vowels in the study it was difficult to conclude how wide spread, and consistent this similarity was.

Another study was presented by C. I. Watson and Hui (2010). For this study, vocal tract data was collected from 5 young and 5 middle aged speakers, each vocalising 9 monophthongs. It was found in this study that the older age group had a pharyngeal cavity with a larger volume than that of the younger group, a result similar to that obtained by Xue et al. Midsagittal MR images were also collected for this study. These images were presented as a visual representation of the vocal tract shape for each of the target vowels.

Thus far, the acoustic reflectometery method has not been used extensively in the study of vocal tract shape during the vocalisation of target vowels, and whether the vocal tract cross-sectional area function obtained from this data would suffice in providing an accurate estimation of formant values has not been determined. The applicability of the results across multiple speakers has yet to be determined. From this stand point, this study aims to investigate these points and provide an indication of the suitability of the AR method for the in modelling the different aspect of the vocal tract.

2.4 Cross-sectional area function of the vocal tract

After the vocal tract data has been collected, it must be transformed into the form of a cross-sectional area function. This is not an issue for the acoustic reflectometer, as the output of the measurement is already in the form of cross-sectional areas. This is not true for the MRI, for which the images obtained need to be processed before the cross-sectional areas for the vocal tract can be found. The challenge here is to build a 3D model of the vocal tract from the 2D MRI images, from which the vocal tract boundaries are identified and the cross-sectional areas are calculated (see for example (Baer et al., 1991) (Clément et al., 2007) (B. H. Story et al., 1998). For this study the approach used was first developed by Bier (2003) and was further adapted by C.I. Watson et al. (2009). The approach uses the open source software tool CMGUI [http://www.opencmiss.org/]. The process of using such a package to extract the cross-sectional area function will be outlined in this section.

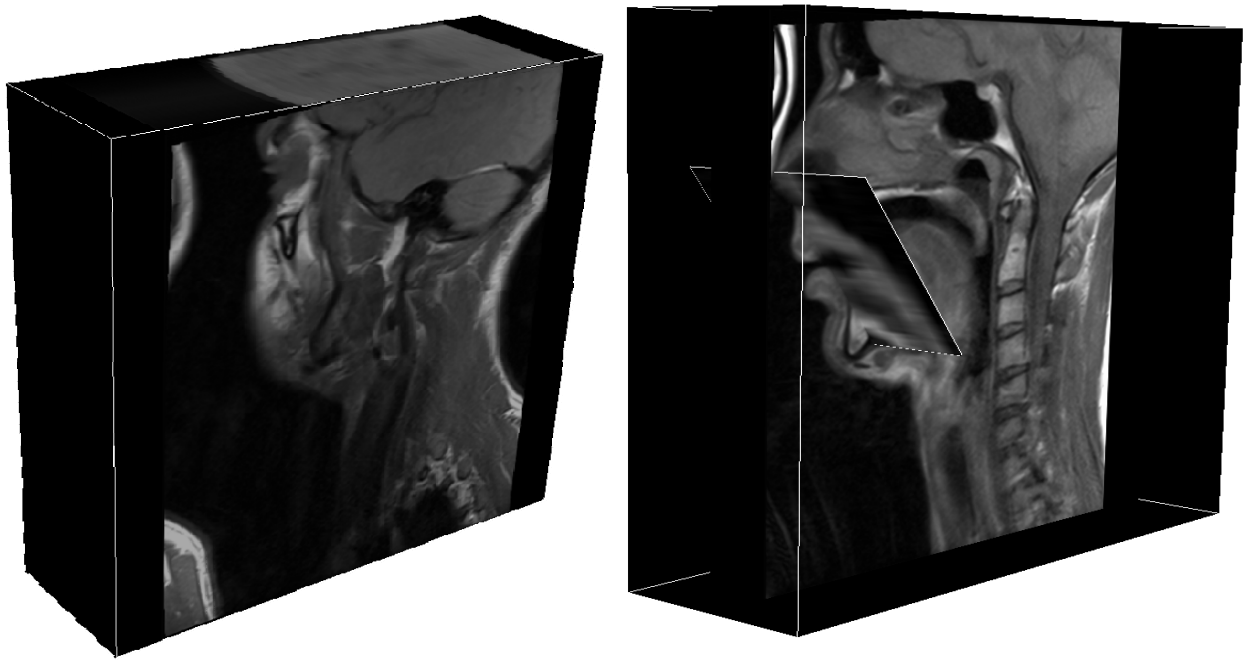
2.4.1 Creating the texture block from the MRI data

The MRI method captures information from a person by taking 2-dimentional images along of any plane of interest. These images can be taken in any orientation, though it is common for studies to take the images along the axial, coronal, or midsagittal planes (Clément et al., 2007) (Wismueller et al., 2008). The limitation of this method is that although images are taken at different depths along the chosen plane, which gives the structural details at different locations, it is inevitable for there to be gaps in between the images where there is missing information.

To compensate for this, by using CMGUI, a texture block is created with the images obtained from the scans by interpolating the information between each of the slices of a set of images. An example of this is shown in Figure 2.2(a). By doing this, a complete 3-dimentional representation of the areas scanned is created so that any point within the texture block can be selected to outline the vocal tract. This texture block removes the necessity to have scans in multiple planes, as the texture block can present data from any planar configuration. It is important to note that this information is the result of interpolation.

Using this texture block, it is possible to specify exact locations within the vocal tract in 3D space and observe structural information in the area. A plane of interest can be defined and the data within that plane can be presented for further analysis. This is shown in Figure 2.2(b), where a plane has been

identified within the texture block. Note that in order for the plane of interest to be seen only the midsagittal plane of the texture block is presented in this figure.

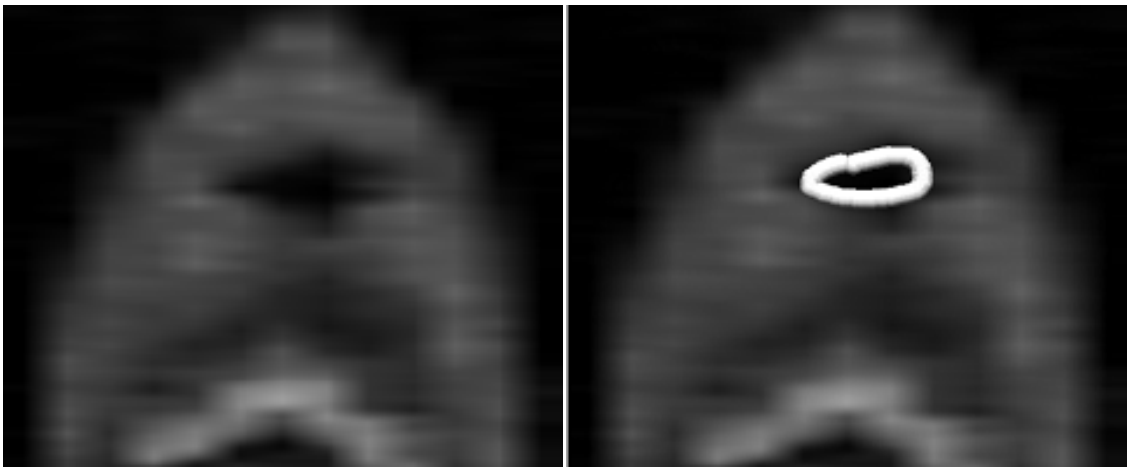


(a) (b)

Figure 2.2: (a): Texture block constructed from the MRI images, which provides a complete 3D representation of the area scanned. (b): A plane is chosen within the texture block which can be used to observe the structural detail of the vocal tract within that plane

In Figure 2.3(a), this defined plane is presented. As it can be seen from the figure, there is a black region in the centre of the image, which indicates the cavity of the vocal tract. It is possible, from this point, for the CMGUI to place nodes around the contour of the vocal tract thus identifying its boundaries. Figure 2.3(b) presents the selected plane with the vocal boundary identified. Once this is completed, the package is able to triangulate the area within the nodes, thus yielding the cross-sectional area.

By repeating this process at various points within the texture block, it is possible to obtain the cross-sectional area at different points along the vocal tract. By performing this on enough points, it is possible to have a cross-sectional area function detailed enough to accurately model the vocal tract for the purpose of this study. This is demonstrated in Figure 2.4 where a wire frame model of the vocal tract boundary identified at various points along the vocal tract is shown.



(a) (b)

Figure 2.3: (a) planar view of the structural data of the vocal tract and (b): the same image with the boundary of the vocal tract identified.

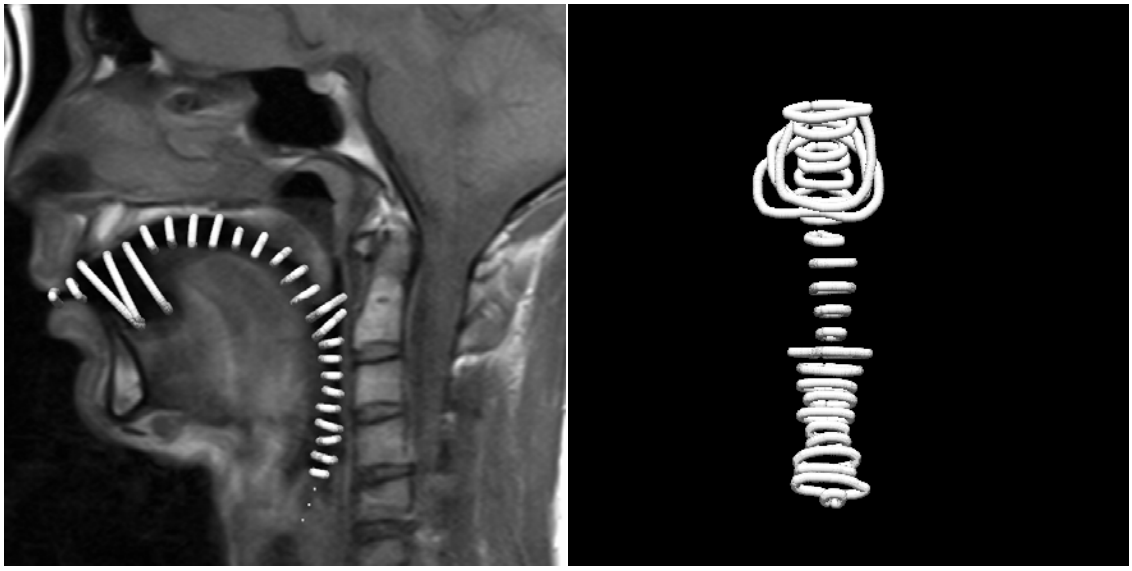


Figure 2.4: Wire frame model of the vocal tract marked in CMGUI.

2.5 Acoustic tubes

As mentioned in the previous chapter, in order to model the vocal tract as an acoustic chamber which takes into account of the various anatomical features such as constrictions at various positions of the tract, the cross-sectional area function needs to be transformed into the form lossless acoustic tubes. To

convert it into this form, the area function is divided into a number of segments, and the area value within each segment is averaged, giving an area value to each tube. This value can be subsequently converted into a tube radius. These tubes stack together to present an approximation of the vocal tract shape. A graphical representation of this is presented in Figure 2.4.

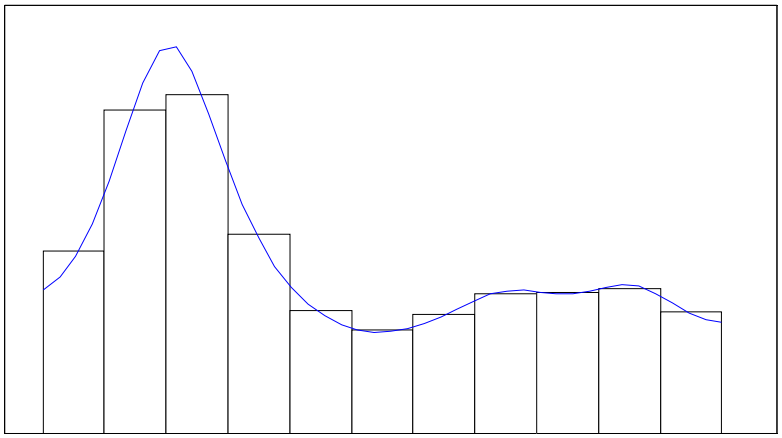
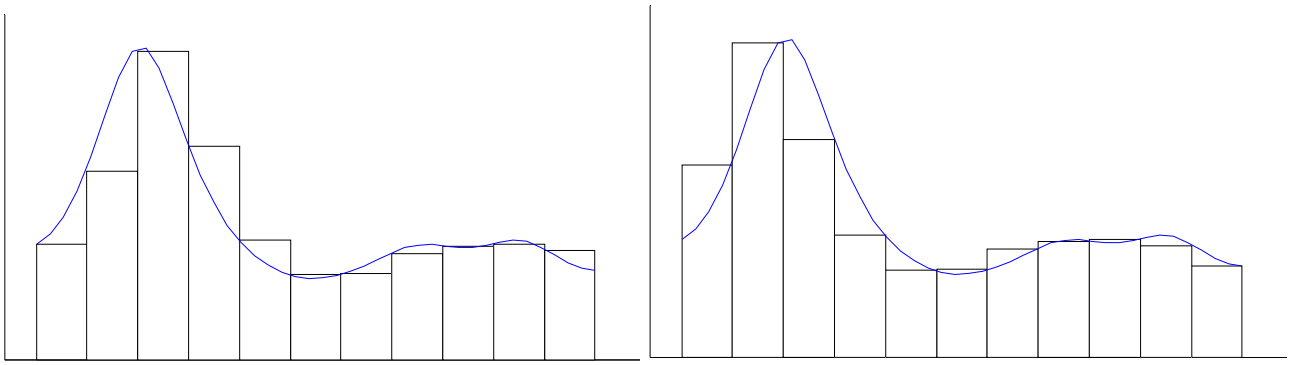


Figure 2.5: Graph of cross-sectional area function (blue) and histogram of the value for each tube (black)

Apart from averaging the data, other methods of determining the tube cross-sectional area may be used. The tube cross-sectional area can be taken straight from the collected data at the rising (Figure 2.5(a)) or falling edge of the tube Figure (2.5(b)). The three methods of finding the tube area were investigated by the author to determine whether the effect of the tube radii had a significant impact on the resulting analysis.



(a) (b)

Figure 2.6: Tube area obtained with rising edge (a), and tube area obtained with falling edge (b)

The three methods of finding the tube cross-sectional areas were compared by finding the associated formant values of the resulting geometry. It was found that the method of averaging yielded the closest results to the formants of actual recorded speech, and thus was chosen as the method implemented in this study.

2.6 The linear predictive coding method

The linear predictive coding (LPC) method is a powerful technique used in speech analysis. The LPC method is a time domain technique which models a signal as a combination of weighted past signals. For a signal 'y' this can be presented as

$$y[n] = a[1]y[n-1] + a[2]y[n-2] + \cdots + a[k]y[n-k] + \mathcal{E}[n]$$

With vowels being the speech signals of interest in this study, it is assumed that these signals are largely determined by the resonances of the vocal tract shape. Having made this assumption, the all-pole filter model of speech can be used to estimate these signals. The all-pole filter models the current signal as the sum of scaled past values and the input value, which can be expressed as

$$y[n] = -a[1]y[n-1] - a[2]y[n-2] - \cdots - a[k]y[n-k] + x[n]$$

It can be seen from this that for the speech signal 'y[n]', the coefficients 'a[k]' encodes the information about the vocal tract filter, and these coefficients can be estimated by the use of the LPC technique, as they both approximate the speech signal as the weighted sum of past values with an added error term (Harrington & Cassidy, 1999).

This method is useful for its ease of computation and the accuracy in estimating key speech parameters such as pitch, vocal tract area function, formants and spectra once the coefficients are obtained. It is used in this study to calculate the vocal tract resonance values and spectrum data from the

acoustic tube representation of the vocal tract. The steps involved in calculating the LPC values and deriving the spectrum are presented in Figure 2.4.

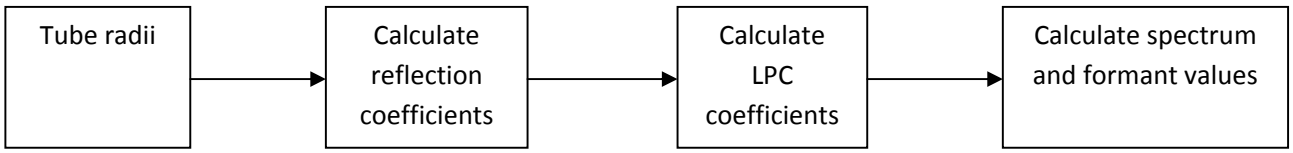


Figure 2.7: Flow diagram of the steps involved in the LPC method.

2.6.1 Calculating reflection coefficients

Reflection coefficients describe the behaviour of the environment with respect to a travelling wave. In this case, these coefficients describe the relationship of the sound being passed through a series of acoustic tubes and how much sound is reflected back towards the source. The reflection coefficient is described by Equation 1, where A is the area of the tube and n is the tube instance

$$r_n = \frac{A_{n+1} - A_n}{A_{n+1} + A_n}$$

2.6.2 Calculating LPC coefficients

LPC coefficients are the filter coefficients which describes the properties of the vocal tract. The LPC coefficients can be calculated from the reflection coefficients using a recursive autocorrelation technique. This has been displayed in equation x and the matrix below.

$$C(i,j) = C(i-1,j) - r_n * C(i-1, i-j)$$

While

$$1 \leq j < i$$

$$i \neq j$$

r_1	0	0	0	0	0	0	0	0	0	0
$C(2,1)$	r_2	0	0	0	0	0	0	0	0	0
$C(3,1)$	$C(3,2)$	r_3	0	0	0	0	0	0	0	0
$C(4,1)$	$C(4,2)$	$C(4,3)$	r_4	0	0	0	0	0	0	0
$C(5,1)$	$C(5,2)$	$C(5,3)$	$C(5,4)$	r_5	0	0	0	0	0	0
$C(6,1)$	$C(6,2)$	$C(6,3)$	$C(6,4)$	$C(6,5)$	r_6	0	0	0	0	0
$C(7,1)$	$C(7,2)$	$C(7,3)$	$C(7,4)$	$C(7,5)$	$C(7,6)$	r_7	0	0	0	0
$C(8,1)$	$C(8,2)$	$C(8,3)$	$C(8,4)$	$C(8,5)$	$C(8,6)$	$C(8,7)$	r_8	0	0	0
$C(9,1)$	$C(9,2)$	$C(9,3)$	$C(9,4)$	$C(9,5)$	$C(9,6)$	$C(9,7)$	$C(9,8)$	r_9	0	0
$C(10,1)$	$C(10,2)$	$C(10,3)$	$C(10,4)$	$C(10,5)$	$C(10,6)$	$C(10,7)$	$C(10,8)$	$C(10,9)$	r_{10}	0
$C(11,1)$	$C(11,2)$	$C(11,3)$	$C(11,4)$	$C(11,5)$	$C(11,6)$	$C(11,7)$	$C(11,8)$	$C(11,9)$	$C(11,10)$	r_{11}

The process starts with the first reflection coefficient. By using the equation described, and making a few assumptions and initial conditions, it is possible to calculate all the coefficient values $C(i,j)$ described in the matrix by iterating the coefficients row by row. Once the iteration has reached the number of the reflection coefficients, the LPC coefficients can be obtained by taking the values in the last row of the matrix (Eason, 2009).

2.6.3 Spectrum and formants

By applying the calculated LPC coefficients back into the all pole filter, it is possible to plot the spectrum of a specific vocal tract configuration. The filter spectrum is calculated the following equation.

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

The spectrum calculated from the data collected allows for a direct comparison between the magnitudes of the frequencies of different target vowels as well as a comparison between the two different scanning methods. The resonances, which can be located at the peaks within the spectrum, indicated the fundamental frequencies of the vocal tract structure for the given target vowel, and this is

used to compare the two different scanning methods. These resonances can also be compared to the formants extracted from a recording of a target vowel to verify the accuracy of the modelling method. This process is implemented in the form of a tool box, which will be discussed in the following chapter.

3 Study

3.1 Chapter overview

In this chapter, the background information on the participants is presented, followed by a description of the target vowels of interest. The processes of collecting the data using the acoustic reflectometer and MRI are described. For the MRI method, the process of extracting the cross-sectional area function from the images is outlined. Following this, the functionalities of the two vocal tract analysis tools used for this study are presented. The various features within the vocal tract tools are outlined. Finally, the repeatability of the two data acquisition techniques is discussed.

3.2 Participant background

All five participants are native speakers of New Zealand English. Their ages ranged from 25 to 45 at the time the data collection took place. For this study, the participants are labelled as SP01 up to SP05. The participant data is summarised in the table below.

	SP01	SP02	SP03	SP04	SP05
Age	25	23	45	45	25
Gender	Male	Male	Female	Male	Male
Height	188cm	171cm	165cm	182cm	190cm

Table 3.1: The age, gender and height of the 5 participants.

3.3 Vowels

For this study, the 11 monophthongs of New Zealand English were the primary vowels of interest. For acoustic reflectometery, data was collected on 9 of the vowels while for the MRI all 11 were collected. For the data collection process, the vowels were placed in an 'HvD' frame. The monophthongs are presented in Table 3.2 with their corresponding 'HvD' framed words.

Phonetic Symbol	æ	a	ɛ	i	ɜ	I	ɔ	ɒ	ʊ	ʌ	u
HvD	HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	HOOD	HUD	WHO'D

Table 3.2: Phonetic symbols of the 11 monophthongs of NZ English and their corresponding words in HvD frame.

3.4 Acoustic reflectometer method

The data used in this study was selected from a database of AR results of 26 speakers collected between 2008 and 2010, part of which were reported by C. I. Watson and Hui (2010). While subsets of the 26 speakers have been presented in various studies, the full analysis for all 26 speakers has not been completed. MRI data was also collected for five of the speakers within this data set. Consequently, these 5 speakers were chosen for the purpose of comparing the two different methods and determining the suitability of using the acoustic reflectometer as a time efficient and reliable substitute to the other techniques of obtaining cross-sectional area data. All the data for these 5 speakers were taken in 2008 apart from SP03, for which the data was collected in 2010.

To take AR measurements, the subjects were required to hold a mouth piece in their mouths while placing their articulators in place for a target vowel. For the duration of the scan, the participants were in a seated upright position. It was important for the subjects to seal their mouths firmly over the mouth piece to ensure the integrity of the sonic pulses being sent down the vocal tract. As the vocalisation of a target vowel would disrupt the sonic sound pulses sent by the acoustic reflectometer and distort the results, the subjects were required to hold their articulators for the target vowels without actually making

the vocalisations. This meant that the subjects were not able to use audio feedback to determine the accuracy of their articulation.

To reduce the impact of this issue the subjects were asked to vocalise the target vowel immediately before each measurement took place. Four instances of each target vowel was then measured and compared for consistency. Any erroneous data observed was removed and the measurement was repeated. The vocalisations made before the measurements were recorded, from which formant data was extracted and compared to the resonances deduced from the vocal tract shapes obtained from the two methods.

The AR data used in this study was collected using the ECCOVISION Acoustic reflectometer. This data was collected from 26 speakers of New Zealand language, from which 5 were selected. Each speaker was asked to vocalise nine of the eleven monophthongs of New Zealand. These were vocalised within a 'HvD' frames in the form of the nine words "HEED", "HEAD", "HAD", "HARD", "HOD", "HOARD", "WHO'D", "HERD" and "HID". The speakers were required to hold their vocal tract shape constant for the three second scanning period. This was repeated four times for each vowel from which an output file was produced after the measurement had been completed, which contained the distance along the vocal tract and its cross area function.

3.5 Magnetic Resonance Imaging

As part of a series of studies, MRI data was collected for all 11 of the New Zealand monophthongs for 12 speakers between 2010 and 2011. A subset of these speakers has been explored by C. I. Watson and Hui (2010), but like the AR, not all the speaker data have been explored. For the purposes of this study, 5 speakers were chosen, the data for which were collected in 2010. These speakers were chosen for their overlap into the AR results.

The MR images used in this study were obtained by the 1.5T Siemens Magnetom Avanto MRI scanner. Scans were performed to obtain images of parallel sagittal planes with 6 mm separations with 1 mm resolution. For SP01, the images were captured with a field of view of 211x260 mm, while SP02 was captured with 211x260 mm field of view. SP03 and SP04 were captured with 199x250 mm field of view, and SP05 was captured with 191x240 mm. 13 slices were obtained for each target vowel, each set requiring the speaker to vocalise the target vowel for 15 seconds. This is significantly longer than the three seconds required for the AR.

One of the draw backs of the MRI method is the fact that the imaging takes place when the speakers are in a lying position. As this would result in the gravity acting on the vocal tract to be in a different direction to when the speakers are upright, there is likely be a discrepancy between the vocal tract structures derived from the scans (lying position) opposed to the upright vocalisation captured by the AR.

For this study, the scanning takes place from the edge of the jaw on one side and ends on the other. The edges of the jaw are determined by a localizing scan at the start of the data collection process. Reducing the scanning to between the edges of the jaw as opposed to the edges of the ears considerably reduced the scanning time in comparison to other vocal tract studies. The speakers vocalised the target vowels within the HvD frame. For the MRI, data for the eleven monophthongs of New Zealand English were collected, which includes all the ones accounted for by the AR, plus "HUD" and "HOOD". The results for two vocalisations of each vowel were recorded from each of the 5 speakers.

The images from the scans were subsequently used to obtain structural information of the vocal tract. To analyse the data, the images were converted to bitmaps and read into the open source program CMGUI [http://www.cmiss.org/cmgui]. CMGUI was used to create a 3-dimensional texture block from the images taken by the MRI, allowing points within the vocal tract to be identified. This process is outlined in the following sections.

3.5.1 Identifying the vocal tract and key landmarks

In this study, the vocal tract is analysed in two segments: the oral and the pharyngeal cavities. The oral segment has been defined to be between the lips and the velar-pharyngeal port, while the pharyngeal segment is defined to be between the velar-pharyngeal port and the glottis. This can be seen in Figure 3.1, where the segment between the two red doted lines is the oral cavity and the rest of the segment marked with the green coloured nodes is the pharyngeal cavity. Note the bottom most node is placed upon the glottis. Analysis of these two segments is conducted independently.

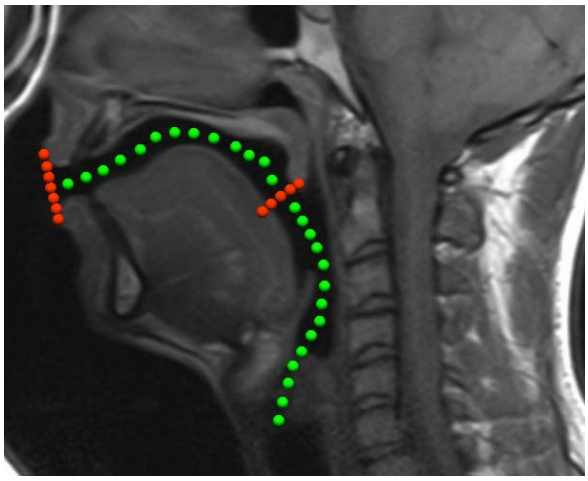


Figure 3.1: The green nodes mark the centre-line of the vocal tract while the red ones mark the lips and the pharyngeal port.

3.5.2 Marking up the midline of the vocal tract

In order to locate and identify the location and shape of the vocal tract, nodes are placed on the centreline of the tract. The mid-sagittal plane of the MRI images is loaded and nodes are placed over the image. For the oral segment, 15 nodes are placed by hand, as evenly spaced as possible, from the outer edge of the lips to the velar-pharyngeal port. This can be seen in Figure 3.2. A similar process is carried out with the pharyngeal segment, with the starting node at the velar-pharyngeal port and the ending node at the glottis. These nodes are referred to as the ‘snake’ nodes. Note the number of snake nodes can be arbitrary but for the purposes of this study fifteen nodes each in the oral and pharyngeal sections gave sufficient detail and was not too time consuming.

Once the snake nodes are placed, a line is fitted to these nodes, and is identified as the centre-line of the vocal tract. This spine is then re-sampled so there are 15 equi-distance nodes along it. At each of these nodes ‘area planes’ are created. On each plane the cross section of the vocal tract is identifiable, enabling the region to marked out and the area of the cross-section calculated.

3.5.3 Creating the area plane and marking the edges of the vocal tract.

At each of the 15 nodes placed on the centre-line, a plane is defined. Each of these planes is normal to the centre-line, and the data within the plane is applied from the texture block. Once the planes are created, the boundaries of the vocal tract can be seen, and subsequently marked. The marking is carried out by placing points along the boundary of the vocal tract as seen on the planes, as shown in Figure 3.3.

Once all the planes have been marked, the area is calculated. Care is taken to account for teeth, which do not appear on MRI images. The CMGUI software allows the vocal tract, and planes to be rotated and zoom into. This also greatly aids the accuracy of the mark-up – being able to compare it to other anatomical landmarks, not necessarily clear on the plane currently being marked up.

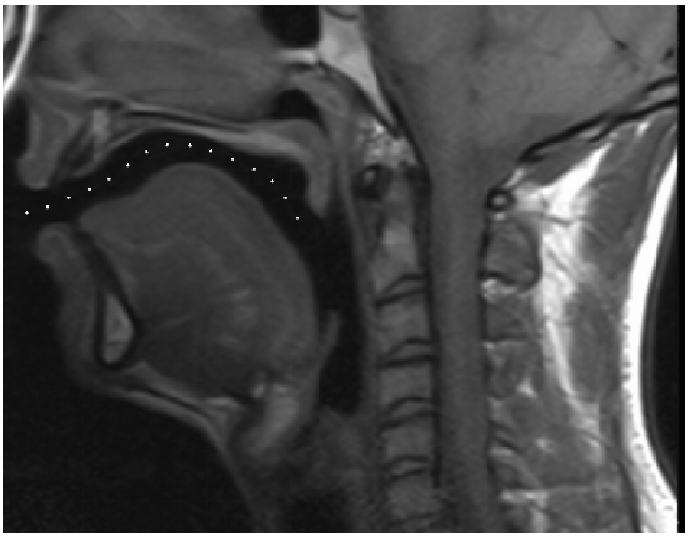


Figure 3.2: Mid sagittal MRI image with snake nodes placed in the oral cavity.

3.5.4 Calculation the vocal tract area

Once the boundary of the vocal tract has been identified on each of the planes, a predetermined pearl script is executed to find the areas. A process of triangulation is used to calculate the area in each cross-section and the results are stored in a text file, along with the distance measure between each cross-section, for later analysis (C.I. Watson et al., 2009).

As mentioned in a previously, the oral cavity and the pharyngeal cavity were analysed separately. This means that there are two sets of results for the area, one for each cavity. As the starting plane of the pharyngeal cavity should coincide with the ending plane of the oral cavity, an average is taken from the two areas which were obtained from the same plane. This now allows the two separate data sets to be combined into a complete area representation of all the planes ranging between the lips and the glottis.

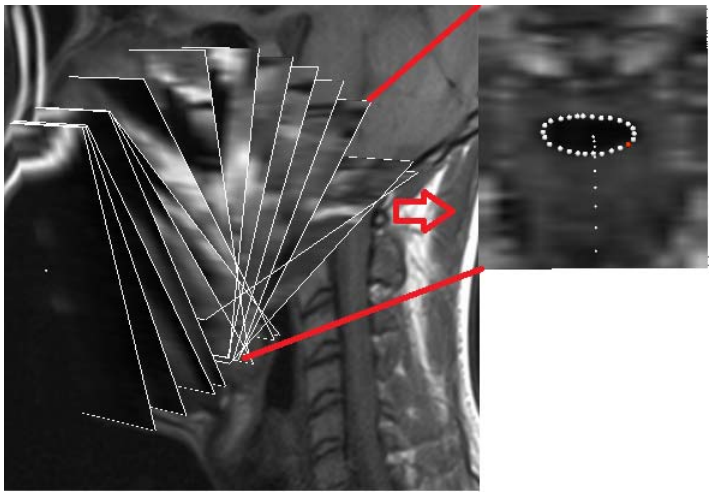


Figure 3.3: Planes are placed at each of the snake nodes and the border of the vocal tract is marked on each node

3.6 Vocal tract tools

In order to efficiently process the data obtained from the AR and MRI method, a software package is required. The function of the software package is to perform the LPC analysis described in section 2.6 of chapter 2 and produce the corresponding resonances for the vocal tract shapes. For this study, two packages were used: the Vocal Tract Tool Mark I and Vocal Tract Tool Mark II. The Vocal Tract Tool Mark I was originally developed for the analysis of AR results only (Eason, 2009). To include the MRI results, the author developed the Vocal Tract Tool Mark II in MATLAB. The two packages are outlined in the following section

3.6.1 Vocal Tract Tool Mark I

The Vocal Tract Tool Mark I (hence forth referred to as VTTMI) was developed in the open source statistical computing software R [www.r-project.org/]. It employs the LPC theory described in the previous sections to calculate the resonances of a target vowel by transforming the cross-sectional area functions obtained from the acoustic reflectometer. It was designed to read in the data and provide a visual representation present the formant results. The tool interface is presented Figure 3.4

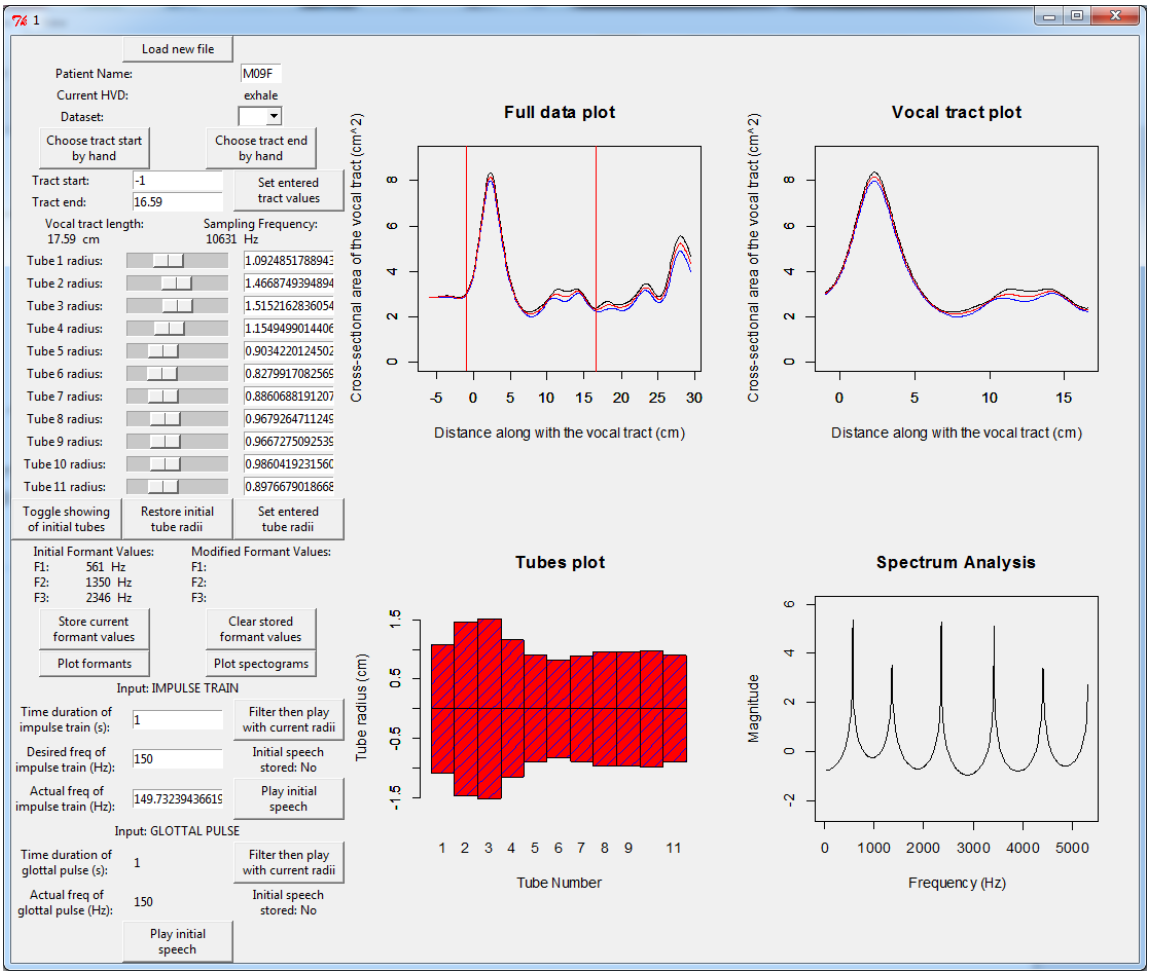


Figure 3.4: Graphical User Interface of the Vocal Tract Tool Mark I

Graphs

This tool box was designed for the ease comparison between the different sets of data collected for the different vowel vocalisation. The primary function of the package is to plot various aspects of the collected data so that a visual comparison can be made between different data sets. For the AR data, four vocalisations of the same vowel is performed by each speaker, and recorded in the data file. This is loaded into the package and the results from the different vocalisations can be viewed by selecting the instance in the drop down menu "dataset".

For each vocalisation, four graphs are presented. On the "Full data plot" graph, the cross-sectional area data obtained from the measurement is plotted against the distance along the vocal tract. This

represents the AR data in its entirety. Much of this data, however, is excess data as the region of interest is only between the lips and the glottis. Therefore, using a predetermined script to locate the position of the lips and glottis within the dataset, a truncated version of the same plot is made under "Vocal tract plot".

The third graph "Tube plot" is the representation of the data in the form of 11 equal width lossless acoustic tubes. This is obtained by dividing the raw data within the area of interest into 11 equal length portions, from which the average cross-sectionals area found. These cross-sectional areas are subsequently converted into radii and plotted on the graph.

For the "Spectrum Analysis" graph, a pre determined script uses the radii of the acoustic tubes to calculate the spectrum of the dataset of interest. This allows the user to quickly see the location of the resonances for a given vowel under inspection. The values for the first three peaks on the graph are displayed under "Initial formant values" in hertz on the user interface.

Dynamic update of tube radii

One of the highly useful functions in the VTTMI is its ability to dynamically update any of the tube radii and reflect the results instantly. To change the radii value from the default, new radii values can be entered directly into text box next to the sliders, or by the adjustment of the sliders. The updated results are displayed with the original, allowing the user to see the impact of the change on the spectrum and resonances. In Figure 3.5, a change in the spectral plot can be seen. The red plot represents the spectrum of the updated tube values.

On "Tubes plot" in Figure 3.9 it can be seen that the sixth and seventh tubes have been altered. The sixth tube radius has been decreased while the seventh has been increased. The “Spectrum Analysis" graph now contains a revised red line, which reflects the changes in the spectrum after the tube radii have been changed. The revised values for the first three resonances are displayed under "Modified formant values", making it easy to see how much impact the altered radii have on the resonances.

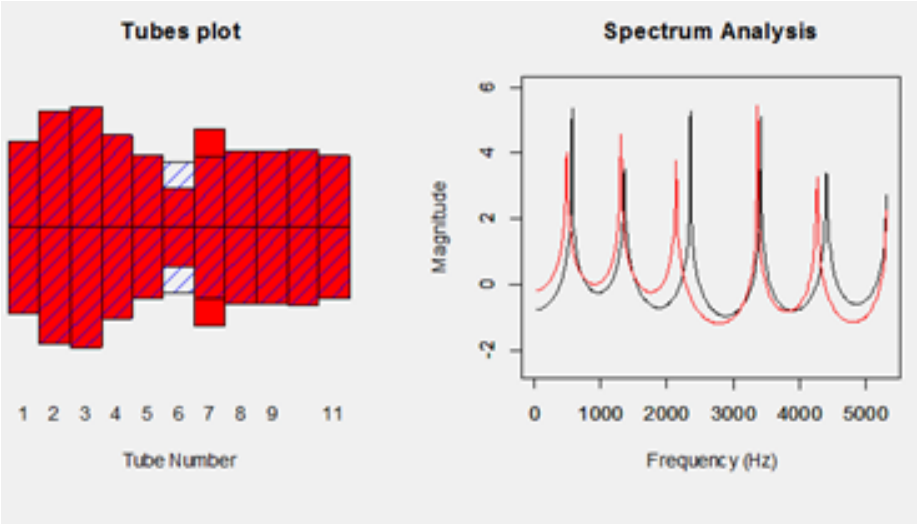


Figure 3.5: Graphs updating to show the changes made to the tube radii and its effect on the spectrum

3.6.2 Vocal Tract Tool Mark II

As VTTMI was developed specifically for the analysis of AR data, it was not able to process the data extracted from the MR images. Therefore the Vocal Tract Tool Mark II was developed (hence forth referred to as VTTMII). This tool box was developed by the author in MATLAB and can load both the AR and MRI data. The interface is shown in the Figure 3.6.

For VTTMII, the functionalities remain the same as the previous version. The cross-sectional areas along the vocal tract is loaded from the data files, and plotted against its distance along the vocal tract. In VTTMII however, all the vowels from one speaker is loaded at once by selecting the appropriate directory, and can be viewed by selecting the vowel in a drop down menu. Similarly, the graphs presented in VTTMII reflect the same functions of their counterparts in the previous version, and the first three resonances are also displayed. However, due to the limitation of time, the dynamic update feature seen in VTTMI was not duplicated in its MATLAB counterpart.

3.6.3 Data collection

For the purposes of this study, the resonance values and the length of the vocal tract for each target vowel are needed. These values need to be collected manually and recorded in a excel spread sheet as data saving is not an automated feature in this package. These results have also been presented in Appendix A.

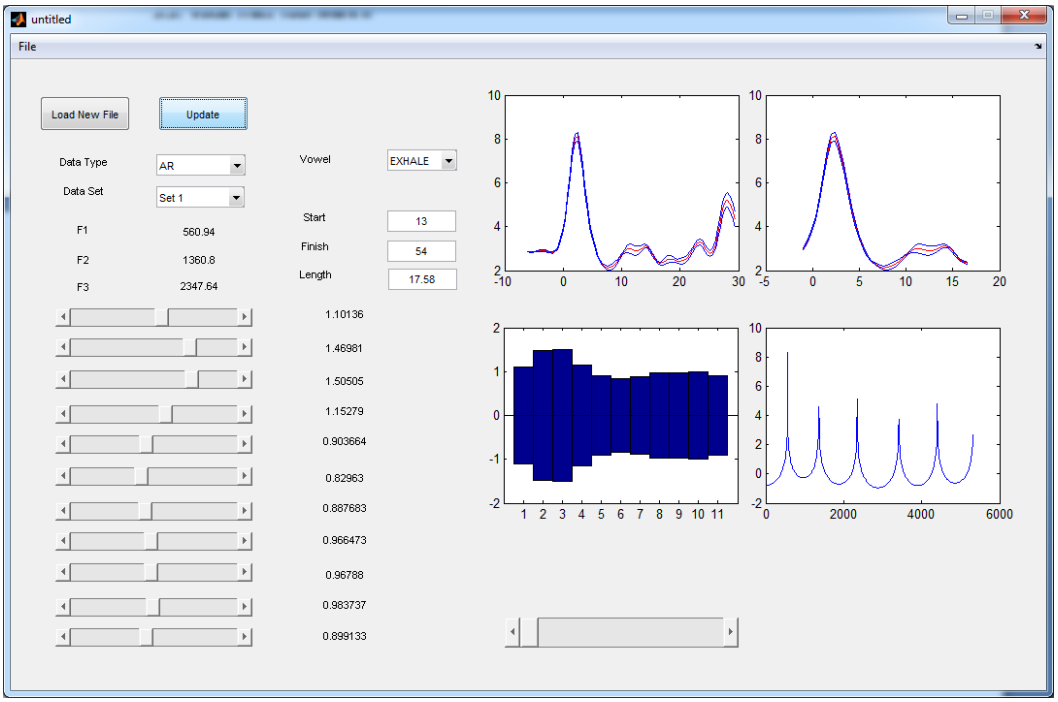


Figure 3.6: Graphical User Interface of the Vocal Tract Tool Mark II

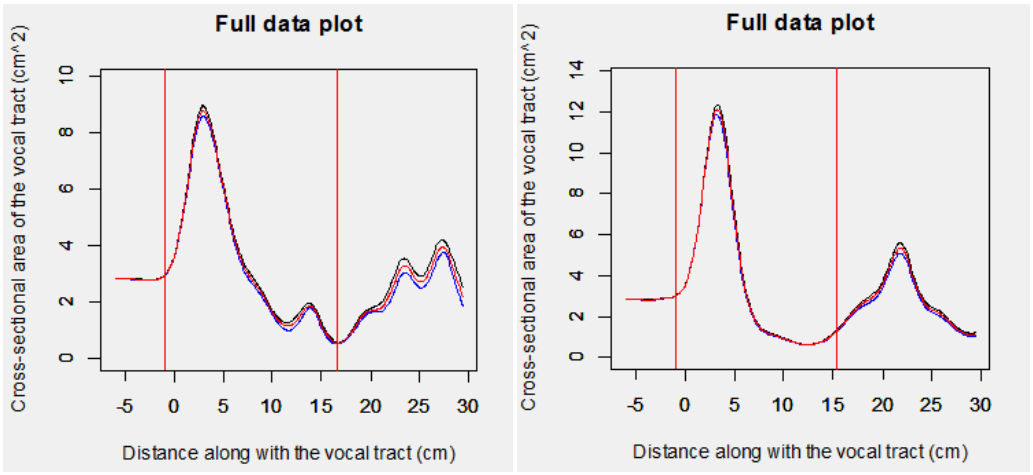
3.7 Issues encountered in data extraction.

For both the AR and the MRI techniques, difficulties were presented within different areas of the data extraction process. Processes such as identifying the glottis on the plot of the AR data and the accurate identification of air-tissue boundaries on the MR images are important in determining the resulting formant values. Therefore, for accurate resonance values results to be obtained, it is essential to employ an accurate process. This section discusses the instances when questionable data is encountered and how to ensure the data is applied to the process accurately.

3.7.1 AR - Indentifying the glottis

As the acoustic reflectometer collects data beyond the distance of the glottis, it is important for the analysis tool to identify its location before further analysis can be made and because of this, the Vocal Tract Tool Mark I (refer to section 3.6.1 for more detail) was designed to automatically detect the location of the glottis. However, as the dimensions of the vocal tract vary between individuals, the glottis cannot always be accurately located.

In Figure 3.7(a) the vocal tract shape of the vowel 'HARD' has been presented. In the graph, the first red line indicates the position of the lips and the second indicates the position of the glottis. In this example, the glottis location looks to be reasonable and yields appropriate formant values. However, in Figure 3.7(b), it can be seen that the glottis position identified by the vocal tract tool is not at an expected location.



(a) (b)

Figure 3.7: Cross-sectional area profile along the vocal tract for the vowel 'HARD' from speakers SP04 (a) and SP02 (b)

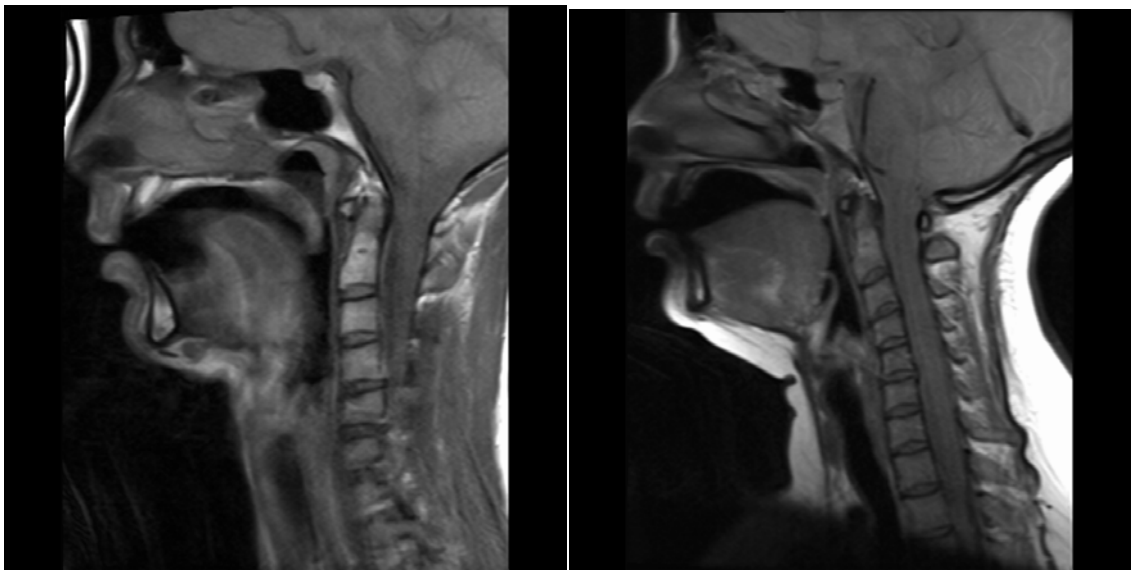
The issue which revolves around visually identifying the glottal position is the lack of consistency between the individuals in terms of the results collected by the acoustic reflectometer. As it is observed in Figure 3.7(a) the glottal position can be identified at the lowest point on the graph, which is expected, as the participants were required to close their glottis during the measurements. However, this is not always the case, as some participants were not able to successfully close their glottises during the measurements. This subsequently affects the vocal tract's ability to correctly identify the glottis.

In Figure 3.7(b) it can be seen that the lowest point on the graph, which is where the glottis should be, is situated around 12.5 cm. It is important to note for the AR method, the measurement does not take the distance from the front teeth to the lips as part of the vocal tract. Instead, because of the mouth piece, the measurement of the tract length starts at the teeth. To compensate for this, the start of the vocal tract is defined to be at -1cm on the graph, making the vocal tract in Figure 3.4(b) 13.5 cm. As this is much too short for a male subject, it is obvious that this is not where the glottis is positioned. In such cases, the position of the glottis needs to be hand selected. By referring to the midsagittal images from the MRI

method and the shape of vocal tract plots from other speakers, it was possible for the author to select the most sensible glottal points on the plot.

3.7.2 MRI - Marking of the vocal tract air-tissue boundary

In the previous section, the process of marking the MR images to extract the cross-sectional areas using CMGUI was described. Due to the nature of the MRI, the air-tissue boundary can be compromised by factors such as movement during the scanning. As this is inevitable and the degree of compromise varies from participant to participant, it can be potentially difficult to accurately mark the boundary. An example of a compromised air-tissue boundary can be seen in Figure 3.5(a), where no clear boundary can be seen between the vocal tract and the surrounding tissue. This is contrasted to Figure 3.5(b), where the boundary can be clearly identified.



(a) (b)

Figure 3.8: Midsagittal MR image of the vocal tract during the vocalisation of the hood vowel for SP05 (3.5(a)) and SP03 (3.5(b))

In Figure 3.5(a), instead of a defined boundary, a gradient ranging from light grey to black is present around the edges of the tongue. This was likely caused by the moving of the tongue during the scanning process. However it is still possible to distinguish where the boundary is by closely observing the image.

This becomes more difficult once the images have been processed. While it is possible to differentiate between the tissue and the gradient caused by movement in the midsagittal image, this

becomes increasingly difficult due to the extrapolation process performed by CMGUI. In Figure 3.9, is can be seen that the majority of the cavity is covered by a grey shade, thus making it difficult to determine it is part of the tract shape.

In order to correctly mark the vocal tract in this case, it is necessary to refer back to the midsagittal image to determine which gives a much more clear indication. Though it is not a fail proof method, it is usually sufficient to determine which regions of the images belongs to vocal tract to allow for the mark up process. How accurate this mark up is depends on an individual's interpretation of the images and his understanding of the vocal tract. Figure 3.9(b) shows the region marked as the vocal tract after cross-referencing the midsagittal image.

(a) (b)

Figure 3.9: Unmarked (a) and marked (b) slice of the hood vowel for SP05

Another issue encountered during the marking of the vocal tract is the determination of the location of the teeth. As mentioned previously, the teeth do not appear in the images taken by the MRI machine, and would appear to be black space much like the vocal tract. This has to be taken into account appropriately during the marking process so that the region where the teeth are located is not mistaken as part of the acoustic chamber.

In Figure 3.10, the area of the vocal tract can be located in the centre of the image as the circular black area. Two small circular patches are present on either side of the centre circle. These small patches represent the teeth, and are not part of the acoustic chamber. Figure 3.7(b) shows the region identified as the vocal tract. Once again, the accuracy of the marking depends on an individual's understanding of the geometric features within the vocal tract and his interpretation of the images. It is likely for the results obtained from the images marked by different individuals to yield slightly different results.

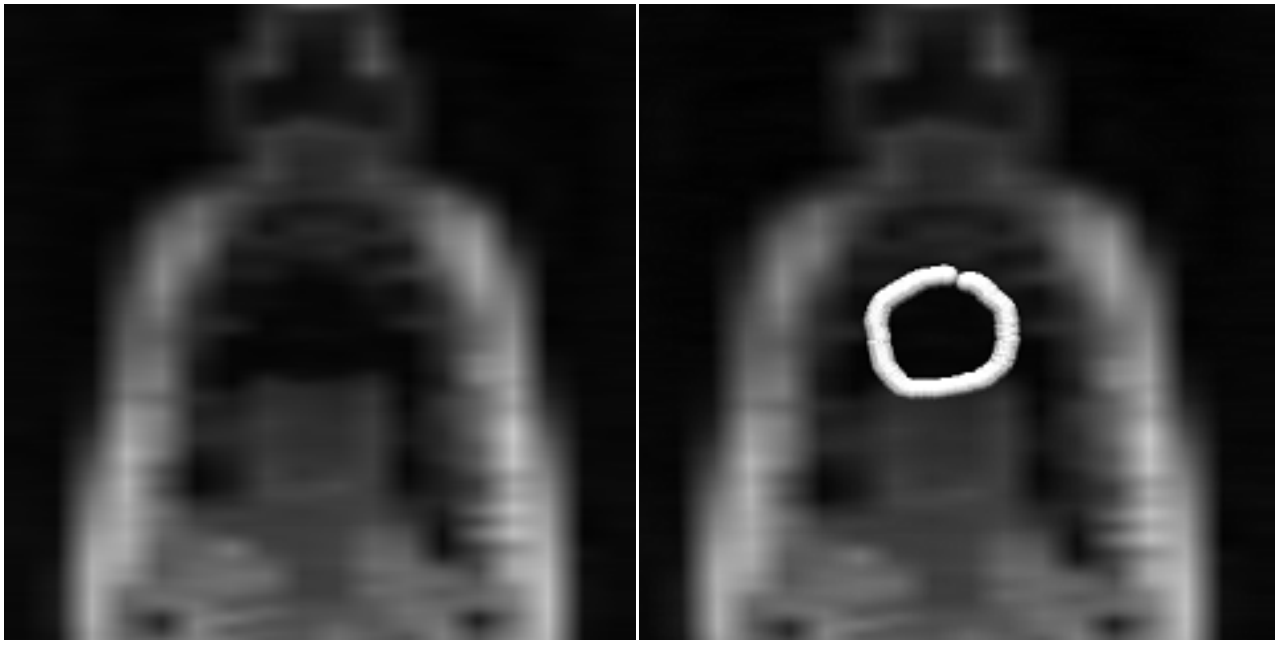


Figure 3.10: Unmarked (a) and marked (b) slice of the hood vowel for SP03

3.8 Repeatability of the experiments

3.8.1 AR

For the AR data, it can be seen from Figure 3.11 that the 4 different vocalisations of the same target vowel is generally relatively similar to each other. This was due to the fact that the output area function could be observed during the measurement process. This meant that the process could be repeated to remove any unreliable data. This allows the data obtained from the AR method to be reliable across different measurement.

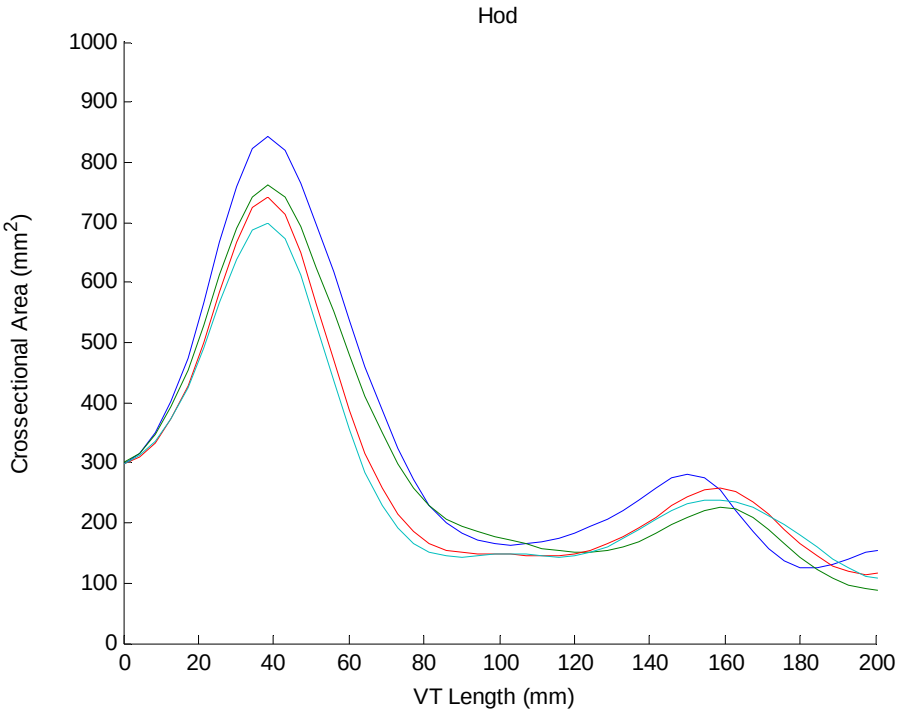


Figure 3.11: Cross-sectional area vs. vocal tract length plot of the HOD vowel for all 4 data sets

3.8.2 MRI

Two aspects of repeatability were investigated in this study for the MRI technique. First of all, repeatability between different vocalisations of the same target vowel is analysed. Two instances of each target vowel were measured and the results were compared against each other. A similarity in the resonance values for the two vocalisations indicates that the 3D structures reconstructed from the two instances of the MRI scans were structurally similar. This suggested that the scanning technique was reliable and the results were consistent across different scans of the same vowel (This will be further illustrated in the results section).

The second aspect of repeatability that was tested was the reliability of the image processing. The quality of the MR images depended on factors such as the participant's ability to accurately articulate the target vowels and how much they moved during the scanning process. The quality affected the process of determining the boundaries of the vocal tract, where the higher quality images had a define boundaries which were easy to label and lower qualities did not.

To test the repeatability of the boundary marking process, the data set for the first vocalisation of each vowel was processed by CMGUI twice. This yielded two sets of results which were compared to their similarity. This is illustrated in Figure 3.12, where two separately marked instances of 'HEED' of SP05's first dataset is shown. As it can be seen in the figure, the two instances of the cross-sectional area mark up resulted in very similar area functions. It was found that multiple markings of the same image set gave similar results, confirming the method's repeatability.

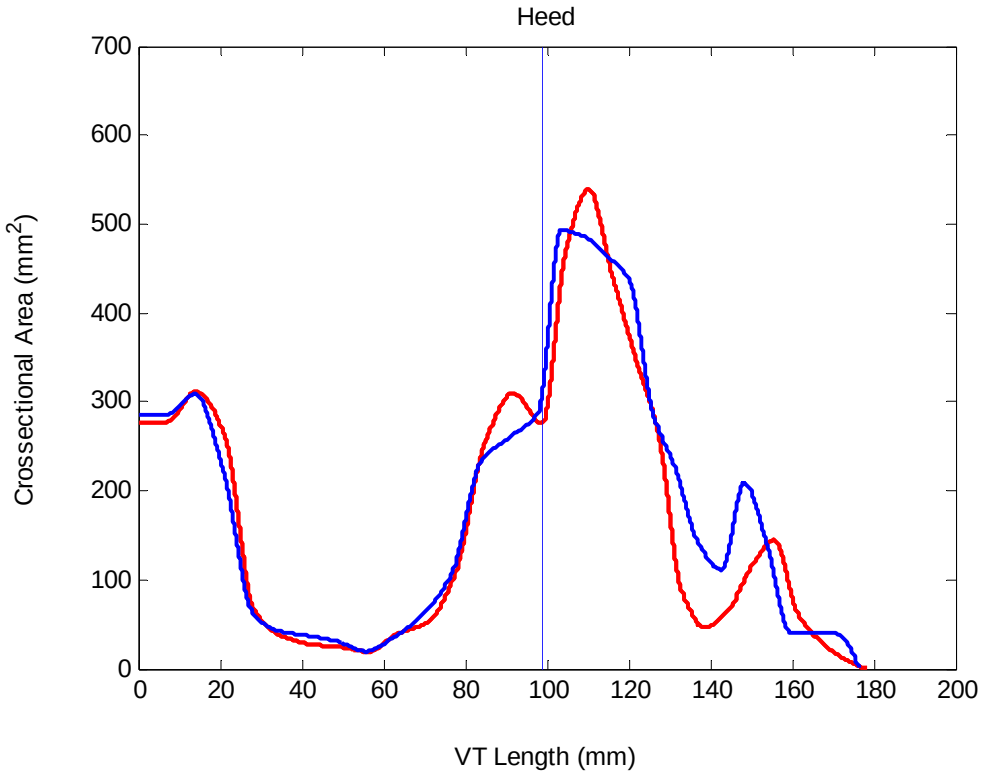


Figure 3.12: Two separately marked instances of the first dataset for SP05 (HEED vowel)

4 Results - Acoustic Reflectometery

4.1 Chapter overview

In this chapter, the results obtained from the AR method will be presented. In the first sections of the chapter, the vocal tract lengths of each vowel are presented and the cross-sectional area of each vowel is plotted and placed into a vowel space arrangement for the 5 speakers. This is accompanied by observational comments. This is followed by the presentation of the formant results obtained from the analysis of the cross-sectional area functions in graphical form. Comments were also made on these results

4.2 Vocal tract length

The vocal tract length for each vowel was extracted from the AR data for each of 4 data sets. The average was taken for each vowel for the 4 datasets. The vocal tract lengths vary from vowel to vowel and the range of these lengths is presented in the table below for each speaker. The mean of the tract length across the vowels are also presented. The vocal tract length for each vowel for all speakers can be found in Appendix A.

	SP01	SP02	SP03	SP04	SP05
Range of lengths (cm)	18.3-22.8	19.3-20.9	17.4-19.6	17.3-21.0	18.3-20.7
Mean length (cm)	20.2	20.3	18.3	19.3	19.6

Table 4.1: The range and average lengths presented in the results of the different vowels of interest for the AR method

4.3 Vocal tract shape

The cross-sectional area results collected using the acoustic reflectometery method is plotted into graphs and presented in this section. In Figures 4.1 the vocal tract cross-sectional area function for each vowel

collected from SP01 is plotted against its distance from the lips. Note that all four datasets collected for each vowel is plotted on the same axis. The cross-sectional area function for each vowel is plotted in its approximate positions within the phonetic vowel space. The data for all 5 speakers were presented in Figures 4.1-4.5.

Comparison of the vocal tract shape across the four datasets

For the acoustic reflectometery method, four sets of data were collected for each target vowel for each speaker. The first thing that can be seen from Figures 4.1-4.5 is the similarity presented in the vocal tract shape across the four datasets for each target vowel. While this is encouraging and would suggest a certain level of reliability in the measurement method, it is important to note that the data was visually assessed for quality during the measurement process, and results which were very different from the expected shape were discarded.

Vocal tract shape: Results vs. Expected geometry

Looking at Figure 4.1, it can be seen that the locations of constrictions and the openings are what would be expected from each vowel. For HEED, being a high front vowel, it is expected for the tongue to be close to the roof of the oral cavity while the tongue is situated towards the front of the vocal tract. This would mean that the oral cavity is expected to be small and the pharyngeal cavity is expected to be large. Looking at the figure, it is shown that the cross-sectional area towards the front of the vocal tract for 'HEED' is indeed much smaller than the second half of the tract, which reflects the expected geometry.

Following the expected change in geometry within the phonetic vowel space, it is expected for vowels further back within the vowel space to have an increasingly large oral cavity due to the tongue shifting towards the back of the mouth. This effect was observed in Figure 4.1 where the 'WHOD' vowel, which is also a high vowel but further back on the vowel space, is seen with a larger cross-sectional area within the front parts of the vocal tract compared to 'HEED'. 'HOARD', which was a high back vowel, had an even larger oral cavity presented on the plot. This indicated that the change in geometry from a high front vowel to a high back vowel presented in the data is concurrent to what would be expected.

Looking at the effect of moving from a high vowel to a low vowel, it is expected for high vowels to have smaller oral cavities compared to a low vowel. Such an effect can be observed through the vocal tract shapes presented in the 'HEED', 'HEAD', 'HAD' and 'HARD' plots. Starting from 'HEED' (a high vowel), the

vowels transition towards a lower setting until' HARD' (a low vowel). This means that the oral cavity in these vowels are expected to grow larger and the pharyngeal cavities smaller as the vowels progress from 'HEED' to 'HARD'. This was indeed observed in Figure 4.1, where the oral cavity of the vowels become larger and the pharyngeal cavity becomes smaller incrementally as the vowels progress from 'HEED', 'HEAD', 'HAD', through to 'HARD'. This effect was also observed for 'HOARD' and 'HOD'.

From Figures 4.1-4.5, it can be observed that the vocal tract geometry collected with the AR method follows the expected behaviours of the vocal tract for different vowels within the vowel space. The relationship between the shapes of the various vowels observed and described in Figure 4.1 was true for all the speakers.

Effect of the mouth piece

As the AR method requires the use of a mouth piece, certain aspects of the vocal tract structure was compromised during the measurement process. First of all, it was required for the participants to tightly seal their lips around the mouth piece, meaning that the opening at the lips was consistent for all the different vowels. This can be observed from the plots, where the starting values for all the plots are very similar (approximately 3 cm^2). This is significant, as it limits the jaw position of the participant, which can in turn affect the resonance values. This is demonstrated in the resonance section (see section 4.4).

It was observed from the plots that certain vowels had very similar vocal tract shapes. The plots for the 'HEAD' and 'HERD vowels can be compared to see that the vocal tract shapes for these two vowels are very similar within an individual speaker. Looking at Figure 4.1, it can be seen that the shape of the area function for these vowels are visually comparable, with similar features (the peaks and troughs) presented in approximately the same positions. This similarity was also observed for 'HOARD' and 'HARD'.

With HEAD and HAD, the similarity was likely to have been caused by the unnatural state in which the mouth piece puts the participant's vocal tract, which may have also affected the positioning of other articulators and compromise the structure of the tract. The tension with which the participants were required to hold the mouth piece in place is likely to limit the shape of the tract within the oral cavity, and may also affect the placement of the tongue. For HOARD and HARD, the difference in vocal tract configuration is reasonably small, with the main difference being the open jaw setting of HARD and the closed jaw setting of HOARD. In this case, the mouth piece limits the jaw opening of both vowels, and subsequently makes the difference between the two vocal tract shapes even less distinguishable.

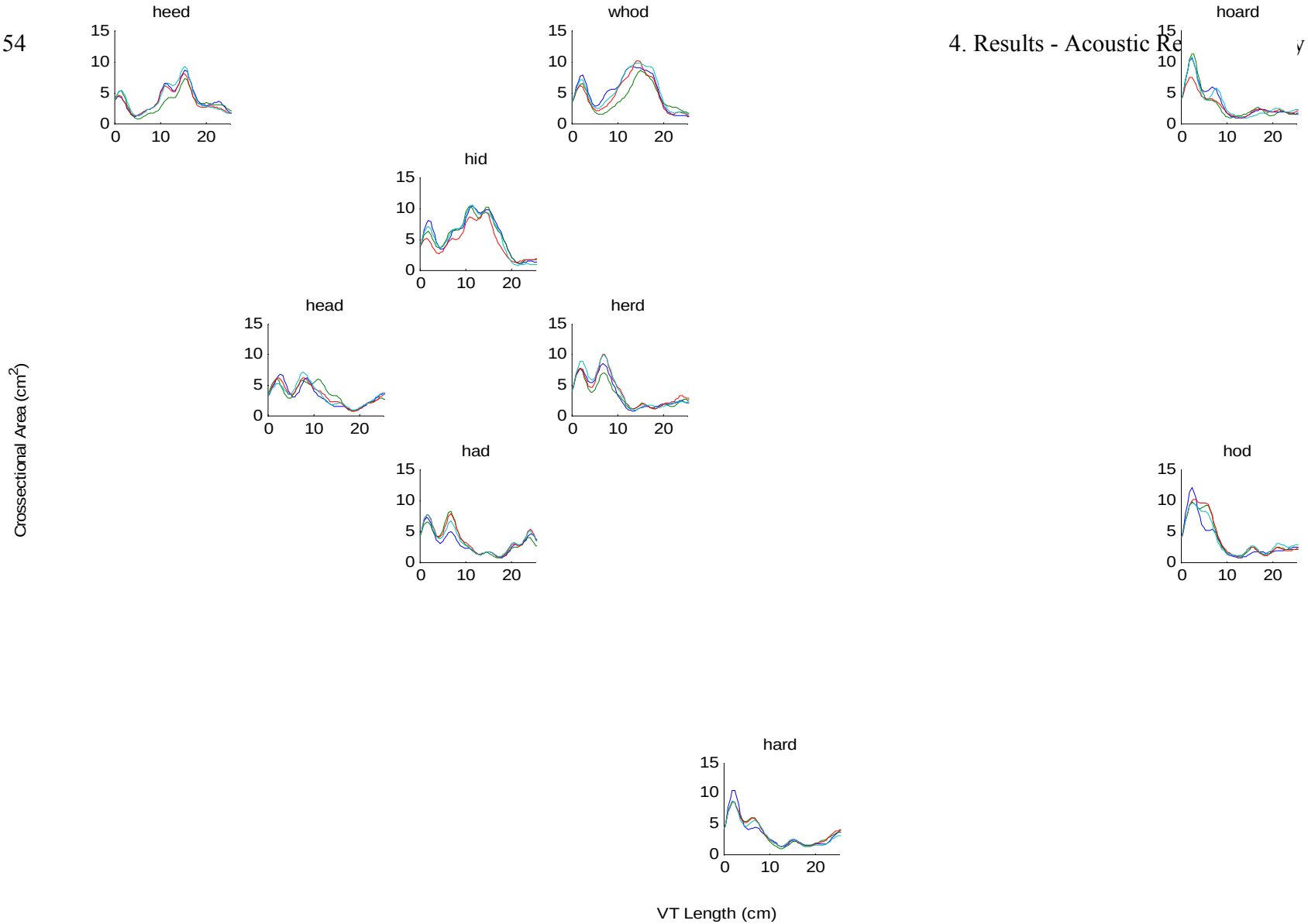


Figure 4.1: Cross-sectional area vs. vocal tract length for the nine monophthongs collected using acoustic reflectometery from SP01

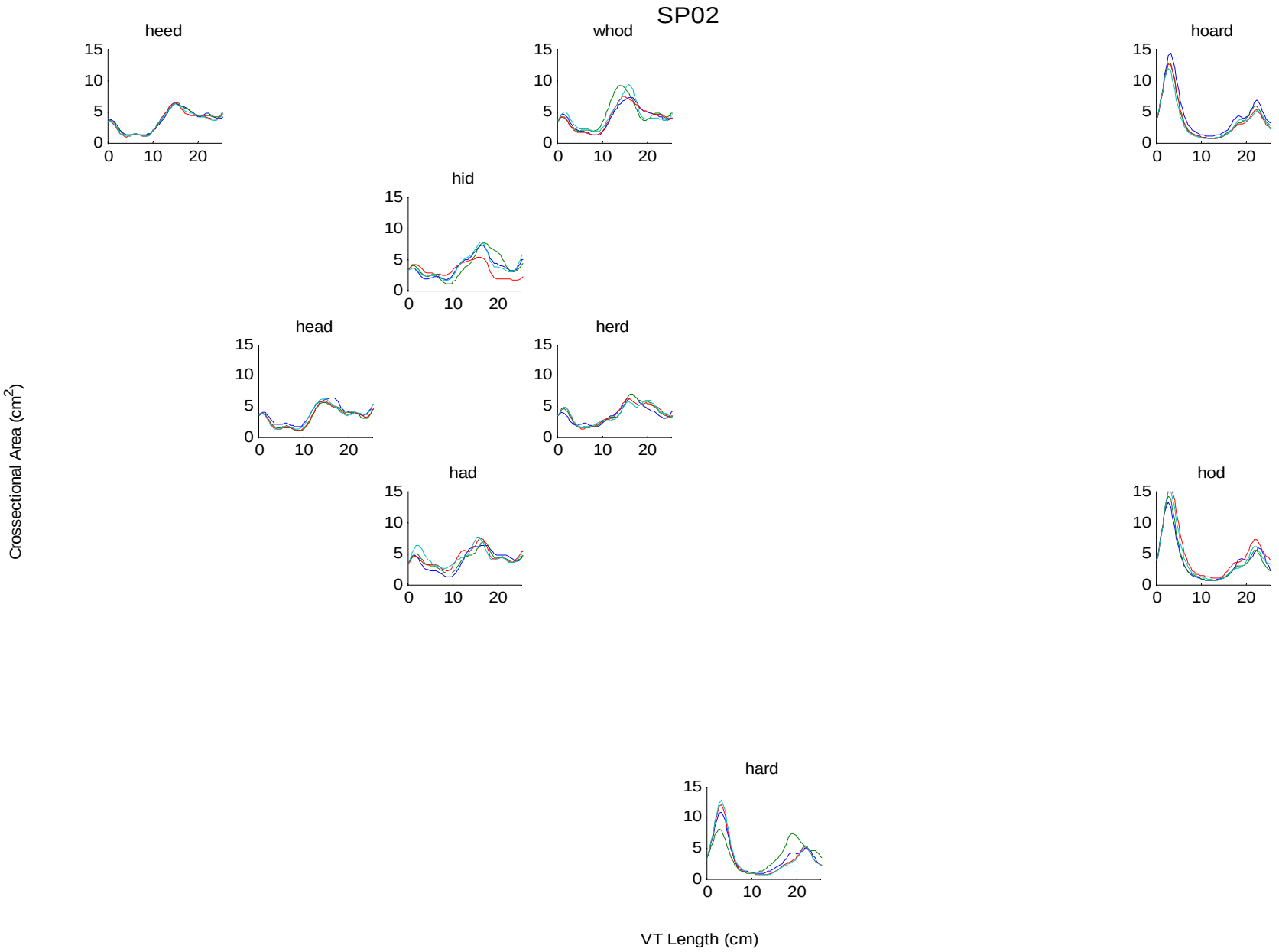


Figure 4.2: Cross-sectional area vs. vocal tract length for the nine monophthongs collected using acoustic reflectometery from SP02

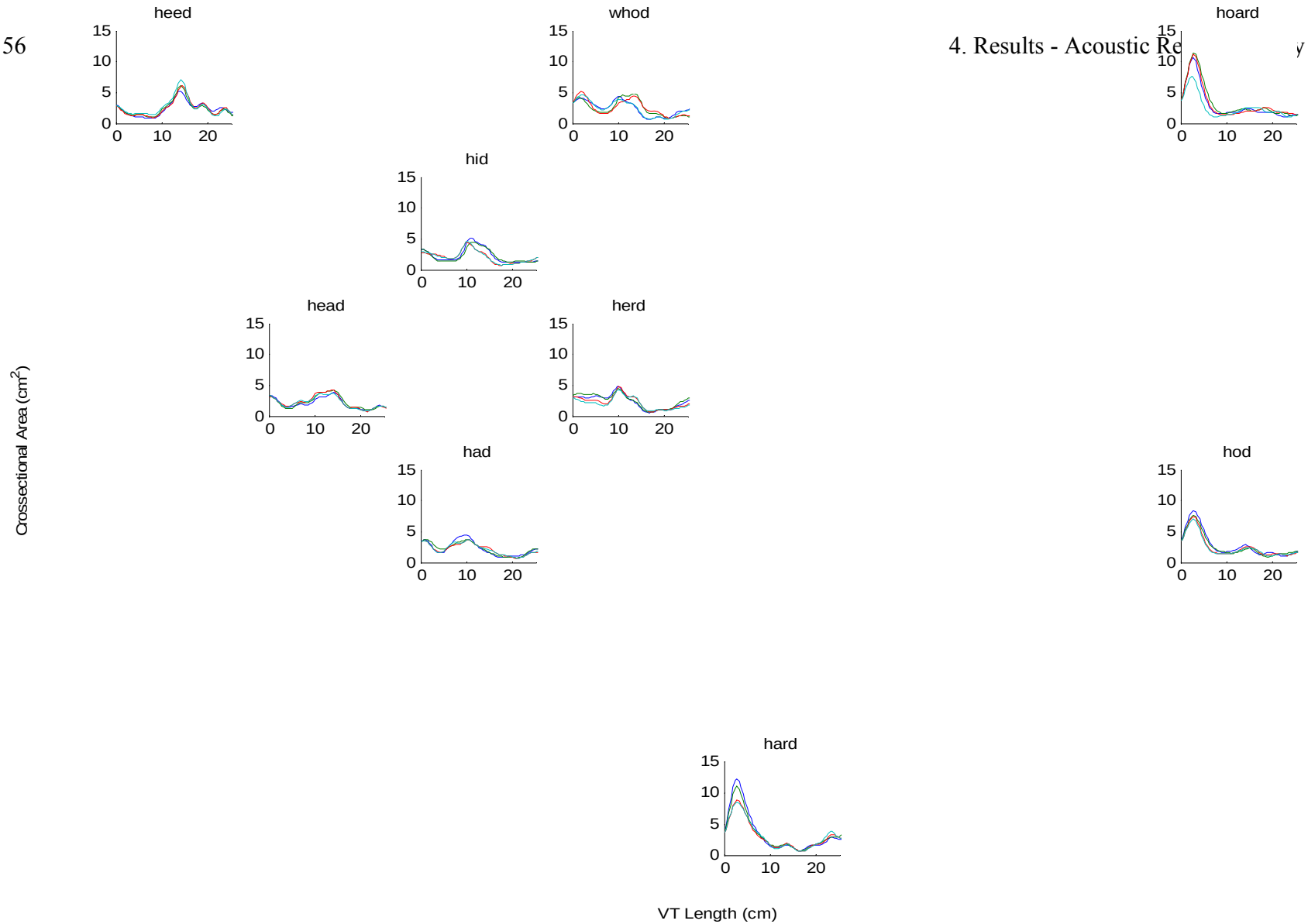


Figure 4.3: Cross-sectional area vs. vocal tract length for the nine monophthongs collected using acoustic reflectometery from SP03

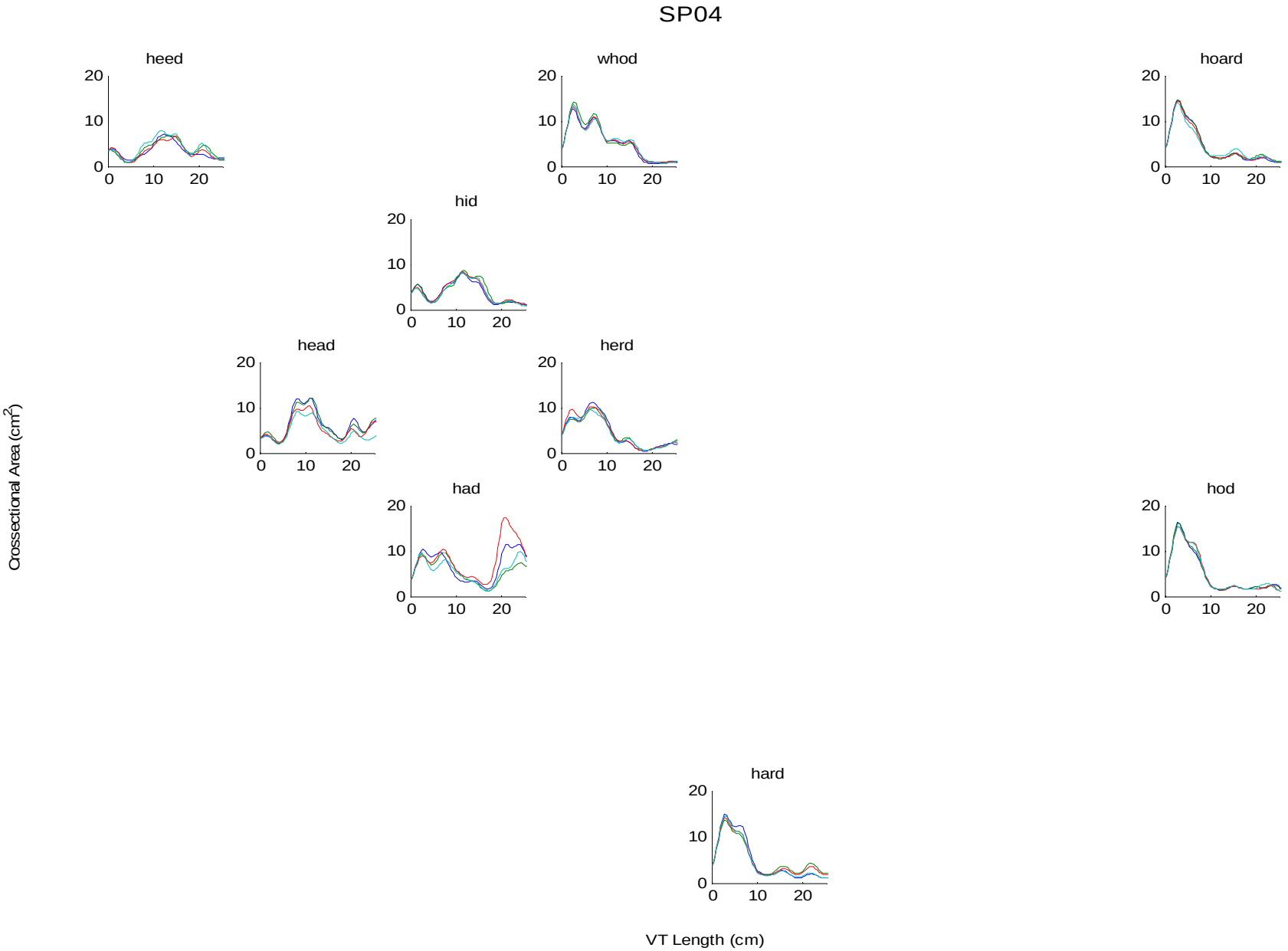


Figure 4.4: Cross-sectional area vs. vocal tract length for the nine monophthongs collected using acoustic reflectometery from SP04

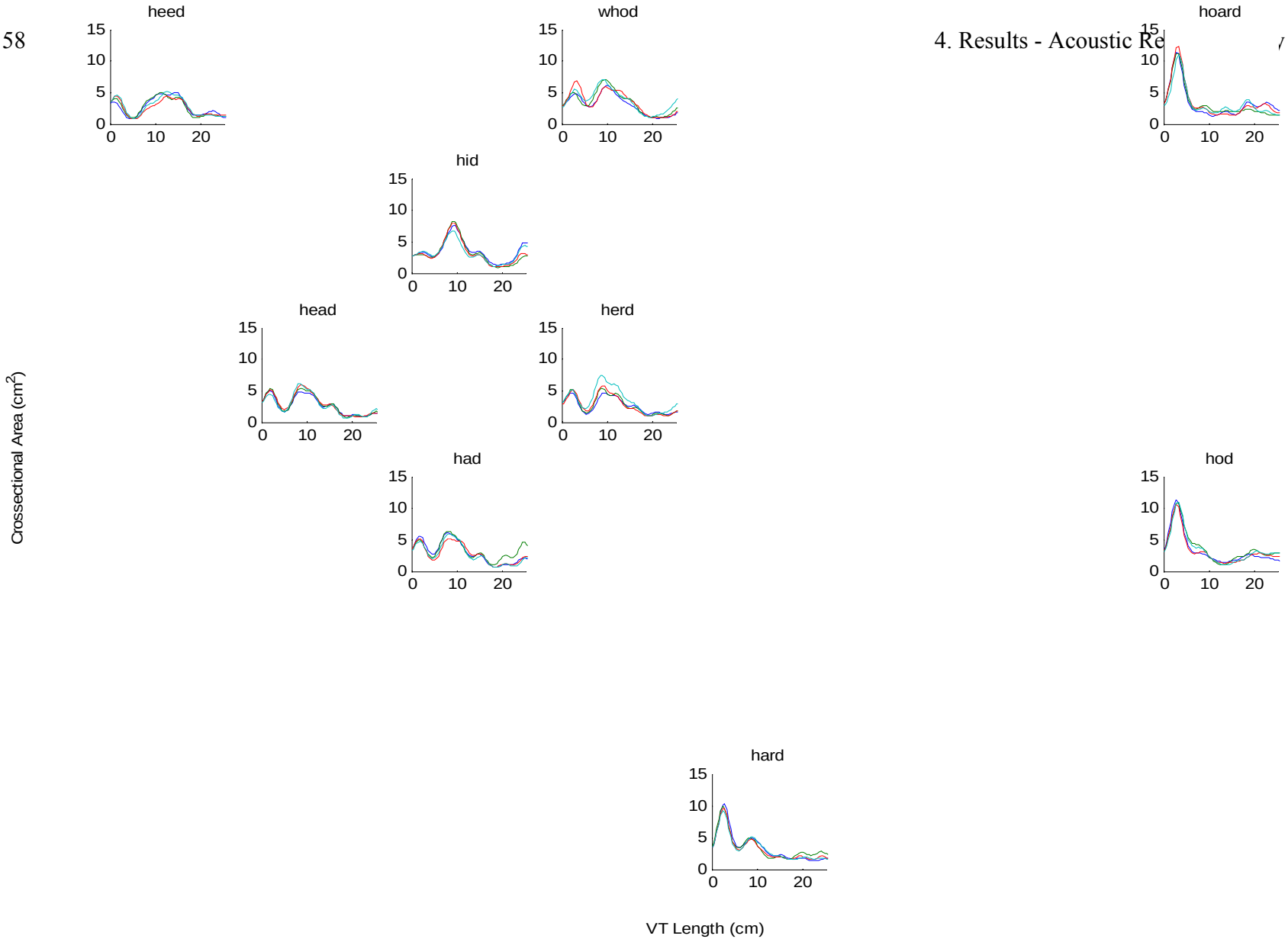


Figure 4.5: Cross-sectional area vs. vocal tract length for the nine monophthongs collected using acoustic reflectometery from SP05

It was speculated, however, that the vocal tract shape for a number vowels were not compromised as much as the other, Considering the expected lip opening of the various vowels presented in this study, and comparing this to the size of the mouth piece, it was speculated that the vowels 'HAD', 'HERD' and 'HOD' had lip openings least influenced by the presence of the mouth piece as opposed to 'HARD', for which the lip and jaw openings are expected to be large.

In terms of the effect the mouth piece has on the resulting resonance values calculated from the vocal tract shape, it was speculated that the second resonance would not match the second formants extracted from recorded speech. It was speculated that for vowels such as HOD and HARD, for which the jaw and lip openings were forced to be smaller than normal by the mouth piece, the calculated second resonance would be much smaller than that of the second formant's. This effect would not be as apparent on the 'HAD', 'HERD', and 'HOD' vowels as they are, as mention before, speculated to be the least affected by the mouth piece.

Similarity in vocal tract shape for vowels within an individual speaker

Upon observation, it can be seen that there is a great similarity in the 'HOARD', 'HOD' and 'HARD' vowels each of the 5 speakers. From Figures 4.1-4.5, it can be seen that for any individual speaker that these three vowels have a very similar shapes. This was expected, as the 'HOARD', 'HOD' (high back vowels), and 'HARD' (low front vowel) are expected to have large oral and small pharyngeal cavities. Their similarity was further increased by the presence of the mouth piece, which limited the variability of the starting area at the lip opening. For these three vowels, as the vocal tract are very similar in shape, it is expected for their formant values to display a degree of similarity. This is investigated in section 4.4

Comparing the vocal tract shape across speakers

Looking at Figures 4.1-4.5, it can immediately be seen that the 'HOARD', 'HOD', and 'HARD' vowels are not only similar in shape within an individual speaker, they also had the same overall shape across the five speakers. As mentioned previously, theses three vowels are expected to have a large oral region and a small pharyngeal region, and this has been captured accurately by the AR method for all the speakers. Looking at the 'HEED' vowel, it was also found that the vocal tract shape this vowel was similar across the different speakers. For the other vowels however, there are several noticeable differences in the tract shapes.

One important point to note is that though the general shape of the vocal tract may be similar for a number of the vowels, the magnitude of the cross-sectional area function differs between the speakers. For instance, while the shape for the 'HOD' vowel is similar for all 5 speakers, the area function of HOD' for SP01, SP03 and SP05 have peaks of around $10cm^2$, while SP02 and SP04 have peaks closer to $15cm^2$. It is important to realise that though the general geometry of the vocal tract may be similar for the same vowels across different speakers, the physical dimensions of the vocal tract structure will vary from individual to individual.

On the other hand, a number of vowels showed very different area functions between the 5 speakers. The target vowels for which the tract shape varied greatly across the speaker were 'HEAD', 'HAD' and 'HERD'. Taking a look at 'HEAD', it can be seen that there are two types of shapes present in Figures 4.1-4.5. First of all, SP02 and SP03 yielded cross-sectional area which were low in the oral region and high in the pharyngeal. As 'HEAD' is a semi high front vowel, this was expected.

However, for SP01, SP04 and SP05, the shape of the 'HEAD' was very different, and does not conform to what is expected. Looking at Figures 4.1, 4.4 and 4.5, it can be seen that the oral region of the vocal tract portrayed by the plot has a large cross-sectional area, while it decreases in the pharyngeal region. The difference in the two shapes can be clearly seen upon comparison of the plots. A similar difference is observed in the 'HAD' and 'HERD' vowels.

Once again, the speculated cause of this discrepancy within the results was attributed towards the mouth piece. In this case, as it was mention before, it is possible for the mouth piece to create tension in the oral region of the vocal tract. With the requirement of sealing the lips tightly around the mouth piece, it is possible that tension is placed in many of the articulators, especially the tongue. With the case of 'HEAD', it is possible that this tension stopped the tongue from moving as far away as the roof of the mouth as it normally would have, and thus causing and unusual vocal tract shape for the three speakers.

4.4 Derived vocal tract resonances

As illustrated in the last section, the jaw opening of the high vowels and the low vowels are likely to have been greatly compromised by the presence of the mouth piece, which would in turn affect the value of the second formant of these vowels. It was speculated that the vowels that were affected the most by the mouth piece were 'HEED', 'WHOD', 'HOARD', 'HARD', and 'HID'. It was speculated that 'HAD', 'HERD' and 'HOD' were affected the least. This is likely to be reflected in the resonance vs. formants results, with

the more compromised vowels having a second formant value very different to that of the second resonance, while the less compromised vowels have more similar values between the second formants and resonances.

For the purposes of this study, the first three resonance values were extracted from the vocal tract geometry. The first three resonance values for each vowel from all 4 datasets are plotted on the graphs. In Figure 4.6, the resonance results have been presented for each speaker, with the circle being the first resonance, square being the second, and triangle being the third. It can be seen that the values for the resonances are reasonably similar. This was particularly true for the first resonance, and this applied to all the vowels for all the speakers. The second and third resonances showed a larger variation in a number of the target vowels.

From Figure 4.6, certain trends in the resonance values can be observed. For the 'HARD', 'HOARD' and 'HOD' vowels, it can be seen that the first and second resonances values for these three vowels are quite similar within an individual participant. This was expected, as it was seen Figures 4.1-4.5 that the cross-sectional area function for these three vowels were very similar in shape. It was interesting, however, that this effect was not observed in the 'HEAD', 'HAD', and 'HERD' vowels when they also had very similar area functions. It is seen from the plots that the resonance values for these vowels are very different from each other.

It can also be seen from these figures that the general trend of the second resonances follows what is expected in relation to the vowel space configuration. Vowels such as 'HEED', 'HID', and 'HEAD', being closer to the high front of the vowel space, are expected to have higher second resonance values than the other vowels. This was indeed the case.

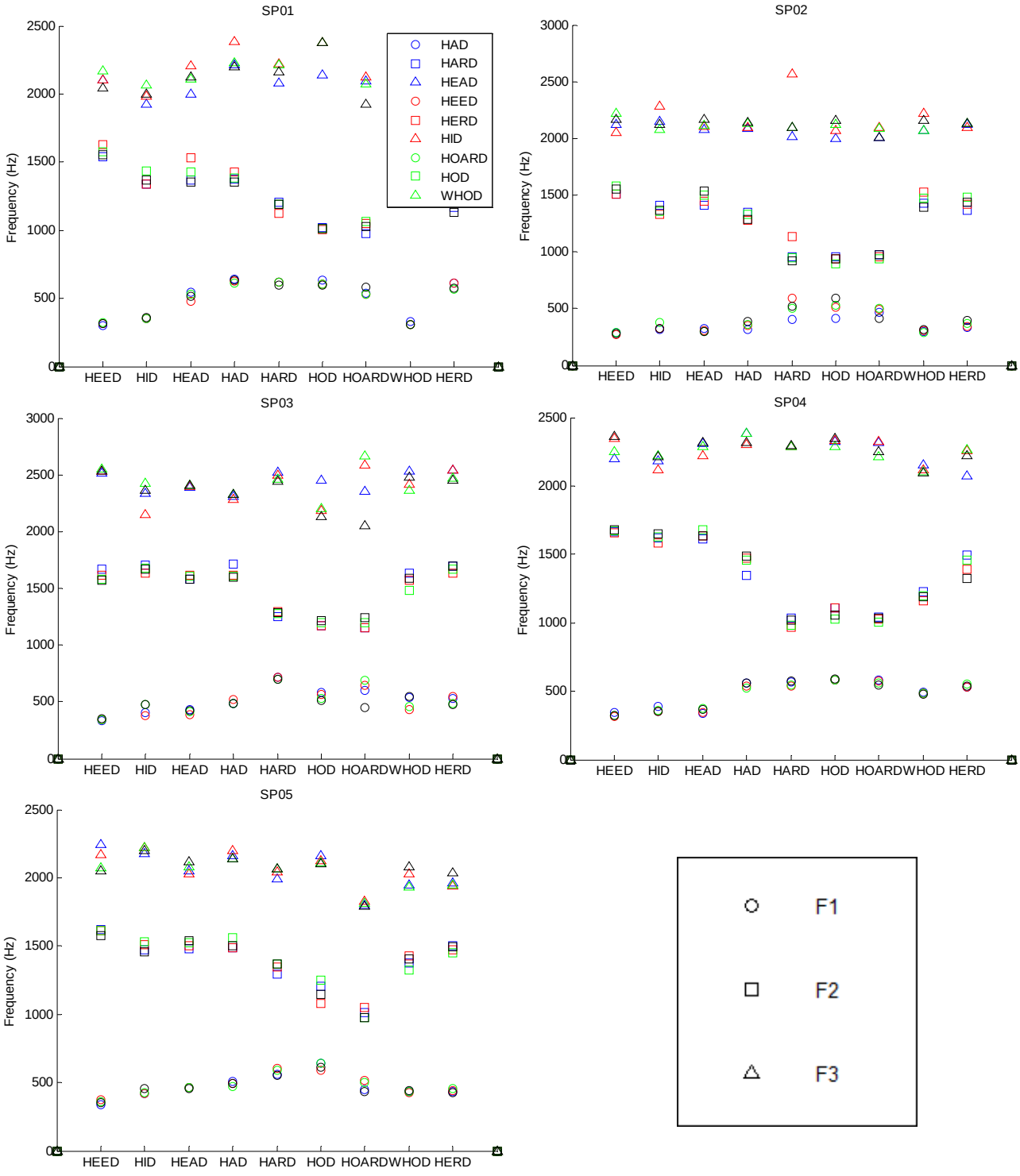


Figure 4.6: Plot of first three resonances calculated from the AR data for all speakers

R1 vs. R2 plots of the target vowels

In Figure 4.7, the first two resonances (R1 and R2) of each vowel are placed in a R1 vs. R2 plot. The scales of the R1 and R2 axis have been reversed, which allows for a direct comparison between the acoustic and articulatory space (C.I. Watson et al., 2009). By plotting the resonances in the described manner, it is possible to visually compare the relative positions the resonances to each other and whether they match their corresponding positions within the phonetic vowel space, and thus determining whether the derived resonances accurately reflect the behaviour of the vowels.

For the ease of data presentation in this section, only the results for dataset 1 from each speaker are presented. (Refer to Appendix B for the plots of the other datasets). Looking at the relative positions of the vowels presented in Figure 4.7, it is possible to see that a number of the target vowels are orientated in the approximate regions of where they would be expected within the vowel space. For example the high front vowel 'HEED', denoted by '*' within the plots, is situated in the top left corner of the graph while the back vowels such as 'HOARD' and 'HOD' are located on the right.

However, the positions of the rest of the vowels are not entirely consistent with what is expected when compared to the phonetic vowel space. As discussed previously, it was speculated that the rigidness the speakers were required to hold their articulators during the measurement process induced a certain amount of distortion in the vocal tract shape. This was perhaps demonstrated the most clearly by the 'HAD' and 'HERD' and 'HID' vowels, where it can be seen in Figure 4.7, the relative position of these vowels compared to the other vowels were very inconsistent across the different speakers

Not only this, it was also seen from the plots that the position presented for the vowel 'HARD' was not consistent with what is expected. 'HARD', being a low front vowel, is expected to be placed around the bottom centre region of the plot for it to match its position within the vowel space. This was clearly not the case for all the speakers except SP01. In general it can be said that vowels plotted in the R1 vs. R2 representation of the AR data did not yield a very similar vowel placement to that of the phonetic vowel space.

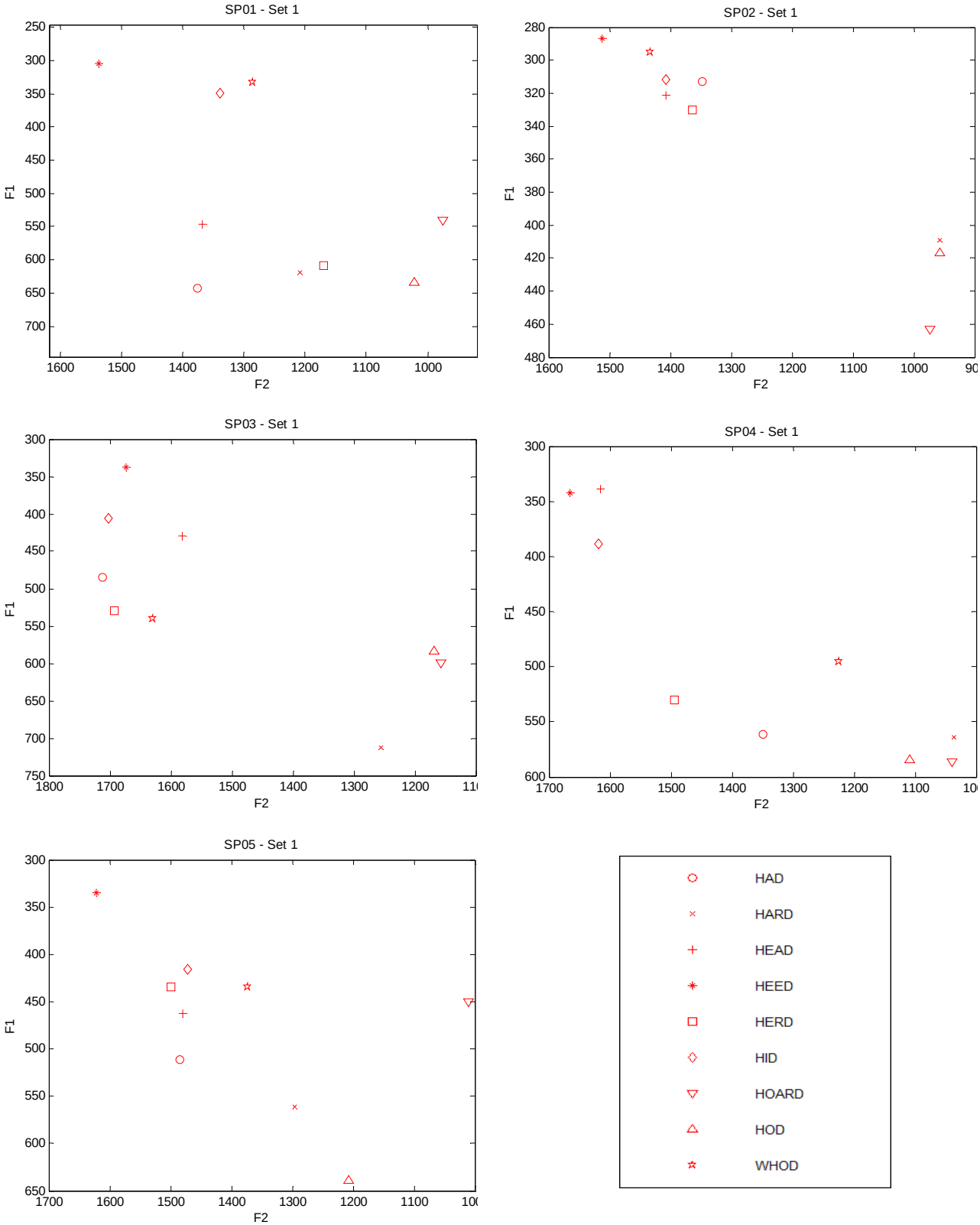


Figure 4.7: First resonance vs. second resonance plot of all vowels collected with AR method

Looking at Figure 4.7 it was noticed that the range of the first resonance values presented was relatively small, with the range being no larger than 350 Hz for any of the speakers. This was expected due to the mouth piece used in the data collection process. As the size of the jaw opening has a direct effect on the first resonance, the restriction in jaw and lips openings caused by the mouth piece also restricted range of the first resonance values. This was also the reason for the many of the vowels to be clustered at approximately the same R1 values.

AR resonance vs. Recorded speech formants

For the purposes of this study, in order to determine the applicability of the AR method, it was important to compare the resonance deduced from the vocal tract shape to the formants extracted from recorded speech. In Figure 4.8, the first two resonances of each vowel obtained by the AR method are plotted against the first two formants extracted from the recordings of the target vowels. These plots visually present the differences between the resonance and formant values for all the five speakers.

At first glance it can be seen from Figure 4.8 that the first resonances and formants were, in general, reasonably similar from each other for all the speakers. A number of discrepancies were observed, such as the 'WHO'D' vowel in SP03, SP04 and SP05, but these differences were not large. As discussed previously, this is likely caused by the mouth piece which compromises the jaw opening and prevents lip rounding.

Looking at the second resonances, it is clear that for many of the vowels the differences between the resonance and extracted formants are large. Evidently, the vocal tract shape obtained using the AR method is not suitable for a reliable estimation of the second formant of all the target vowels. However, it was observed that for certain vowels the calculated and extracted formants bear a certain similarity. For instance, it was observed that for 3 out of 5 speakers (SP01, SP03, SP05), the second resonance for 'HARD' and 'WHO'D' closely resembles the formants extracted from the recording. This was also observed for 'HOD', where it was seen in SP01, SP03 and SP04.

It was discussed previously in section 4.3 that the 'HAD', 'HERD' and 'HOD' vowels were speculated to have been influenced the least by the effects of the mouth piece. If this was true, it would be likely for the resonance values for these vowels to match the formant values more closely than the other vowels. Looking Figure 4.8, it can be seen that this was not true for the 'HAD' vowel, for which the second formant and resonance values were very different for all 5 speakers. The 'HERD' vowel, however, had a

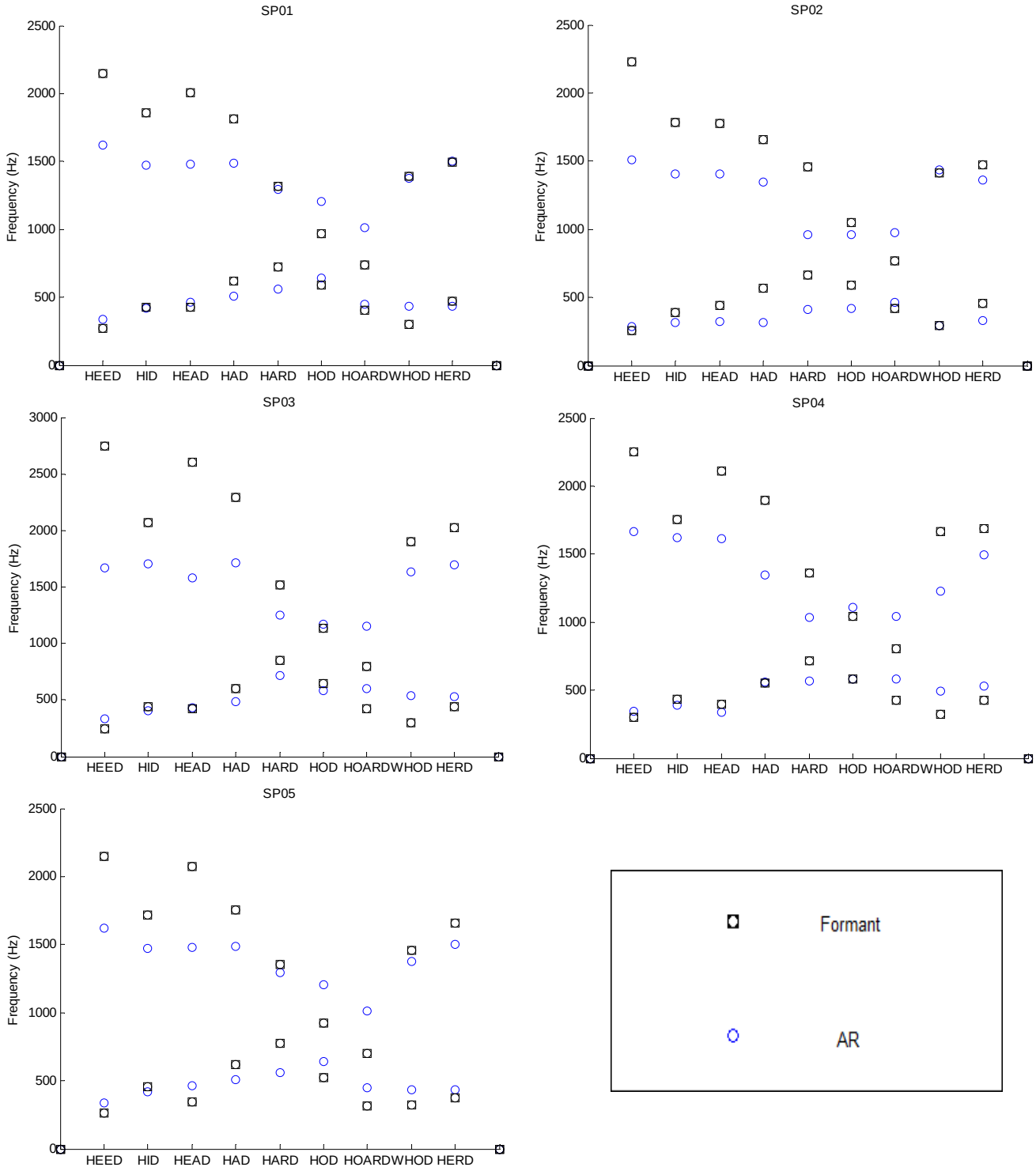


Figure 4.8: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech.

smaller difference between the second formant and resonance values, with the values SP03 being matched almost perfectly. The difference is even smaller for 'HOD'.

An interesting observation made from Figure 4.8 was the fact that the 'HARD' vowel, for which the lip and jaw openings should have been affected greatly by the mouth piece, had a reasonably close match of the formant and resonance values for 3 out of 5 speakers (SP01, SP03 and SP05. Looking at the results, it is found that only 3 out of 9 vowels for which the resonance was deduced for were match to the formants with a decent level of accuracy. This indicated that while the AR method does not yield accurate estimations of the formant values for all target vowels.

5 Results - MRI

5.1 Chapter overview

In this chapter, the results obtained from the MRI technique will be presented. In the first sections of the chapter, the vocal tract lengths of each vowel are presented and the vocal tract cross-sectional area for each speaker is plotted and presented in a vowel space arrangement, which is accompanied by observational comments. In the following section, the formant values deduced from the presented cross-sectional area function will be presented for each vowel for each speaker in graphical form. This is also accompanied by comments.

5.2 Vocal tract length

The vocal tract length for each vowel was extracted from the MRI data for the two datasets. The average is taken from the two sets for each vowel. The vocal tract lengths vary from vowel to vowel and the range of these lengths is presented in the table below for each speaker. The mean of the tract length across the vowels are also presented. The vocal tract length for each vowel for all speakers can be found in Appendix A.

	SP01	SP02	SP03	SP04	SP05
Range of lengths (cm)	15.1-16.9	15.1-15.9	13.9-15.5	18.2-21.2	18.2-19.6
Mean length (cm)	15.9	15.6	14.6	19.9	18.8

Table 5.1: The range and average lengths presented in the results of the different vowels of interest for the MRI method

5.3 Vocal tract shape

As with the AR results, the cross-sectional area function for each vowel is plotted in its approximate positions within the phonetic vowel space. This is repeated for all speakers, and the plots are presented in Figures 5.1-5.5. Both sets of data for each vowel are plotted on the same axis. The midsagittal MR image for each vowel is also presented in the results next to the graph of the area function.

Comparison of the vocal tract shape across the MRI datasets

For the MRI method, two sets of data were collected for each target vowel for each speaker. In Figures 5.1, 5.2 and 5.5, it can be seen that the area functions are relatively consistent across the two different image sets. Where deviations did occur, the area functions between the two sets still shared a similar shape. In Figure 5.4, the vocal tract shapes for the two datasets are almost an exact match for all the target vowels. It was noted that of all the speakers, SP03 presented results with the largest variation between the two data sets. This can be observed in Figure 5.3, where none of the vocal tract shapes from the two datasets were an exact match.

Upon investigation, it was found that there were two main sources of deviation. The first cause of difference in area function between the two data sets is due to the variations in area that are inherent to the marking method described in Chapter 3. The size of the cross-sectional area at a given point is dependent on an individual's judgement of the air-tissue boundaries, which is open to interpretation, which may cause one set to be marked differently to the other. The other source of the deviation was an actual difference in the vocal tract shape captured by the MR images. It was found that for different instances of the same vowel the vocal tract can have different orientations, though the general shape of the shape stays reasonably similar. The general similarity in shape is encouraging and would suggest a certain level of reliability in the measurement method, as well as the reliability of the cross-sectional area extraction process.

Vocal tract shape: Results vs. Expected geometry

Like with the AR results, the MRI data presented the locations of constrictions and the openings within the vocal tract that are what would be expected from each vowel. This is demonstrated with the high front vowel 'HEED', as seen in Figure 5.1, clearly shows the expected configuration of a small oral cavity and a

large pharyngeal one, or the high back vowel 'HOARD', which has the expected large oral cavity and small pharyngeal cavity.

As with the AR results, a clear increase the oral cavity size can be observed moving from the front vowels through to the back vowels. In Figure 5.1, it can be seen that the oral cavity in 'WHOD' is larger than that of 'HEED's. 'HOARD', which is further back than 'WHOD', has an even larger oral cavity. It can be seen that the MRI method has captured the increase in oral cavity size and decrease in pharyngeal cavity size accurately as vowel move further back within the vowel space.

As mentioned previously, when transitioning from a high vowel to a low vowel, it is expected for high vowels to have smaller oral cavities compared to a low vowel due to the fact that low vowels have larger jaw openings which increase the size of the oral cavity. Like in the AR results, this effect was observed vocal tract shape presented in the 'HEED', 'HEAD', 'HAD', 'HUD' and 'HARD' plots. Starting from the high vowel 'HEED', the vowels transition towards a lower setting until the low vowel 'HARD'. This means that the oral cavity in these vowels are expected to grow larger and the pharyngeal cavities smaller as the vowels progress from 'HEED' to 'HARD'. This was observed in Figure 5.1, where the oral cavity of the vowels become larger and the pharyngeal cavity become smaller incrementally as the vowels progress from 'HEED', 'HEAD', 'HAD', 'HUD' through to 'HARD'. From Figures 5.1-5.5, it can be observed that the vocal tract geometry collected with the MRI method have vocal tract shapes similar to what would be expected for each vowel. The trends observed between the vowels were observed for all 5 speakers.

Like the acoustic reflectometery results, there are a number of similarities that can be seen in the plots. An example of this is the vocal tract shape presented for 'HOARD' and 'HOD', which are visually very similar. This was true for all the speakers. Similarities were also presented between 'HUD' vs. 'HARD', and 'HEAD' vs. 'HAD'. These similarities in the shapes indicate that these vowels have similar vocal tract configurations. As the resonance values for this study are calculated from the vocal tract structure, it would be interesting to observe whether the visual similarity in the tract shape will result in a similarity in the calculated formants. This will be further outlined in the derived resonance section (section 5.4).

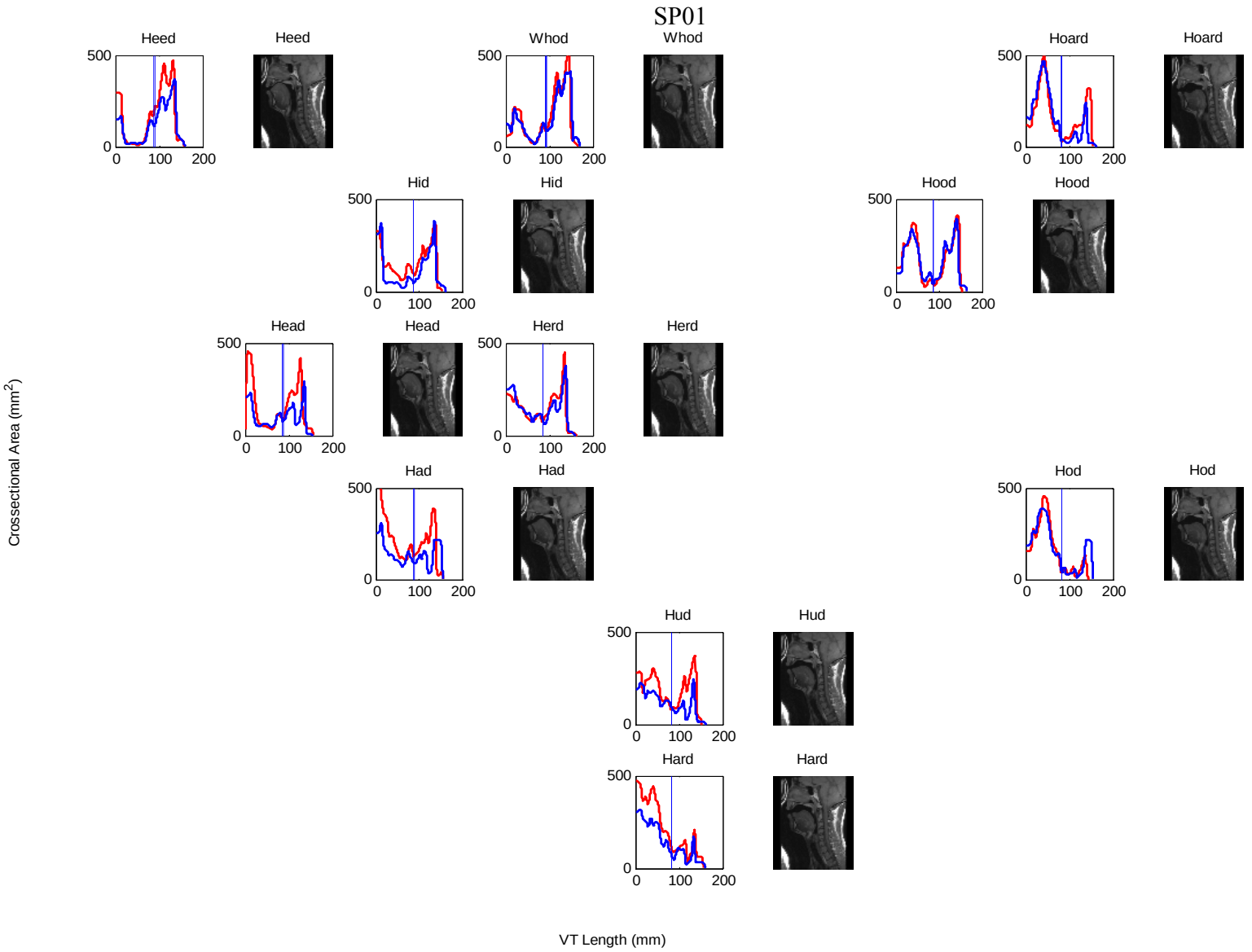


Figure 5.1: Cross-sectional area vs. vocal tract length for the eleven monophthongs collected using the MRI from SP01

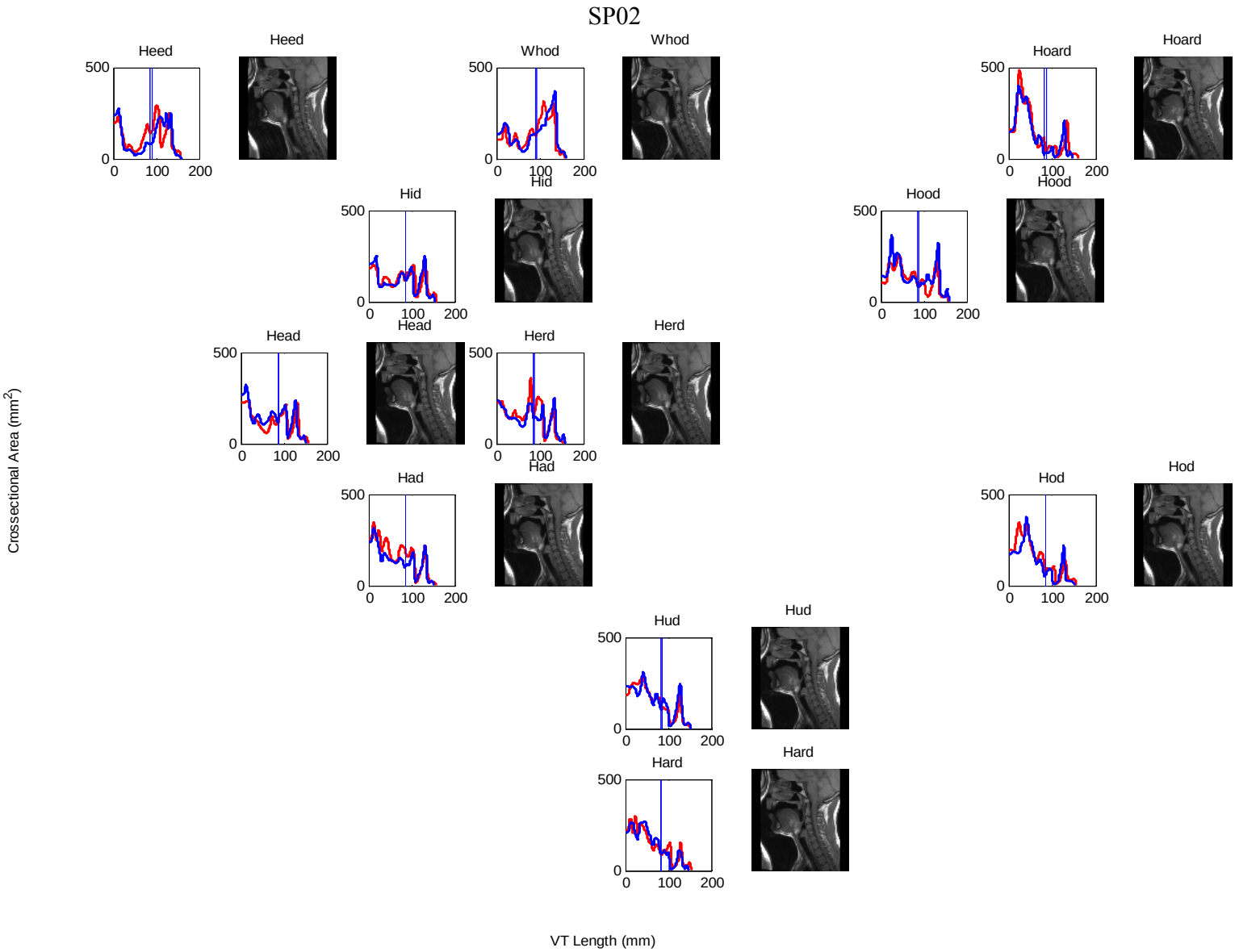


Figure 5.2: Cross-sectional area vs. vocal tract length for the eleven monophthongs collected using the MRI from SP02

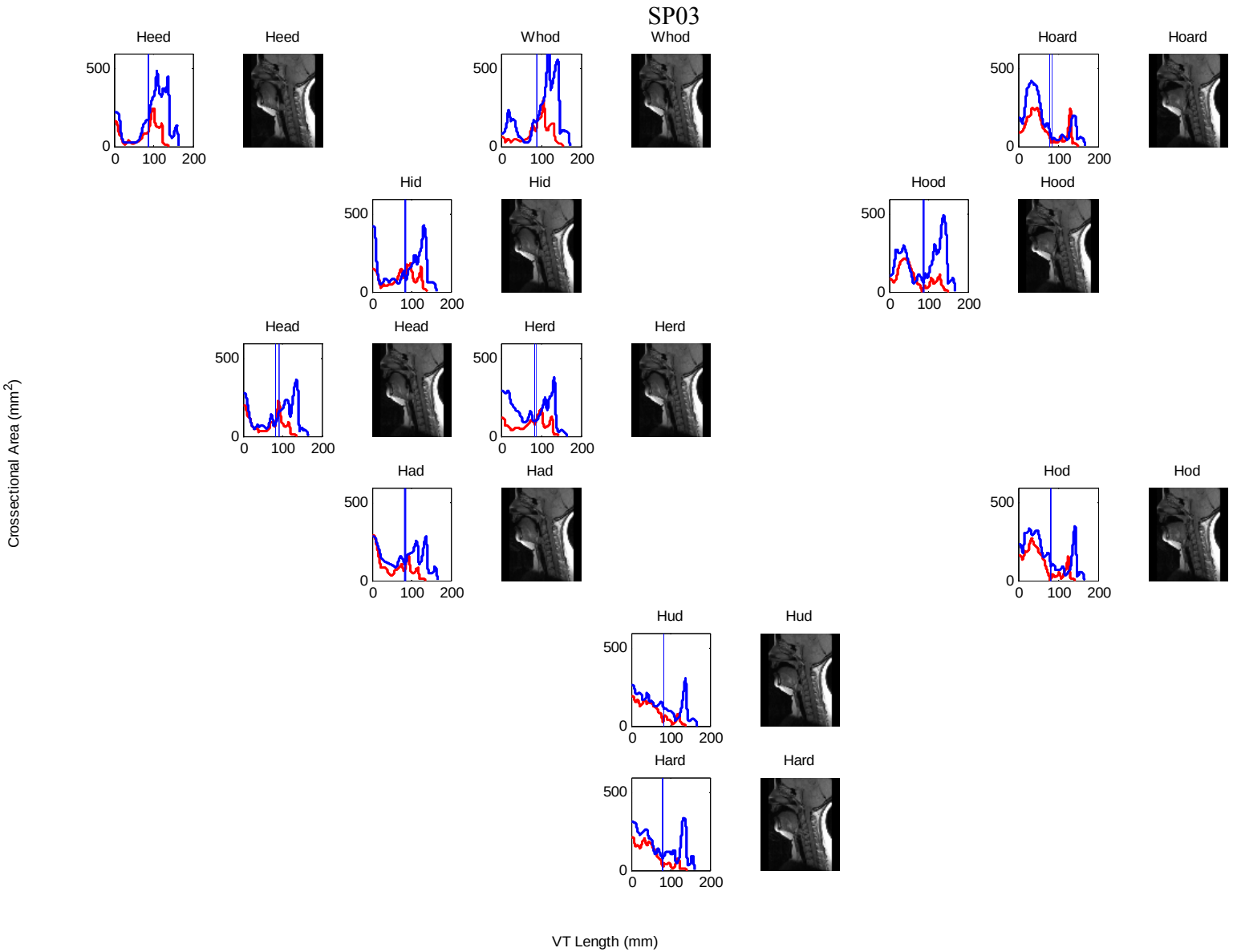


Figure 5.3: Cross-sectional area vs. vocal tract length for the eleven monophthongs collected using the MRI from SP03

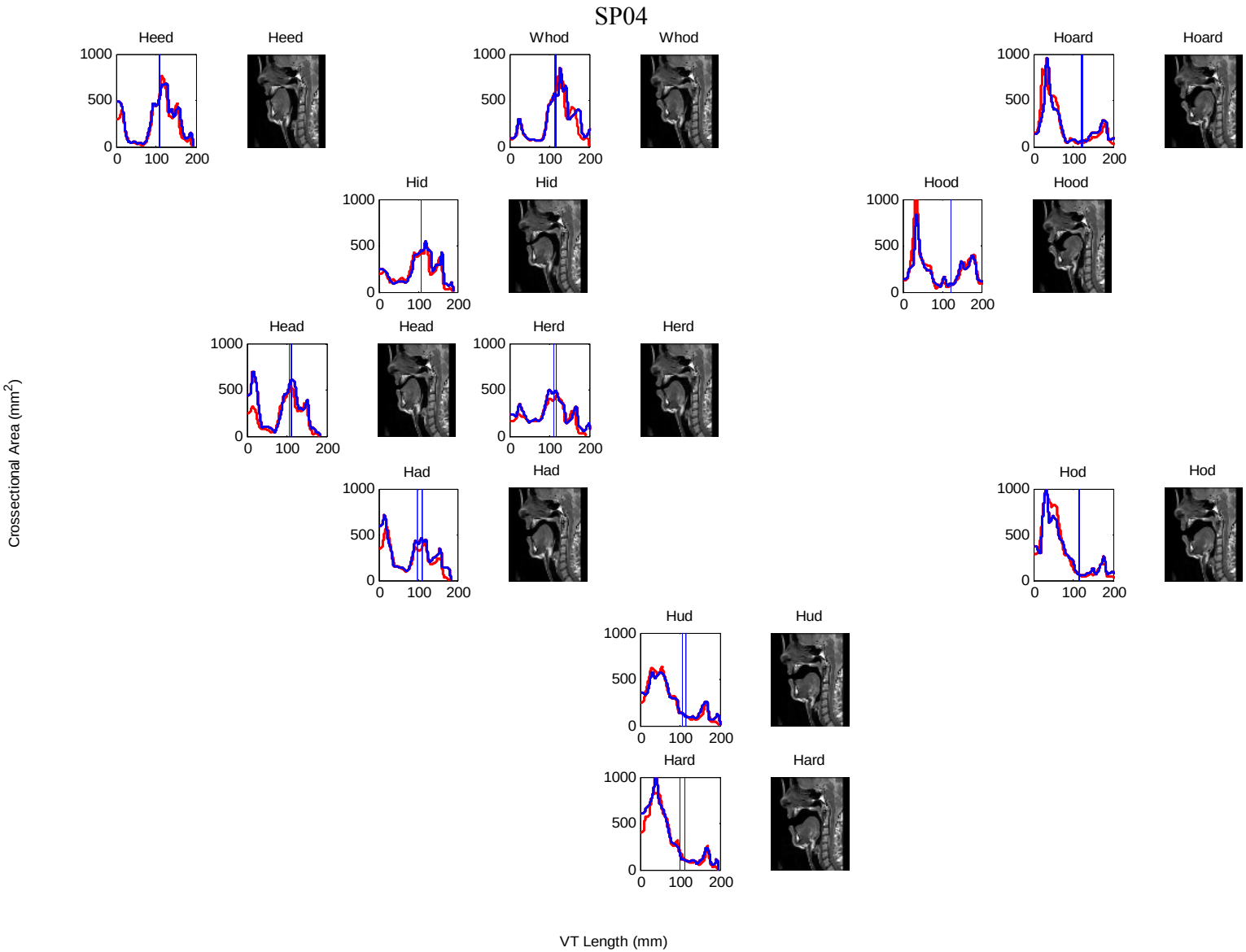


Figure 5.4: Cross-sectional area vs. vocal tract length for the eleven monophthongs collected using the MRI from SP04

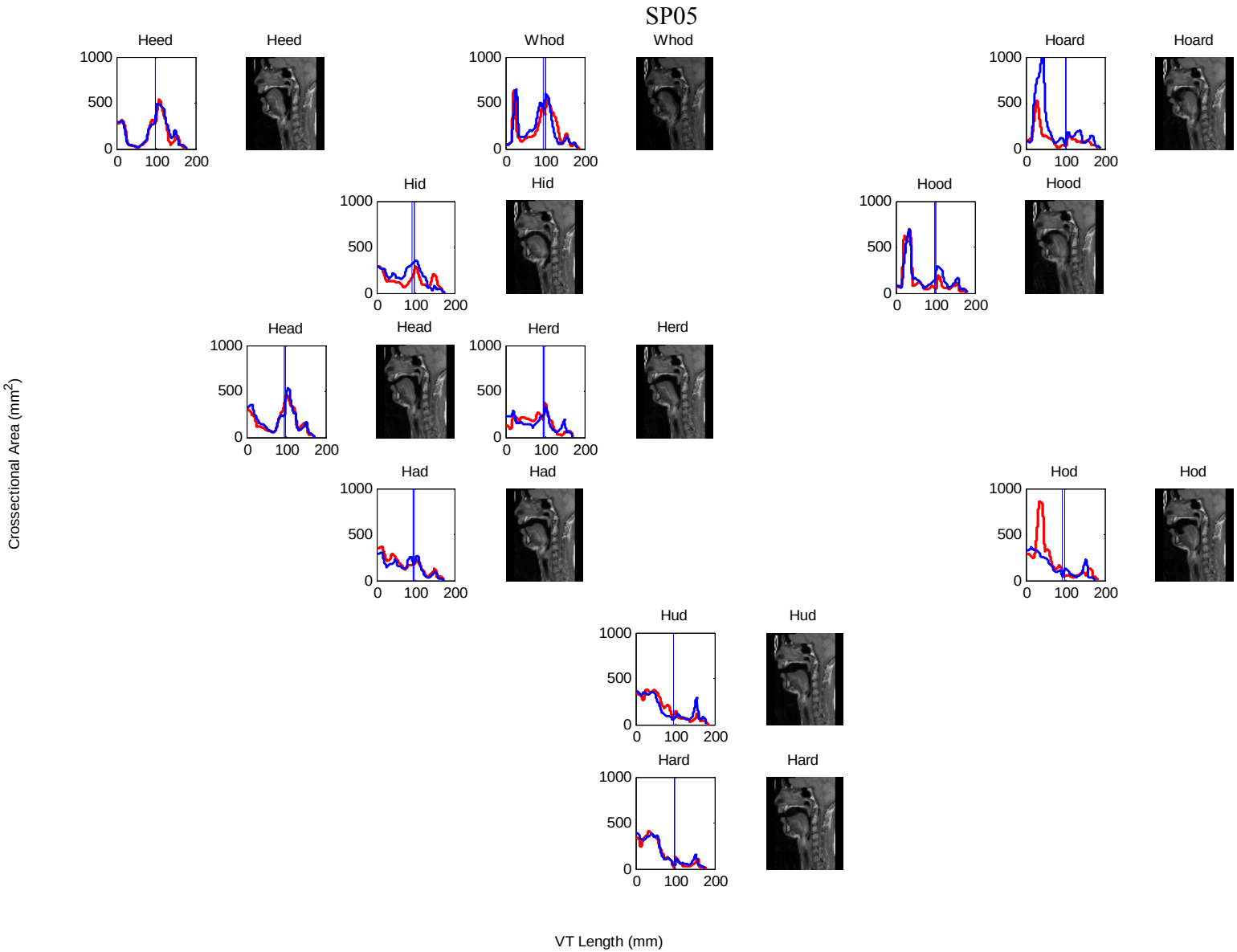


Figure 5.5: Cross-sectional area vs. vocal tract length for the eleven monophthongs collected using the MRI from SP05

Similarity in vocal tract shape for vowels within an individual speaker

While certain similarities in the general shape of the vocal tract across different vowels can be observed from Figures 5.1-5.5, it can be seen that the differences between the tract shapes were substantial enough to be noticed. For instance, looking at Figure 5.1, it can be seen that 'HOARD' and 'HOD' has very similar vocal tract shapes. However, it can clearly be seen that that 'HOD; has a smaller pharyngeal cavity than that of 'HOARD's. Characteristics such as this which can be used to distinguish between the tract shapes are much more apparent in the MRI method as opposed to what was shown in the AR results

Comparing the vocal tract shape across speakers

Unlike the AR method, where only a selection of the vowels had the same shape across the different speakers, the MRI method yielded similar shaped vocal tracts between the speakers for all the vowels. This was not surprising, as the general shape of the vocal tract for any given target vowel is reasonable consistent across individuals.

For example, for a high front vowel, the jaw and mouth opening is small and the tongue is at the front of the oral cavity, and regardless of the individual producing the vowel, his or her articulators will be placed accordingly. As there are not compromising factors affecting the position of the articulators and the MR imaging technique is accurate, it is no surprise that the same vowel articulated across different speakers yielded the same shape

However, this is not to say that no differences were observed in the vocal tract shape across the different speakers. As mentioned before, the physical size of the vocal tract will vary from person to person, and thus the magnitude of the area functions displayed in the plots would obviously be different.

Another point of difference to take note of is the localised details present in the vocal tract shape. Looking at a vowel such as 'HOD' for any of the speakers, it can be seen that in the pharyngeal region, there is a small increase in the cross-sectional area right towards the end of the vocal tract. While details like this can be observed in all the speakers, the size of these feature vary in size, shape, and sometimes location, giving each speaker defining characteristics to their vocal tract shape.

5.4 Derived vocal tract resonance

The first three resonance values for each vowel from both sets of data are plotted on the graph. In Figure 5.6, the resonance results have been presented for each speaker, with the circle being the first resonance, square being the second, and triangle being the third. The two datasets are distinguished by the colours blue and red.

For the purposes of this study, the first three resonance values were extracted from the vocal tract geometry. The first three resonance values for each vowel from the two datasets obtained with the MRI method are plotted on the graphs. In Figure 5.6, the resonance results have been presented for each speaker, with the circle being the first resonance, square being the second, and triangle being the third. It can be seen that the values for the resonances are reasonable similar. This was particularly true for the first resonance, and this applied to all the vowels for all the speakers. Taking a closer look at the values of the second resonances in Figure 5.6, it was seen that the second resonance values collected across the two datasets for each vowel were also quite similar, thought the difference was slightly bigger than that observed from the first resonances.

Though the value of the resonance frequencies of a signal is not unique to one specific vocal tract configuration, it was intuitive to assume that people generally produce the same sound using similar vocal tract shapes. However it was interesting to note that for a number of vowels produced by speakers within this study, this was not the case. Referring back to Figure 5.3, it is seen that for a number of vowels such as 'WHO'D' and 'HEED', the vocal tract shape for the two datasets are drastically different for SP03 but looking at the comparison of the deduced resonance values for SP03 in Figure 5.6, there is no noticeable difference in the first and second resonances for either the ‘WHO’D’ or the 'HEED' vowel.

Looking at Figure 5.4 on the other hand, SP04 had two datasets which had very similar vocal tract shapes for every vowel, which means that very little variation is expected in the resonance values. However, the plot for SP04 in Figure 5.6 displays a clear difference in the second formants in a number of vowels. This indicates that the formant value calculations are significantly impacted by the finer details of the vocal tract shape. It is obvious from the results that a visually perceived similarity in the vocal tract shape, as observed in Figure 5.6, does not guarantee a similar resonance value.

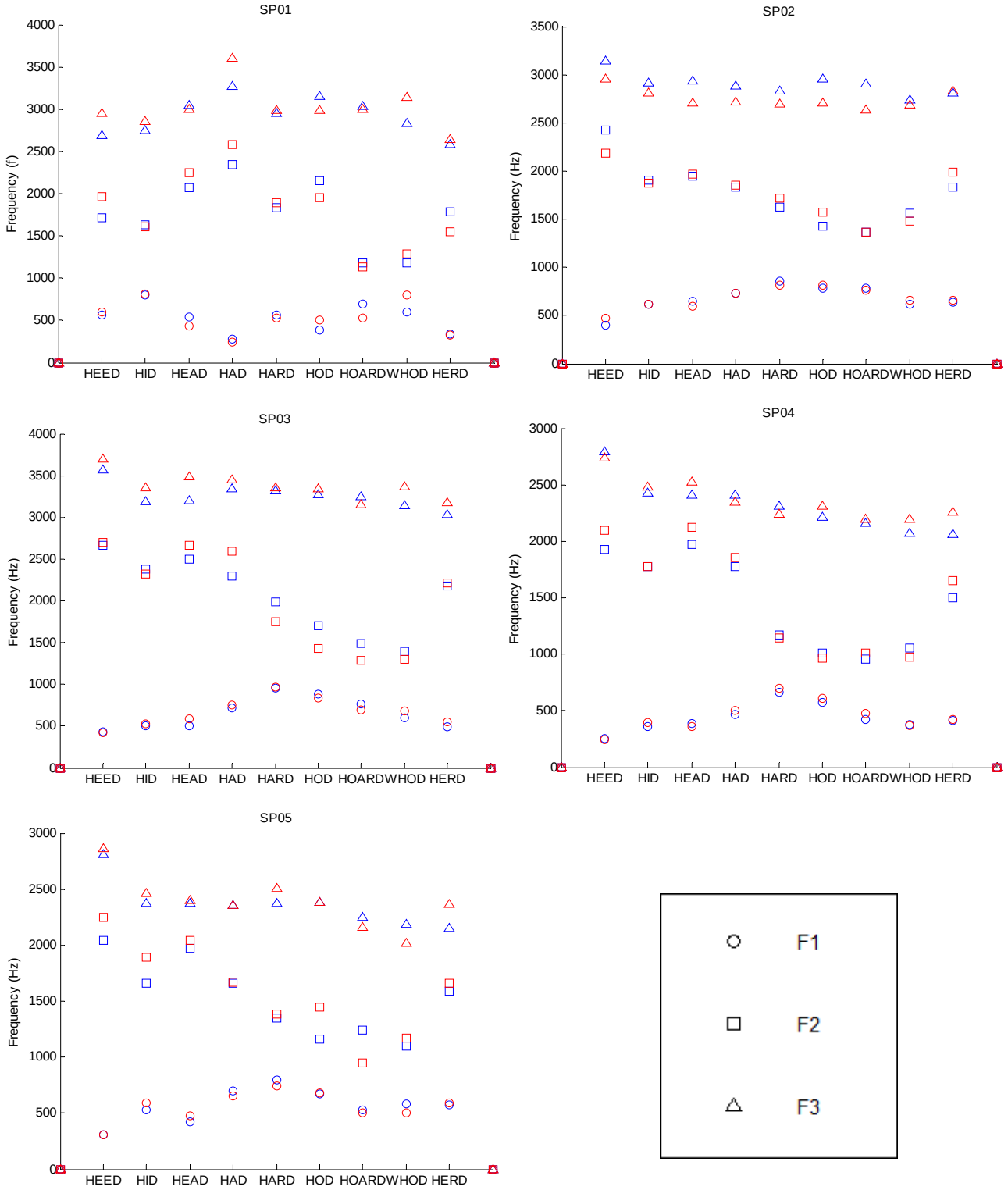


Figure 5.6: Plot of the first three resonances calculated from the MRI data for all speakers

R1 vs. R2 plots of the target vowels

In Figure 5.7, like in the AR results, the first two resonances (R1 and R2) of each vowel are placed in a R1 vs. R2 plot. For the ease of data presentation only the results for dataset 1 from each speaker is presented in this section, but it was found that the behaviour of the results were relatively consistent across data sets. (Refer to Appendix B for the plots of the other datasets).

Unlike the AR data, the R1 vs. R2 plot of the resonances obtained by the MRI method yielded most of the target vowels in their respective regions within the phonetic vowel space. In Figure 5.7, it can be seen that the 'HEED' vowel, which is a closed front vowel, is located in the top left region of the plots for all speakers and back vowels such as 'HOARD' and 'HOD' are located to the right.

Moving towards the lower front vowels from 'HEED', it is expected for 'HEAD', 'HAD' and 'HARD' to be seen. On the phonetic vowel space, 'HEAD' is positioned below and slightly to the right of 'HEED', this is followed by 'HAD', which is below 'HEAD', also towards the right, and finally 'HARD'. A similar configuration can be seen in Figure 5.7, with 'HEAD' 'HAD' and 'HARD' being below 'HEED', and moving increasingly towards the right. In general, most of the vowels were in their respective positions within the vowel space.

The range of the first resonances displayed in Figure 5.7 is significantly larger than what was shown for the AR method. With no restriction in the jaw opening in the MRI method, the first resonance values now range across up to 600Hz between the high vowel 'HEED' and low vowel 'HARD' opposed to the 350Hz shown in the AR results. This suggests that the resonance values calculated from the MRI results would likely be a better estimation of the formant values compared to the AR method.

MRI Resonance vs. Recorded speech formants

In order to evaluate how applicable the MRI method is in producing resonance values comparable to the formants extracted from recorded speech, the two sets of values are compared for each of target vowels. Like in the AR results chapter, the resonance and extracted formants are plotted on a graph for comparison. Figure 5.8 displays the MRI resonances vs. recorded speech formants for the 5 speakers.

From Figure 5.8, it is seen that the first resonance values for the target vowels in the MRI method are similar to that of the first formants'. From the plots the majority of the first resonance values are

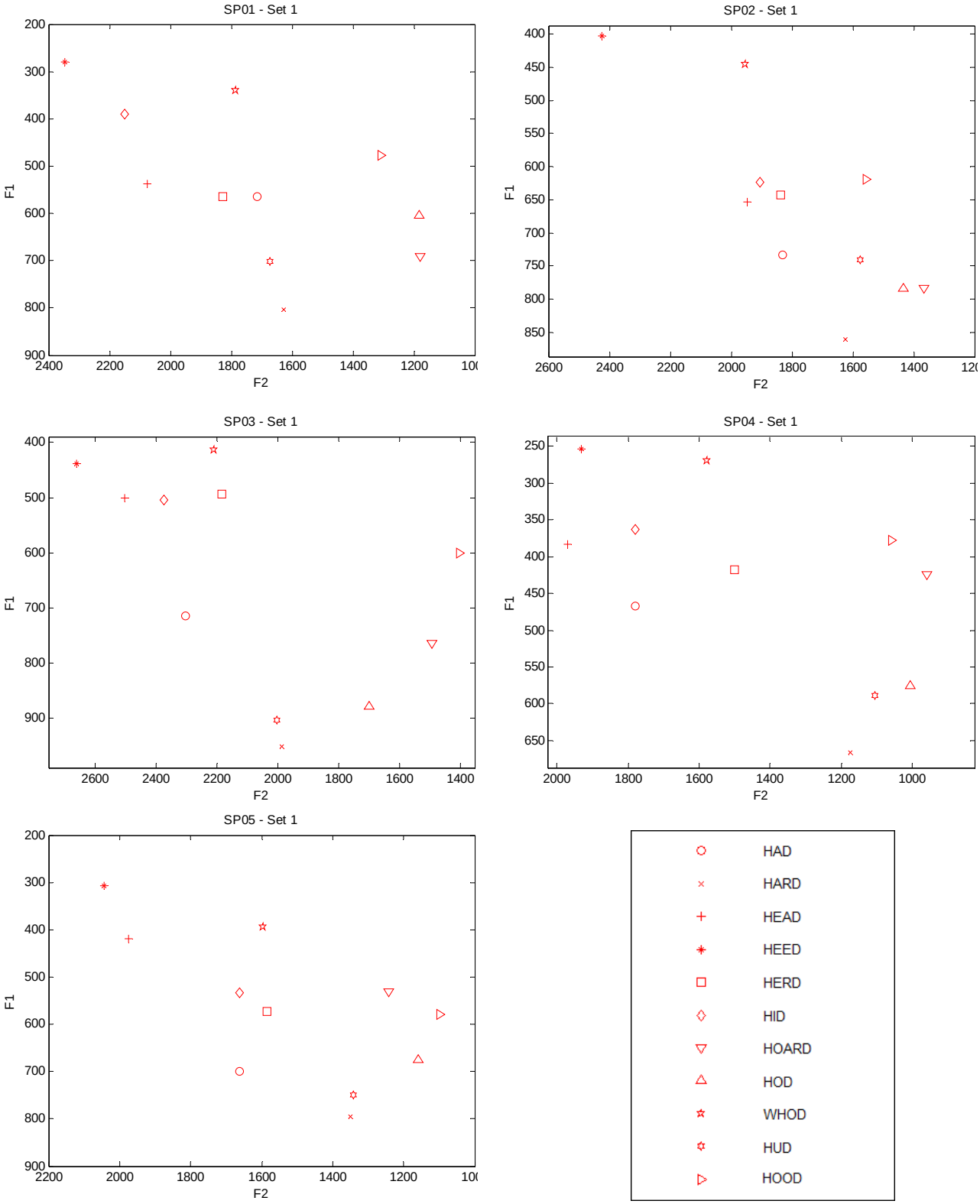


Figure 5.7: First formant vs. second formant plot of all vowels collected with MRI method

close to or overlaps the first formants values. The exception was the 'HOARD' vowel, where for 3 out of 5 speakers the calculated first resonance values were significantly different to the formants'. Looking at the second resonances, it can be seen that there is a similarity between the resonances and extracted second formants for many of the vowels, though clearly some discrepancies are present.

Looking at both the first and second resonances together, it can be seen that for all the speakers, the resonance values matched the formants reasonable closely for most vowels (Note that SP02 preformed the worst out of all the speakers). The exceptions were the 'HOD' and 'HOARD' vowels, which were badly matched for most of the speakers. From this it can be seen that the MRI method can be used to estimate the formants of recorded speech with a reasonable degree of accuracy.

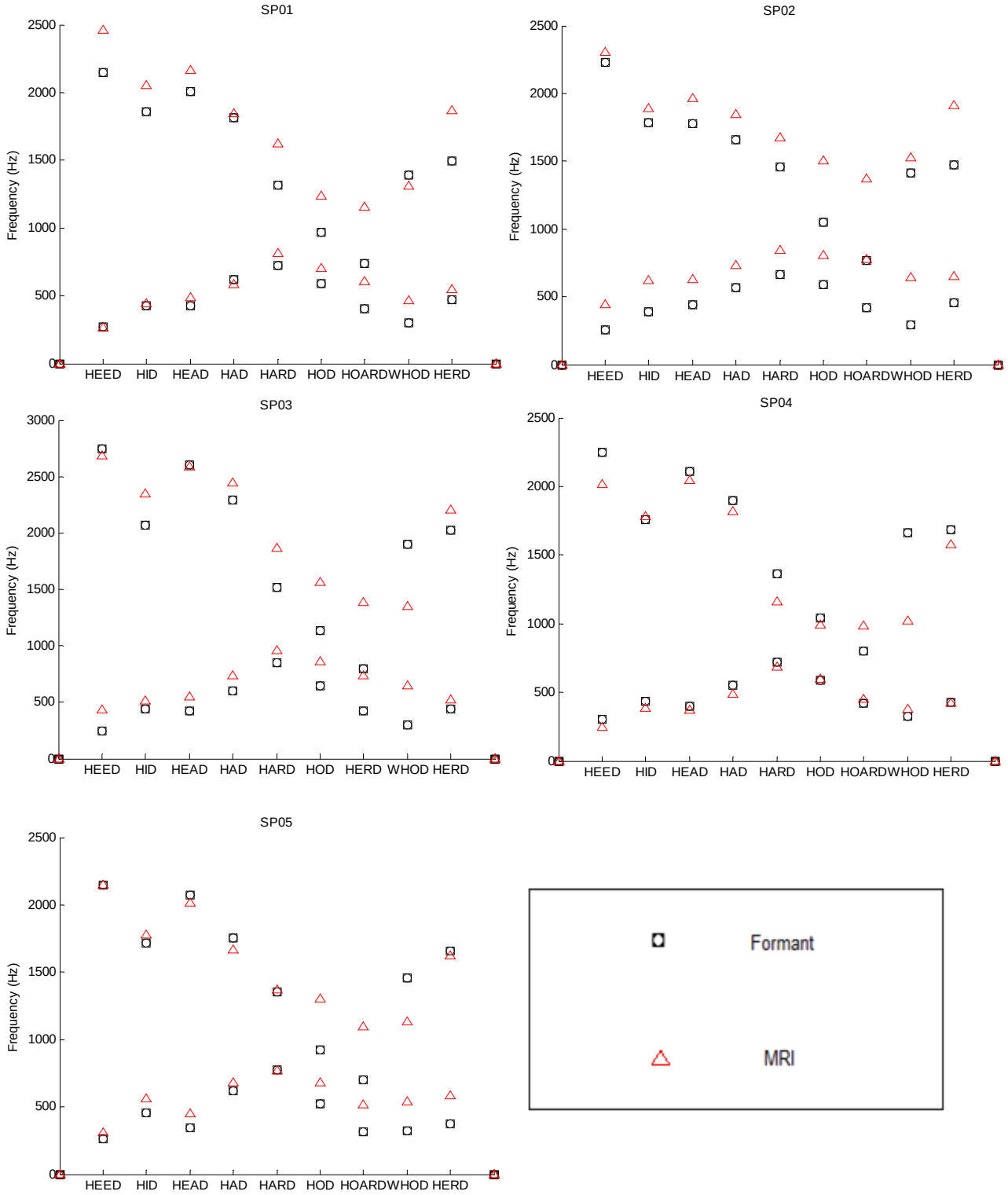


Figure 5.8: Plot of first and second formants calculated from MRI cross-sectional area function vs. first and second formants extracted from recorded speech.

6 Discussion

6.1 Chapter Overview

In this chapter, a comparison is made between the acoustic reflectometery (AR) method and the magnetic resonance imaging (MRI) method. The average vocal tract length of each vowel for each speaker for the two methods are compared and discussed. This is followed by a comparison of the vocal tract shapes and the resonances obtained from the two methods. A general discussion and comparison of the AR and the MRI methods is presented.

6.2 VT lengths

A graphical summary of the vocal tract length has been presented below. The average vocal tract length measured from the MRI is plotted against the average vocal tract length measured by the acoustic reflectometer. Refer to Appendix A for a full table of results.

It can be seen from Figure 6.1 that the vocal tract length obtained from the acoustic reflectometer was generally larger than that obtained from the MRI. This was true except for SP04 and SP05, where the lengths for the two methods were similar. The speculated reason for the difference in the vocal tract length for the AR and MRI methods was the fact that the participants were required to be in a seated upright position for the AR, opposed to the supine position of the MRI method. It has been shown in a previous study that the seated and lying position of the speakers caused differences to the length of the vocal tract.(C.I. Watson et al., 2009)

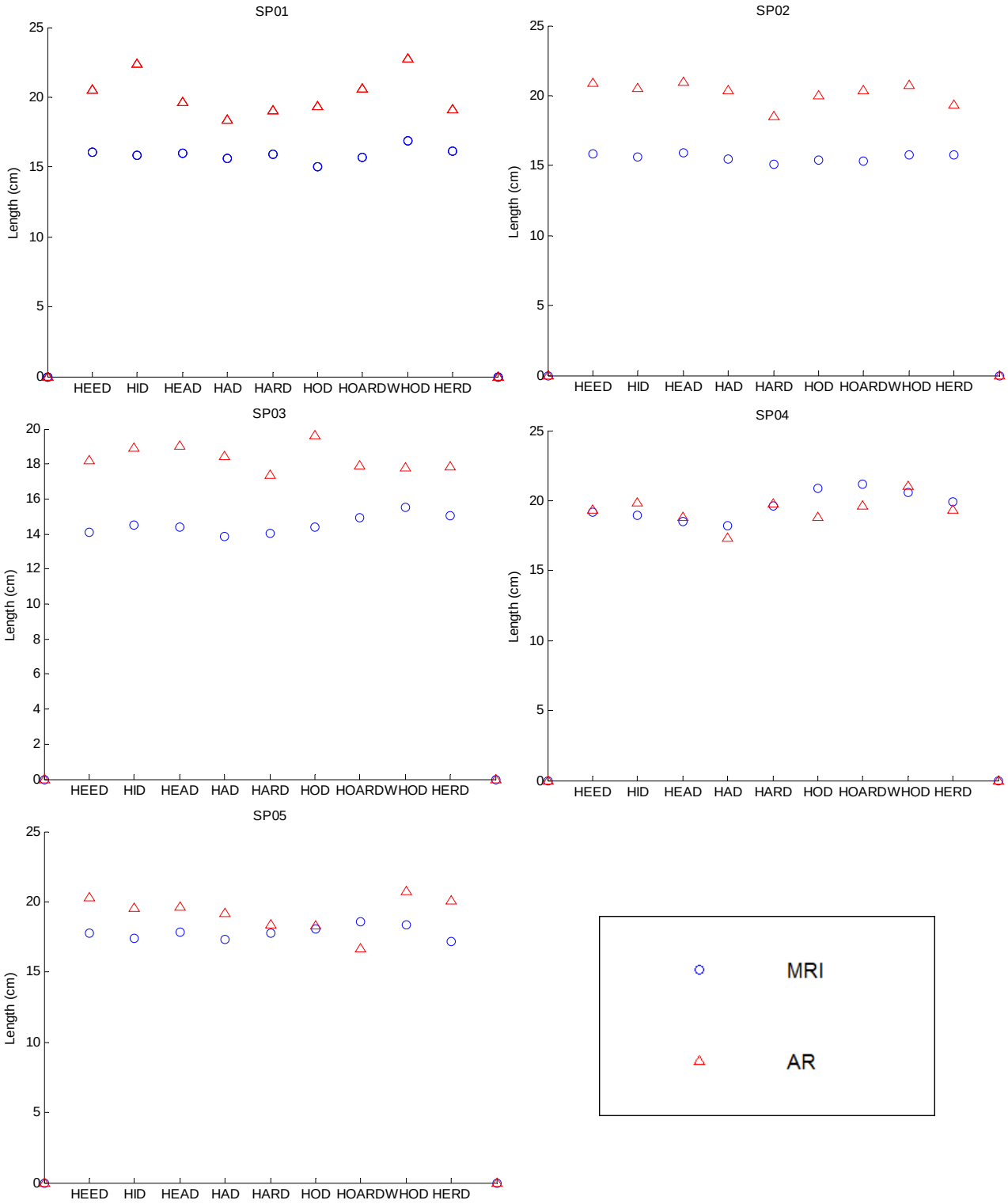


Figure 6.1: Plots of vocal tract lengths obtained from the AR and MRI methods for all 5 speakers.

A point of concern with the length obtained from the MRI method is that it appears to be shorter than what would be expected. It can be seen from Figure 6.1 that lengths from the MRI for SP01 and SP02 are both around 15-16cm. This is lower than what the expected 17 cm for males. The opposite can be said about SP04 and SP05, where the tract length is around 19 cm, which is too higher than expected.

SP04, which was the only female speaker in the study, had tract lengths in the 14-16 cm region, which is close to the expected average of 14 cm for females (Stevens, 2000). From Figure 6.1, the lengths obtained with the AR method can also be observed for each speaker. In the case of the AR, the lengths are longer than the expected values. For all the male speakers (SP01, SP02, SP04, SP05), the vocal tract lengths vary around 20 cm, which is higher than the expected 17 cm. For the female speaker (SP03), the average lengths were around 18 cm, which is larger than the expected 14 cm.

It was noted that one possible reason for the overestimation of the vocal tract length obtained from the AR method was the fact that there was a compensation process of adding 1 cm onto the length acquired from the raw AR data to account for the distance from the teeth to the lip opening. This was originally done as the vocal tract length extracted from the raw AR data was calculated from the start of the mouth piece, which was at the edge of the teeth. It was observed in other studies that this compensation process was not common practice, and the tract length as the raw value presented from the raw recordings (Steve An Xue & Hao, 2006) (S. A. Xue, Cheng, & Ng, 2010). In these studies, the male and female subjects had vocal tract lengths close to the expected lengths of 17 and 14 cm.

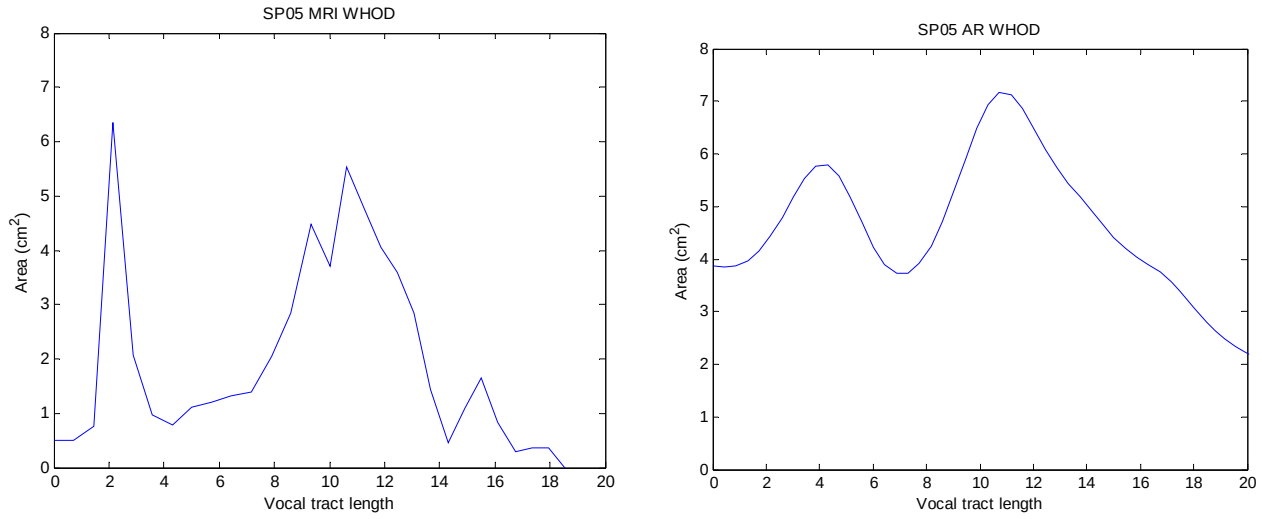
However, even by taking this into account, the lengths presented in the AR data in this study are still much larger than the commonly expected vocal tract lengths for both males and females. It was assumed that an unknown reason was present which caused the lengths from the AR to be overestimated. If it was assumed that the MRI lengths were underestimated, the large differences between the lengths of the two methods can be addressed. In SP01 and SP02, the differences in length are as large as 5 cm, which is highly unrealistic. The reasons for the over and under estimation were not discovered in this study, and can be a topic for future research.

6.3 Vocal tract shape

As observed in the results sections, both methods used in this study have provided cross-sectional area functions which represent the vocal tract shape in an expected manner. It was seen from both methods

that the area function plots accurately indicate the locations of the openings and constrictions present in the vocal tract when it is configured to produce a target vowel.

One of the differences observed between the two different imaging methods is the area representing the opening at the lips. As mention previously, the mouth piece restricts the movement of the articulators, which means all the AR results start with the same area. Due to this restriction, vowels such as 'WHO'D' and 'HEED', which are expected to have a small cross-sectional area at the lips, has a much larger starting area than what would be expected for the AR method. This is illustrated below: in Figure 6.2(a), the area at the lips is approximately $0.5cm^2$, while in Figure 6.2(b) it is almost $4cm^2$. By the same reason, vowels such as such as HARD and HOD would have smaller lip openings displayed for the AR compared to the MRI.



(a) (b)

Figure 6.2: Area function plot of the 'WHO'D' vowel for the MRI and AR methods

Another difference observed from the area function plots was that the AR method often yields a larger cross-sectional area for any given point along the vocal tract. From Figure 6.2(a), it can be seen that the point of the largest cross-sectional area is around $6.4cm^2$, while its AR counterpart in Figure 6.2(b) is slightly more than $7cm^2$. This is displayed in Figure 6.3, where the shape of the vocal tract from the two methods is similar but the data from the MRI is smaller than that of the AR.

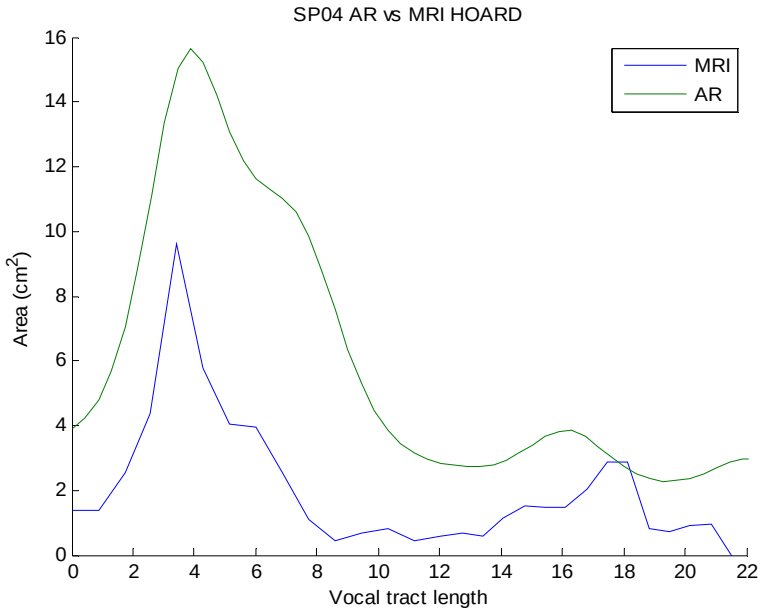


Figure 6.3: Area function plot of the vowel 'hoard' for SP04

As it can be seen from Figure 6.3, the area functions collected from the AR and MRI are reasonably similar in shape if not in magnitude. This was especially true for the back vowels, for which the structure of the tracts were not as affected by the mouth piece as significantly.

In Figures 6.4-6.9, the cross-sectional area obtained for each vowel from the 5 speakers with the AR and MRI methods are presented side by side. In Figure 6.4, it can be seen that for a number of vowels, the vocal tract shape obtained from the AR method were reasonably similar in to the shape obtained from the MRI method. This was shown in vowels such as 'HEED', 'WHOD', 'HOARD', 'HARD' and 'HOD', where the shape of the plot for the MRI is of a similar shape to that of the AR. For 'HID', 'HEAD', 'HERD' and 'HAD', the AR and MRI data did not always yield the same vocal tract shapes.

As suggested before, one of the main speculated reasons for the differences between the vocal tract contours obtained by the two methods is the structural compromise caused by the acoustic reflectometer mouth piece. A main aspect of difference between the vocal tract shapes obtained by the two methods is the size of the lip opening. For low vowels such as 'HARD', where the lip and jaw openings are expected to be relatively large, the plot of the MRI is clearly different to that of the AR. In Figures 6.4-6.9, it can be seen that the starting cross-sectional area for the MRI plot is large for 'HARD',

which is expected, while for the AR plot, even though the oral cavity is large as expected, the starting area is small.

One very interesting note of difference between the area function obtained from the two methods was that the MRI method seemed to yield a greater amount of detail for localised structures. For instance the distinctions between the areas of high or low volume were much more defined in the MRI results compared to that of the AR results. Looking at 'HOARD' in Figure 6.4, it can be seen that both the AR and MRI plots have a large oral cavity, and the area decreases towards the pharyngeal region. Followed by this, there is an increase in the area toward the end of the vocal tract, until it decreases again to the glottis. By comparing the two graphs, it can be seen that the increase in the area near the end of the pharyngeal region was not as pronounced in the AR data. This effect was observed for a number of vowels across all the speakers.

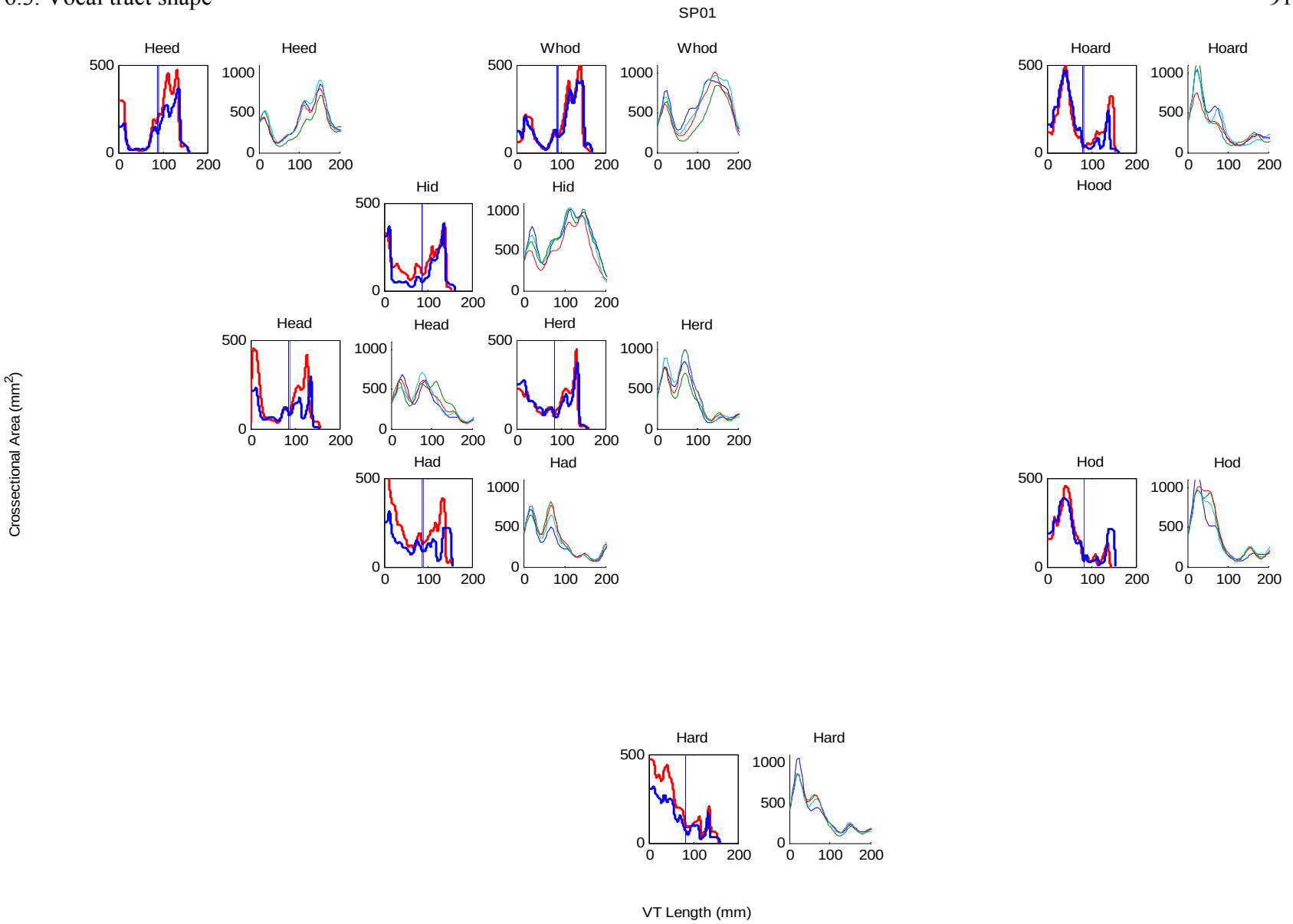


Figure 6.4: Cross-sectional area vs. vocal tract length plots for the AR and MRI methods SP01

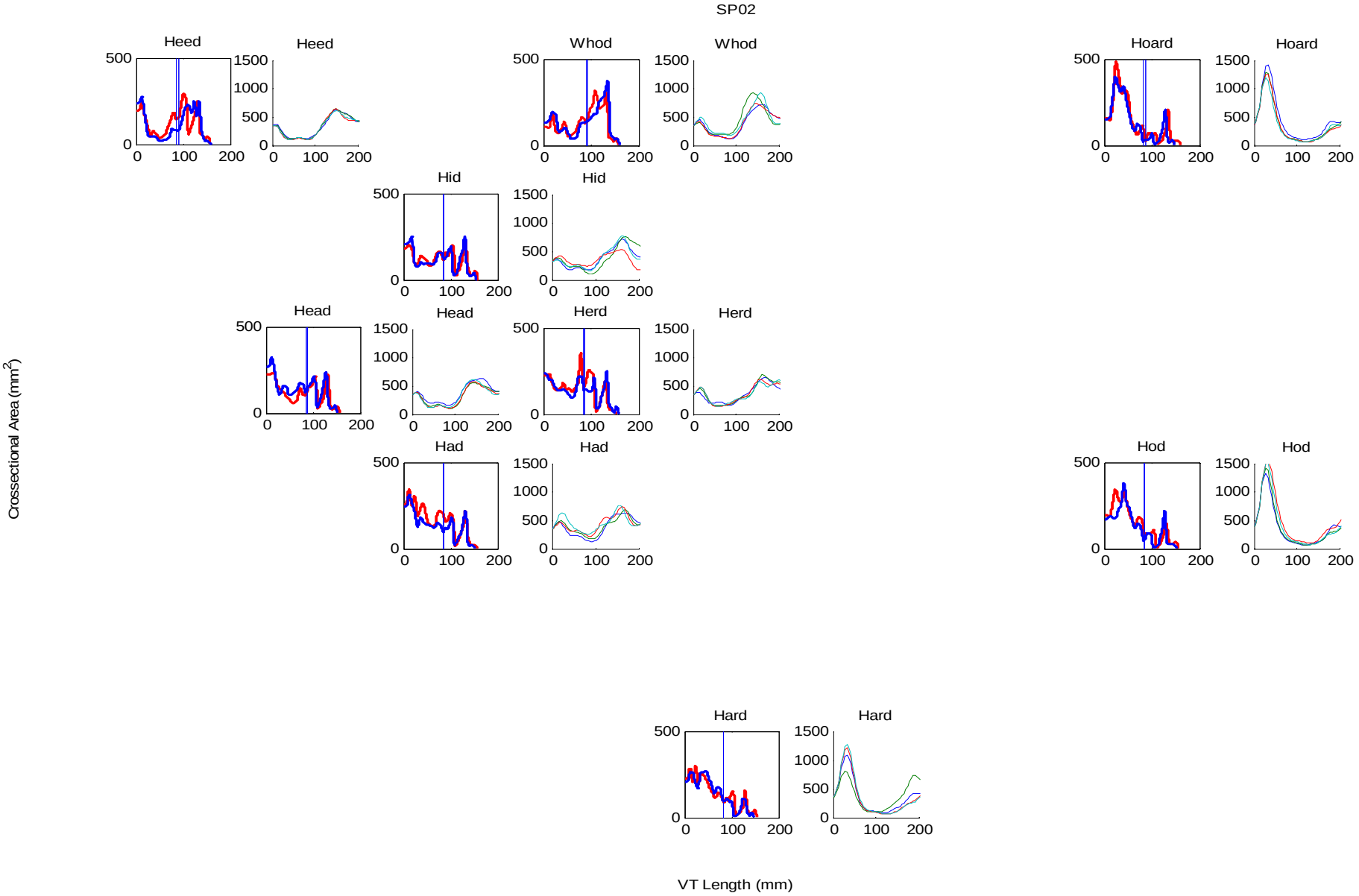


Figure 6.5: Cross-sectional area vs. vocal tract length plots for the AR and MRI methods SP02

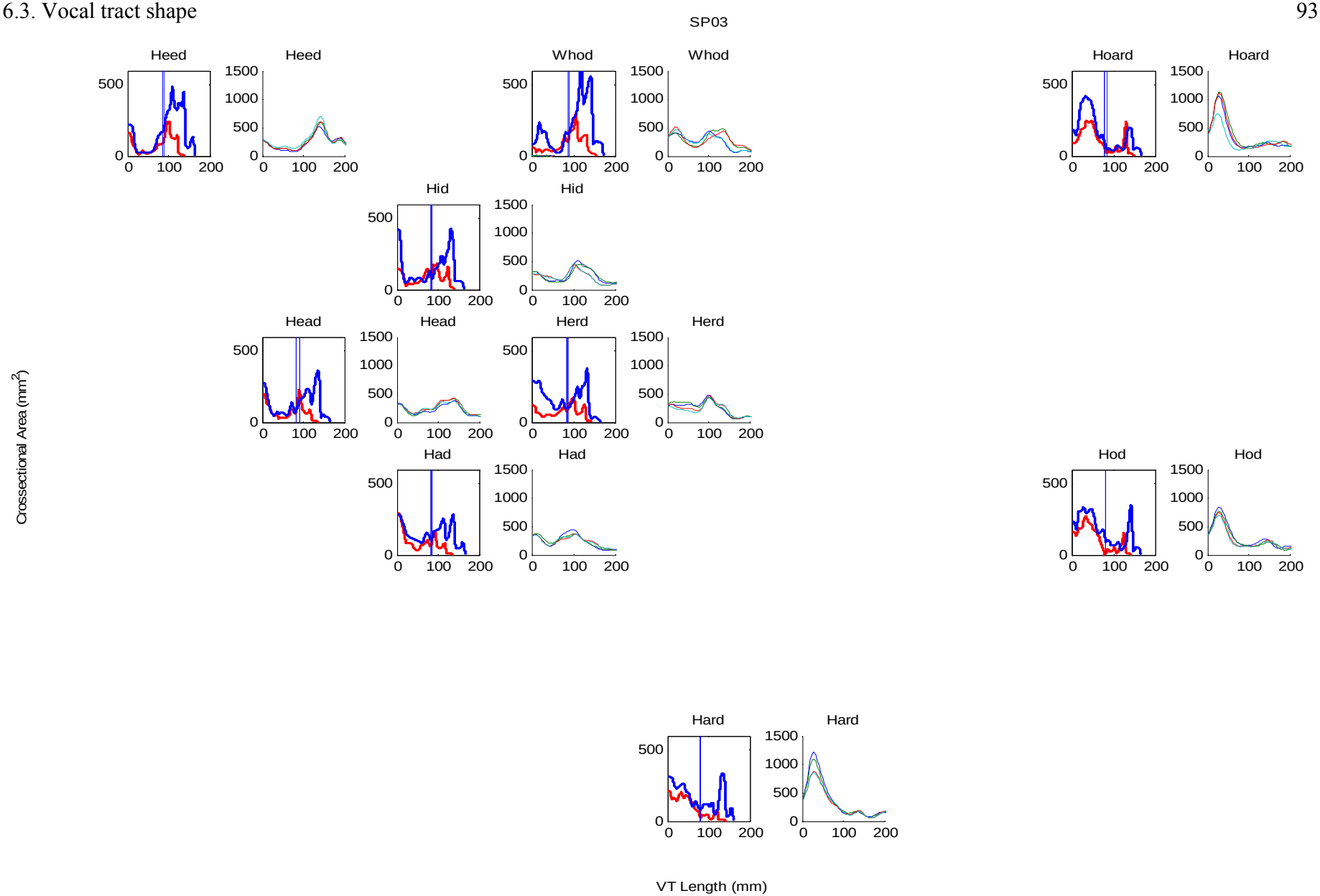


Figure 6.6: Cross-sectional area vs. vocal tract length plots for the AR and MRI methods SP03

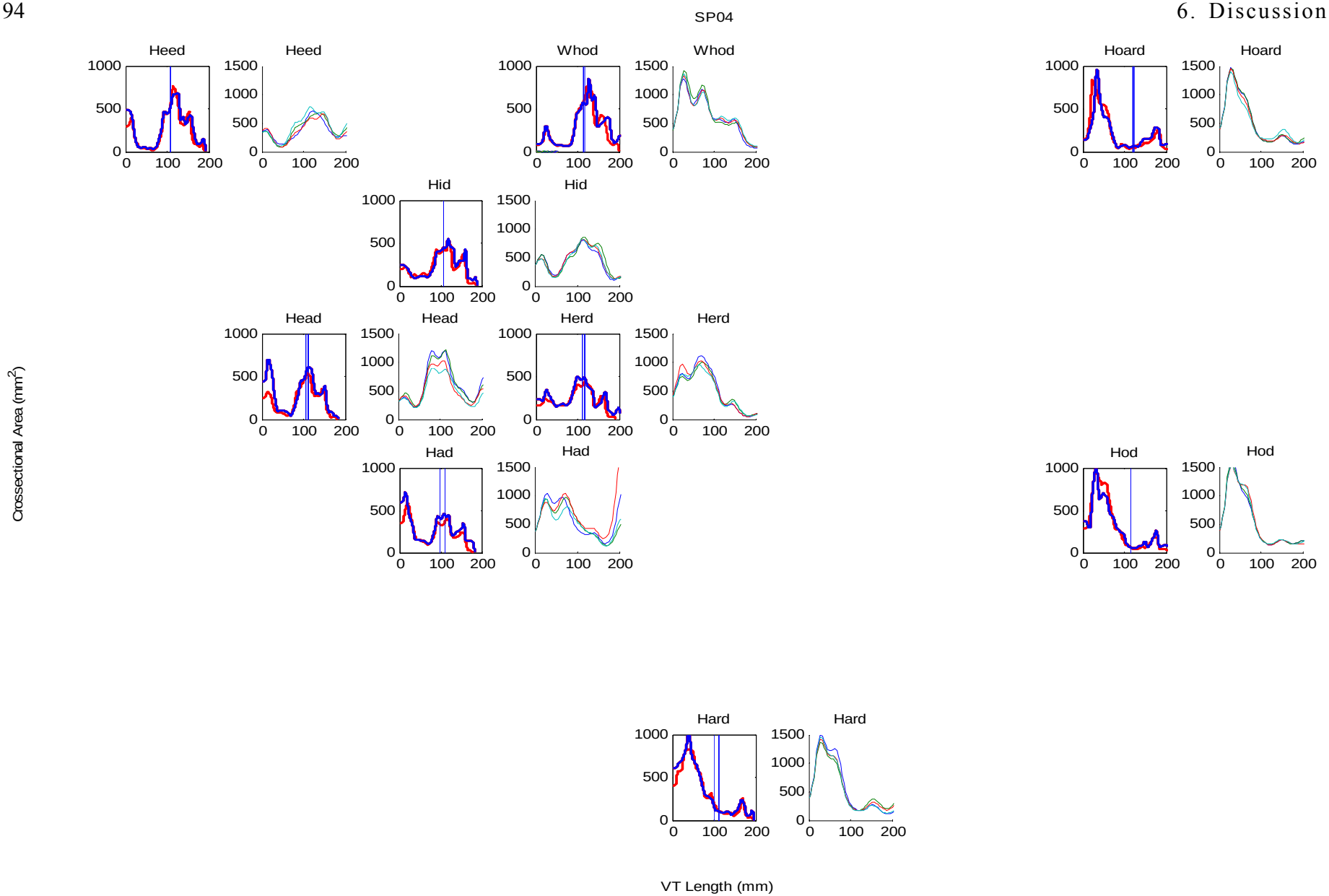


Figure 6.7: Cross-sectional area vs. vocal tract length plots for the AR and MRI methods SP04

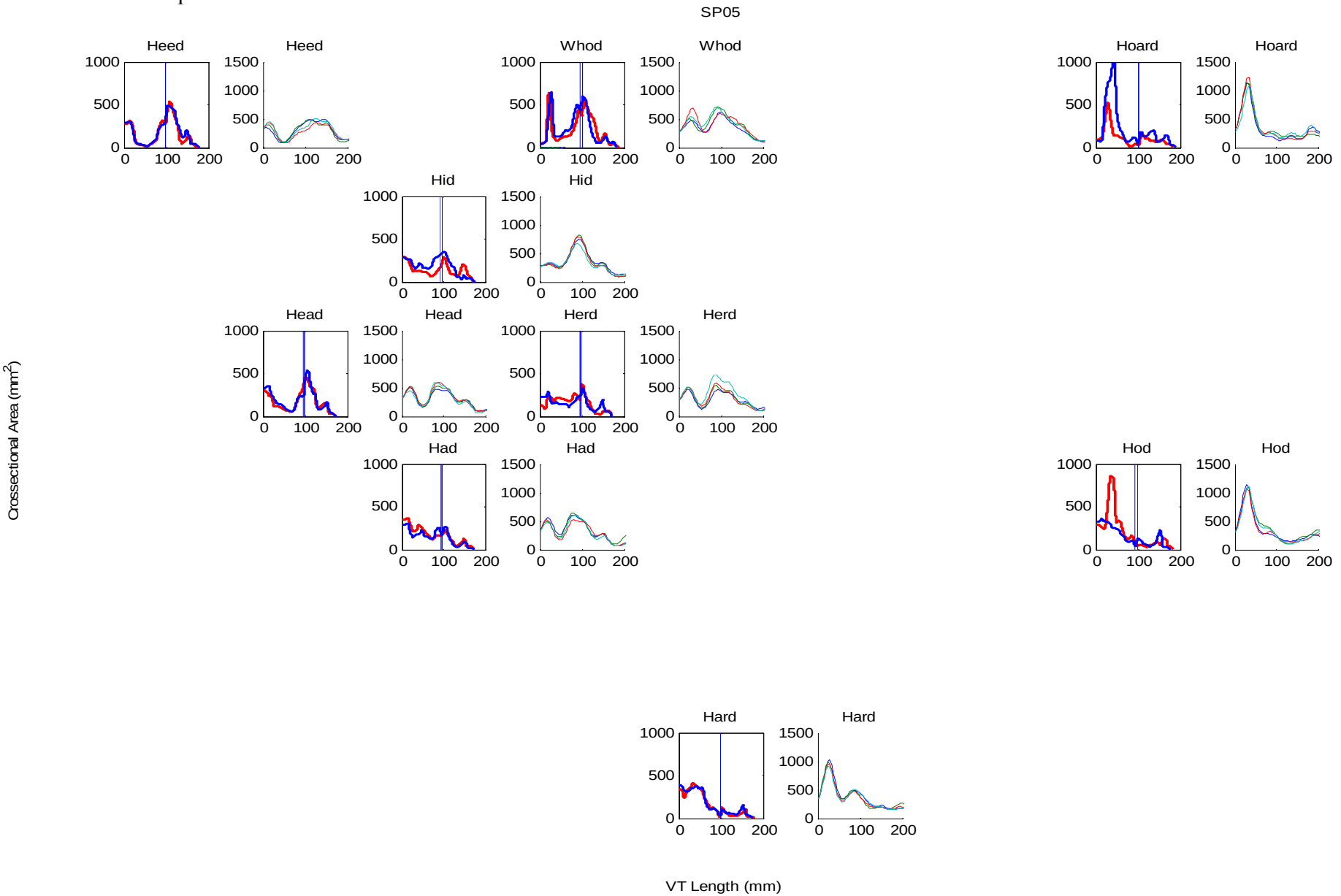


Figure 6.8: Cross-sectional area vs. vocal tract length plots for the AR and MRI methods SP05

6.4 Resonances vs. formants

Specific characteristics could be attributed to the resonances values collected from the two different methods. For the AR data, the resonances were found to be consistent across the different datasets for any given vowel of interest. This is due to the fact that the resonance can be calculated reliably once an accurate vocal tract shape has been obtained. For the acoustic reflectometery method, this is an advantage, as the area function obtained from the measurement can be visually confirmed on a computer while the measurement is taking place. As long as the operator of the equipment has a working knowledge of the expected shape of the tract for each vowel, the measurement can be repeated until the right vocal tract shape is acquired

This is not the case for the MRI method. While MRI give more structural detail in the images acquired from the scanning process, these images need to be processed before an area function is obtained. While the shape of the vocal tract can be verified by looking at the midsagittal image of the tract, marking discrepancies can cause a noticeable impact in the extracted area function which may subsequently affect calculation of the resonance values. Indeed, the differences of the resonance values between datasets for the MRI were larger than that of the AR method.

Comparing the resonances to the formants extracted from the vowel recordings, it was seen that the MRI's first and second resonances (R1 and R2) were more similar in values to the first and second formants (F1 and F2) than the AR's resonances were. As mentioned in the results chapters, it was found that the first resonances of both the AR and MRI methods were very close in value to the formants of recorded speech. The MRI's derived second resonances match the second formants well, while the AR's derived second resonances had noticeable lower values than that of the second formants' for many of the target vowels.

From the comparison of the resonances and the formants, it was found that the MRI method yielded resonances which were a reasonable estimate of the formant values for all the target vowels, except 'HOARD' and 'HOD'. The performance of the AR method was not as impressive, with only the three vowels 'HOD', 'HARD' and 'WHOD' providing a reason estimate of the formant values.

6.5 Acoustic Reflectometery vs. Magnetic Resonance Imaging

The aim of this study was to identify the strengths and weaknesses of two data acquisition techniques - acoustic reflectometery and magnetic resonance imaging - for the use of modelling the vocal tract. Both the methods have its merits and contribute very specifically to different aspect of the modelling process. In this section, the two different methods will be contrasted and discussed.

The first difference in the two measurement techniques is the accessibility. The MRI method requires a highly specialised machine which requires a trained technician to operate. Due to its high price tag, the MRI is not readily available, and often requires booking and expensive hiring fees. In comparison, the acoustic reflectometer is economic, and easy to operate.

The second aspect to consider is the convenience of the two methods. In this aspect, the acoustic reflectometer is preferable to the MRI for two reasons, the first being the duration of the measurement. While it takes 15 seconds to collect one set of data for the MRI, the AR requires only 3 seconds. Apart from this, the MRI requires the participant to lie on their backs in the interior of a large machine while the AR can be performed in an upright seated position. Having a more comfortable measuring position and shorter scanning time means the AR method can be used while causing less discomfort for the participants, making it a more preferable method.

Another the key aspect of difference separating the AR and MRI methods is the form of the data collected. For the AR, the results from the measurement are already in the form of cross-sectional areas while the MRI results yield images that need to be processed. As discussed previously, the cross-sectional area extraction method for the MR images is open to interpretation and can be a source of discrepancy between the area function of different data sets. In this aspect the AR's area function can be guaranteed to be similar between datasets, suggesting less variability in the calculated formant values. In contrast, the MRI method is likely to increase the variability of the formants between the datasets.

This being said, the results showed that the vocal tract area function extracted from the MR images were relatively consistent across the datasets for most of the speakers. The vocal tract shape is also comparable between the AR and MRI methods, with the AR being slightly higher in magnitude. From this stand point, both methods provided a reasonable process for reliably collecting the vocal tract shape.

Once the vocal tract cross-sectional area function was obtained, they were process and chamber resonances specific to the particular vocal tract orientation was calculated. For the two methods, the resonances were compared to the formants extracted from recorded speech, and it was found that the both methods yielded reasonable first formant values for the target vowel. In terms of the second formant, it was found that the AR method was less reliable. While the MRI produces similar second resonance values to the formants for most vowels, the AR ones were generally significantly lower.

One of the possible reasons for this difference was the mouth piece used in the AR data collection process. It was seen in the AR results chapter that both the shape of the vocal tract and the resulting resonance values were compromised due to the structural restrictions imposed on the lip region by the mouth piece, though the general trend of the vocal tract shape geometry was still captured

The interesting aspect of the result of the AR method was that while all the evidence pointed toward the mouth piece being the main reason for the distortion of the vocal tract, and subsequently the reason why the resonance values derived did not match the formants well, the mouth piece's restriction in mouth and jaw opening was supposed to have affected the first resonance and not the second. However, for the results, it was seen that the second resonance values were very different from the second formants while the difference between the first resonance and first formant was comparatively very small. For some reason, the second resonance was affected by the mouth piece instead of the first resonance

With all the compromising effects of the mouth piece, the resonances derived from the AR method for many of the vowels did not accurately reflect the formants of actual speech. It was found that the AR method was only accurate in estimating the formants of the 'HOD', 'HARD' and 'WHOD' vowels, while the MRI provided accurate estimations for all the vowels except 'HOARD' and 'HOD'

7 Conclusions and Future Work

7.1 Conclusion

In this study, the vocal tract shapes of the 9 monophthongs, obtained via the MRI and AR methods, were compared. This was followed by a comparison of the resonance values calculated from the vocal tract geometry determined by each method to the formant values extracted from recorded speech.

In order to process the cross-sectional areas extract from the MRI images to obtain the resonances, the Vocal Track Tool Mark II (VTTMII) was developed in MATLAB by the author as an iteration of the Vocal Track Tool Mark I (VTTMI) previously developed in the open source mathematical platform R. VTTMII was capable of loading the MRI cross-sectional data for analysis as well as the data obtained from acoustic reflectometery, while the VTTMI supported only the AR data.

Upon analysis it was found that the resonances for the AR did not match the formants from actual speech as well as the resonances obtained from the MRI, as the mouth piece used in the AR method distorted the second resonance values. It was concluded that the second resonance values from the AR method cannot be used as an accurate estimate to the formant values of real speech.

This being said, the acoustic reflectometery method has advantages such as a faster data acquisition time, and it is more comfortable for the participants to be in the upright seated position for the duration of the measurement opposed to lying position for the MRI. The output data of the AR was also in a more desirable form. As it is in cross-sectional areas, it does not require extra processing as with the MR images.

The AR method has demonstrated characteristic which could suggest a more convenient and time efficient method for the collection of vocal tract structural data, especially for studies which require vocal tract cross-sectional areas, while causing less discomfort for the participants. The time efficient nature of the method could increase the sample space of future studies, which could stretch the

applicability of the result to a wider population. Though the resonances values derived from the AR method was not an accurate estimation of the actual formant values for all the vowels of interest, it was concluded that, with future research into reducing the compromising aspect of the measurement process, the AR could potentially be an efficient and reliable method for collecting structural vocal tract data for the modelling of the vocal tract.

7.2 Future Work

The next step in this topic of research would be to perform statistical analysis on the data presented in this study. So far, for the purposes of this study, the analyses of the results have been confined to the qualitative aspects. Observations being made on the various aspects of the vocal tract such as the shape and length are based solely on a visual interpretation of the graphs in which the results are presented. By performing statistical analysis on the results, it is possible to observe the significance of the differences in the resonance and formant values, and thus have a more definitive conclusion on the AR's applicability in accurately modelling the vocal tract.

Another direction of interest would be investigating how the mouth piece compromises the vocal tract shape and methods of reducing these effects. It was seen that not all vowels which were expected to be affected by the mouth piece to a large degree were presented with a compromised second resonance value. The AR results for the low vowel HARD, which was expected to have a compromised second resonance value due to the restricted jaw opening during its vocalisation, were presented with resonances which were reasonable estimates of the formant values deduced from recorded speech. By further investigating how the mouth piece affects the articulation of the target vowels, it may be possible to remove the compromising effect of the mouth piece.

Appendix A - Tabled results

A.1. AR vocal tract length

SP01

	Scan 1	Scan 2	Scan 3	Scan 4	Average
HAD	18.45	18.02	18.45	18.45	18.34
HARD	18.88	18.88	18.88	19.41	19.01
HEAD	19.74	19.31	19.74	19.74	19.63
HEED	21.02	19.74	20.17	21.02	20.49
HERD	18.88	18.88	18.88	19.74	19.10
HID	22.50	22.50	22.00	22.50	22.38
HOARD	20.50	20.60	20.60	20.60	20.58
HOD	19.31	19.31	19.31	19.31	19.31
WHO'D	23.00	22.00	23.00	23.00	22.75

Table A. 1: Vocal tract length for SP01 (AR)

SP02	Scan 1	Scan 2	Scan 3	Scan 4	Average
HAD	21.02	20.6	20.17	19.74	20.38
HARD	20.00	17.00	20.00	20.00	19.25
HEAD	21.02	21.02	21.02	20.60	20.92
HEED	21.02	21.50	20.40	20.60	20.88
HERD	20.50	19.50	18.88	18.45	19.33
HID	20.50	19.60	21.02	21.02	20.54
HOARD	20.50	20.00	20.00	21.00	20.38
HOD	21.00	20.00	19.50	19.50	20.00
WHO'D	21.02	20.17	21.02	20.60	20.70

Table A. 2: Vocal tract length for SP02 (AR)

SP03	Scan 1	Scan 2	Scan 3	Scan 4	Average
HAD	18.45	18.45	18.45	18.45	18.45
HARD	17.17	17.17	17.59	17.59	17.38
HEAD	18.70	19.31	19.00	19.00	19.00
HEED	18.00	18.20	18.45	18.20	18.21
HERD	17.59	17.59	18.02	18.02	17.81
HID	18.88	20.17	18.02	18.45	18.88
HOARD	18.00	17.10	16.31	20.17	17.90
HOD	18.45	20.17	19.74	20.17	19.63
WHO'D	17.59	18.00	18.00	17.59	17.80

Table A. 3: Vocal tract length for SP03 (AR)

SP04	Scan 1	Scan 2	Scan 3	Scan 4	Average
HAD	17.59	18.02	17.17	16.59	17.34
HARD	19.74	19.74	19.74	19.74	19.74
HEAD	18.88	18.88	18.45	18.88	18.77
HEED	19.74	19.31	19.31	18.88	19.31
HERD	19.31	19.31	18.88	19.74	19.31
HID	19.74	20.17	19.74	19.74	19.85
HOARD	19.31	19.31	20.17	19.74	19.63
HOD	18.45	18.45	19.31	18.88	18.77
WHO'D	21.02	21.02	21.02	21.02	21.02

Table A. 4: Vocal tract length for SP04 (AR)

SP05	Scan 1	Scan 2	Scan 3	Scan 4	Average
HAD	19.31	18.88	19.31	19.31	19.20
HARD	18.88	18.02	18.02	18.45	18.34
HEAD	19.74	19.74	19.74	19.31	19.63
HEED	20.17	20.17	20.17	20.6	20.28
HERD	19.74	20.17	20.17	20.17	20.06
HID	19.74	19.74	19.31	19.31	19.53
HOARD	20.50	20.00	20.00	21.00	20.35
HOD	18.00	18.50	18.00	18.50	18.25
WHO'D	21.02	20.70	21.02	20.17	20.73

Table A. 5: Vocal tract length for SP05 (AR)

A.2. MRI vocal tract length

SP01	HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	HOOD	HUD	WHOD
Scan1	15.5	16.1	15.3	16.3	15.9	16.3	16.1	15.4	16.5	16.2	17.1
Scan2	15.7	15.7	16.6	15.8	16.4	15.4	15.3	14.7	15.5	15.3	16.7

Table A. 6: Vocal tract length for SP01 (MRI)

SP02	HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	HOOD	HUD	WHOD
Scan1	15.2	14.8	15.7	15.9	15.9	15.5	14.7	15.2	15.9	15.3	15.9
Scan2	15.7	15.4	16.1	15.8	15.6	15.8	16.0	15.6	15.8	15.2	15.7

Table A. 7: Vocal tract length for SP02 (MRI)

SP03	HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	HOOD	HUD	WHOD
Scan1	14.0	13.8	15.0	14.1	15.6	14.8	14.8	14.5	15.5	14.1	15.5
Scan2	13.7	14.3	13.8	14.1	14.5	14.2	15.0	14.3	15.2	14.1	15.5

Table A. 8: Vocal tract length for SP03 (MRI)

SP04	HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	HOOD	HUD	WHOD
Scan1	18.4	19.7	19.1	19.4	20.5	19.1	21.5	21.2	21.7	20.2	21.0
Scan2	18.0	19.6	17.9	18.9	19.3	18.8	20.9	20.5	21.7	19.9	20.1

Table A. 9: Vocal tract length for SP04 (MRI)

SP05	HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	HOOD	HUD	WHOD
Scan1	18.2	18.1	19.3	18.9	18.2	18.5	19.2	19.4	19.0	19.3	19.6
Scan2	18.5	18.5	18.4	18.7	18.1	18.3	20.0	18.7	19.2	18.9	19.1

Table A. 10: Vocal tract length for SP05 (MRI)

A.3. Derived AR resonances

SP01		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D
Set 1	F1	643	619	546	304	609	349	541	634	333
	F2	1376	1209	1369	1538	1170	1339	975	1021	1286
	F3	2217	2080	1998	2102	2086	1924	2092	2137	1842
Set 2	F1	628	619	482	315	609	349	532	605	307
	F2	1429	1122	1532	1628	1267	1339	1046	1002	1361
	F3	2382	2225	2204	2100	2157	1988	2128	2374	1884

Set 3	F1	614	619	527	326	571	349	532	605	310
	F2	1386	1190	1425	1575	1296	1436	1064	1012	1389
	F3	2227	2215	2109	2173	2176	2067	2074	2374	1866
Set 4	F1	633	596	518	313	574	357	585	596	310
	F2	1356	1192	1351	1555	1129	1372	1028	1012	1326
	F3	2197	2166	2128	2042	2091	1997	1924	2374	1834

Table A. 11: Derived resonance values for all target vowels SP01 (AR)

SP02		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D
Set 1	F1	313	409	321	287	330	312	463	417	295
	F2	1347	957	1407	1512	1363	1407	974	957	1433
	F3	2085	2017	2076	2120	2120	2147	2009	2000	2068
Set 2	F1	355	591	304	272	347	326	493	511	308
	F2	1277	1139	1451	1512	1423	1332	959	940	1530
	F3	2092	2567	2102	2056	2098	2283	2100	2073	2218
Set 3	F1	362	502	295	286	368	382	502	534	287
	F2	1331	950	1503	1576	1480	1355	940	899	1477
	F3	2128	2100	2111	2220	2128	2076	2091	2126	2068
Set 4	F1	389	520	301	284	396	321	417	596	319
	F2	1286	922	1534	1551	1435	1364	974	942	1392
	F3	2137	2100	2172	2172	2128	2120	2009	2163	2163

Table A. 12: Derived resonance values for all target vowels SP02 (AR)

SP03		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D
Set 1	F1	485	713	430	337	529	406	599	584	540
	F2	1712	1255	1582	1673	1692	1702	1157	1168	1630
	F3	2306	2521	2393	2514	2544	2341	2354	2455	2533
Set 2	F1	525	713	388	346	550	380	641	561	436
	F2	1613	1298	1617	1613	1630	1630	1153	1168	1573
	F3	2286	2499	2402	2534	2544	2146	2584	2182	2415
Set 3	F1	485	696	413	346	486	476	694	527	457
	F2	1603	1277	1605	1584	1672	1682	1198	1193	1481
	F3	2326	2461	2412	2554	2473	2422	2665	2202	2364
Set 4	F1	485	696	423	356	476	475	453	516	550
	F2	1594	1287	1576	1574	1692	1673	1240	1213	1588
	F3	2326	2440	2412	2534	2452	2366	2055	2128	2481

Table A. 13: Derived resonance values for all target vowels SP03 (AR)

SP04		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D
Set 1	F1	561	564	339	342	530	389	586	584	495
	F2	1350	1036	1615	1665	1494	1619	1040	1109	1225
	F3	2388	2285	2312	2202	2070	2183	2317	2326	2155
Set 2	F1	537	537	348	312	530	353	577	584	487
	F2	1469	971	1635	1655	1390	1584	1031	1109	1164
	F3	2300	2285	2225	2345	2260	2119	2326	2336	2120
Set 3	F1	521	546	376	322	551	361	561	586	487
	F2	1457	981	1683	1674	1461	1628	1005	1031	1199
	F3	2382	2285	2286	2251	2263	2220	2218	2289	2102
Set	F1	561	574	368	319	537	352	546	590	478

4	F2	1485	1018	1635	1683	1323	1647	1036	1054	1190
	F3	2315	2294	2321	2360	2220	2211	2248	2350	2094

Table A. 14: Derived resonance values for all target vowels SP04 (AR)

SP05		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D
Set 1	F1	511	561	463	335	435	416	450	639	434
	F2	1485	1296	1480	1621	1499	1471	1011	1207	1373
	F3	2166	1993	2054	2245	1961	2174	1801	2161	1946
Set 2	F1	493	608	463	371	444	416	513	592	426
	F2	1490	1348	1499	1612	1476	1508	1052	1076	1427
	F3	2196	2047	2026	2173	1938	2220	1826	2122	2030
Set 3	F1	473	588	463	362	453	426	504	639	434
	F2	1560	1368	1523	1612	1449	1532	974	1248	1321
	F3	2137	2067	2081	2073	1947	2222	1817	2110	1929
Set 4	F1	492	554	454	355	426	454	432	612	444
	F2	1504	1366	1542	1578	1494	1456	979	1145	1403
	F3	2137	2069	2118	2048	2037	2196	1793	2103	2082

Table A. 15: Derived resonance values for all target vowels SP05 (AR)

A.4. Derived MRI resonances

SP01		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D	HUD	HOOD
Set 1	F1	564	803	537	281	564	391	691	604	341	701	477
	F2	1715	1629	2076	2348	1829	2149	1178	1183	1784	1673	1308
	F3	2690	2749	3043	3269	2945	3157	3036	2828	2586	2894	2937

Set 2	F1	605	815	429	242	524	499	525	796	327	586	447
	F2	1967	1606	2255	2577	1898	1947	1133	1282	1548	1602	1305
	F3	2955	2851	3002	3602	2980	2985	2995	3136	2639	3145	3009

Table A. 16: Derived resonance values for all target vowels SP01 (MRI)

SP02		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D	HUD	HOOD
Set 1	F1	734	861	653	402	643	623	783	783	445	741	619
	F2	1830	1624	1947	2425	1837	1905	1367	1433	1952	1577	1558
	F3	2877	2829	2938	3138	2813	2916	2907	2951	2774	2831	2738
Set 2	F1	731	817	601	475	658	613	763	820	465	803	661
	F2	1858	1717	1972	2187	1985	1873	1366	1571	1975	1630	1484
	F3	2717	2699	2709	2951	2830	2810	2630	2707	2928	2817	2689

Table A. 17: Derived resonance values for all target vowels SP02 (MRI)

SP03		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D	HUD	HOOD
Set 1	F1	715	952	500	439	493	504	764	879	413	904	600
	F2	2301	1984	2501	2661	2183	2374	1491	1700	2208	2002	1400
	F3	3341	3320	3196	3565	3028	3186	3252	3266	2976	3294	3142
Set 2	F1	749	969	593	426	555	528	696	840	400	945	683
	F2	2593	1747	2664	2700	2219	2317	1282	1426	2166	1760	1295
	F3	3448	3354	3481	3695	3177	3348	3151	3336	2990	3209	3369

Table A. 18: Derived resonance values for all target vowels SP03 (MRI)

SP04		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D	HUD	HOOD
Set 1	F1	467	666	383	254	418	363	424	576	270	589	378
	F2	1780	1175	1970	1931	1502	1779	959	1006	1577	1105	1059
	F3	2406	2313	2410	2788	2063	2429	2155	2209	1874	2336	2067
Set 2	F1	507	698	357	241	425	399	473	613	273	635	371
	F2	1855	1145	2120	2100	1653	1780	1006	969	1639	1094	978
	F3	2342	2243	2528	2736	2258	2481	2197	2311	2058	2299	2193

Table A. 19: Derived resonance values for all target vowels SP04 (MRI)

SP05		HAD	HARD	HEAD	HEED	HERD	HID	HOARD	HOD	WHO'D	HUD	HOOD
Set 1	F1	699	796	419	307	573	533	531	674	394	750	580
	F2	1663	1351	1975	2045	1585	1661	1242	1160	1595	1341	1099
	F3	2352	2369	2374	2813	2149	2371	2244	2380	2086	2351	2187
Set 2	F1	657	740	472	310	588	590	500	679	403	622	503
	F2	1670	1387	2046	2245	1657	1896	952	1451	1452	1386	1167
	F3	2359	2502	2403	2866	2362	2465	2163	2378	1794	2374	2012

Table A. 20: Derived resonance values for all target vowels SP05 (MRI)

Appendix B - Graphical results

B.1. AR data - R1 vs. R2 plot

SP01

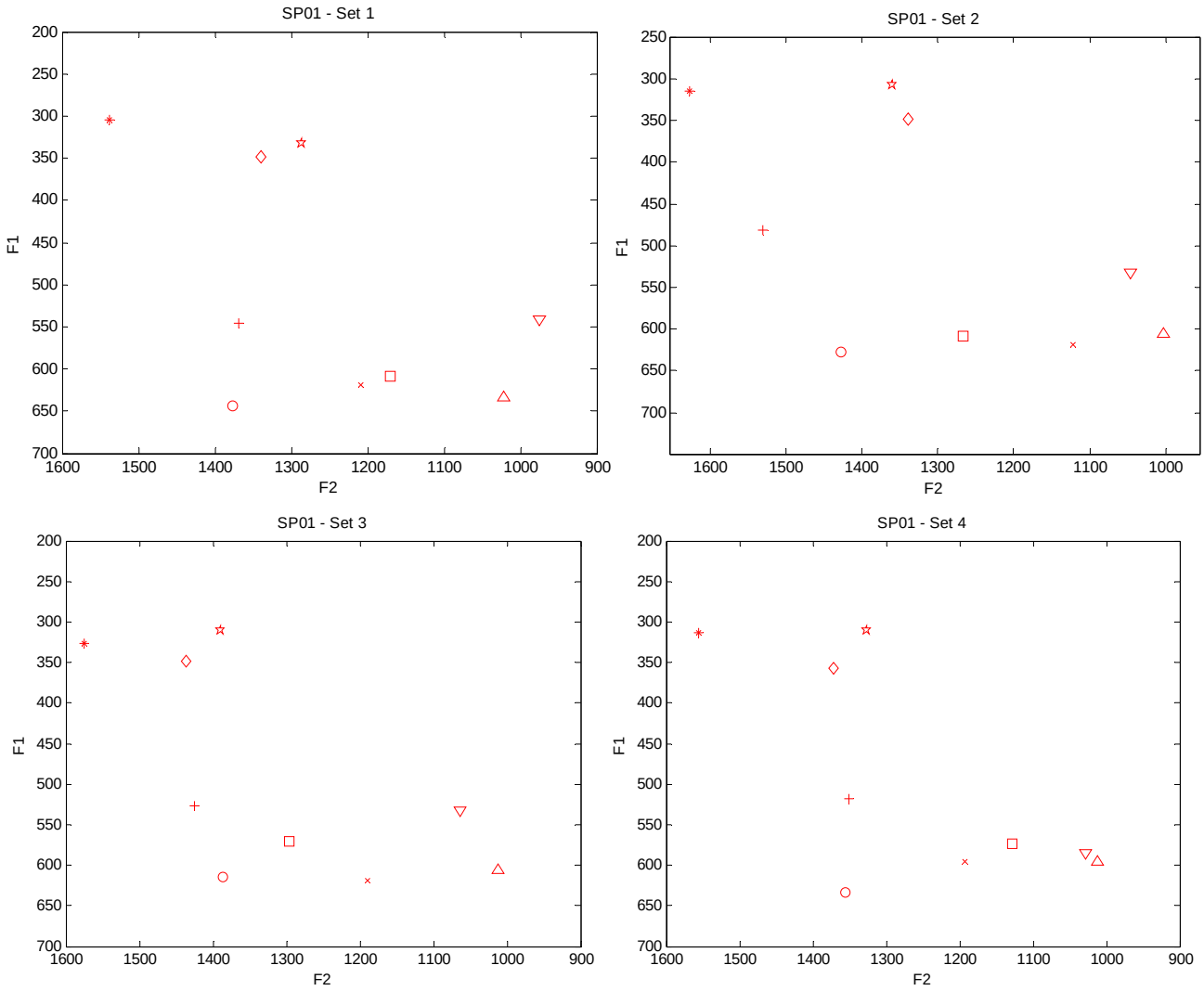


Figure B. 1: First resonance vs. second resonance plot of all vowels collected with AR method (SP01)

SP02

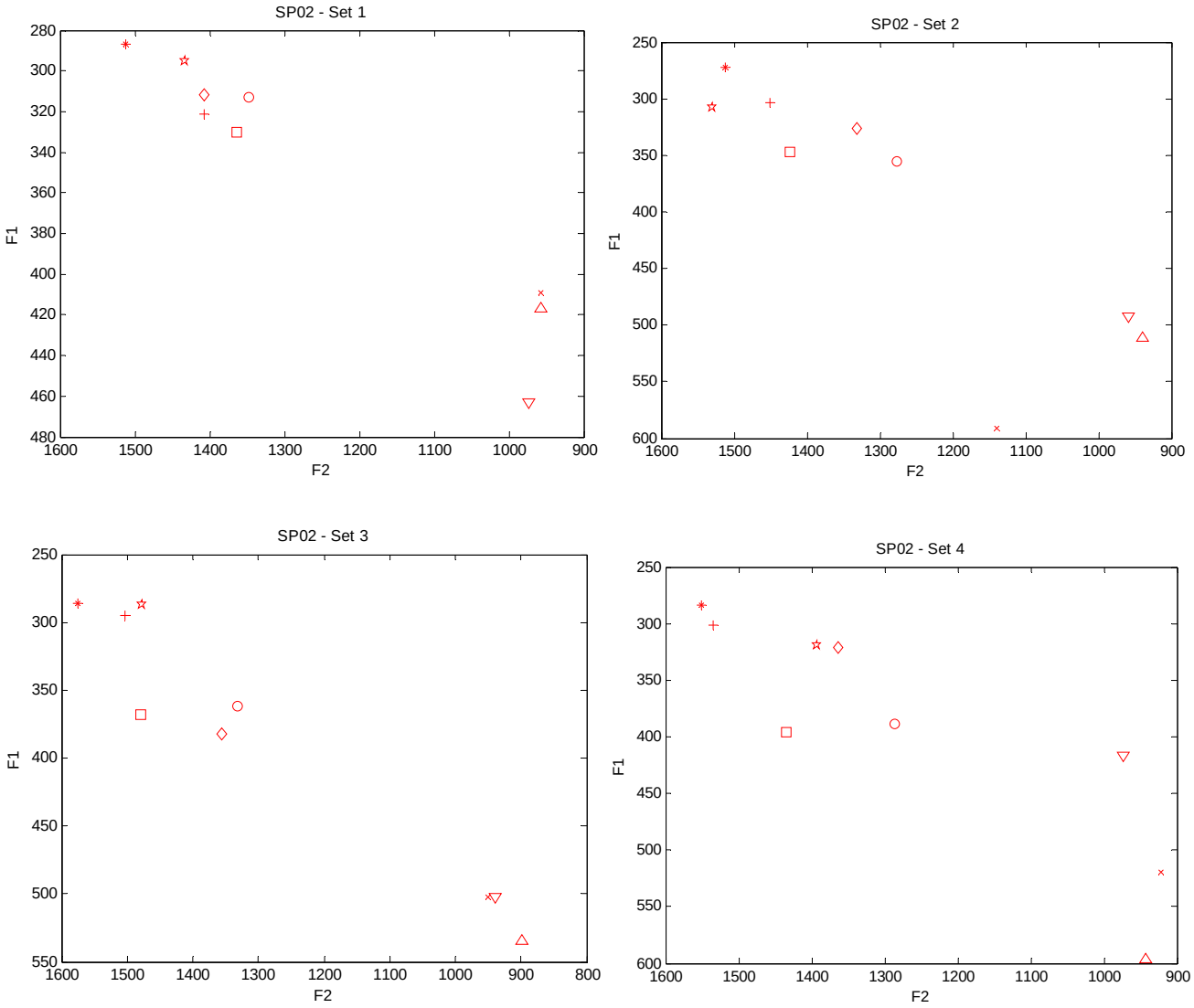


Figure B. 2: First resonance vs. second resonance plot of all vowels collected with AR method (SP02)

SP03

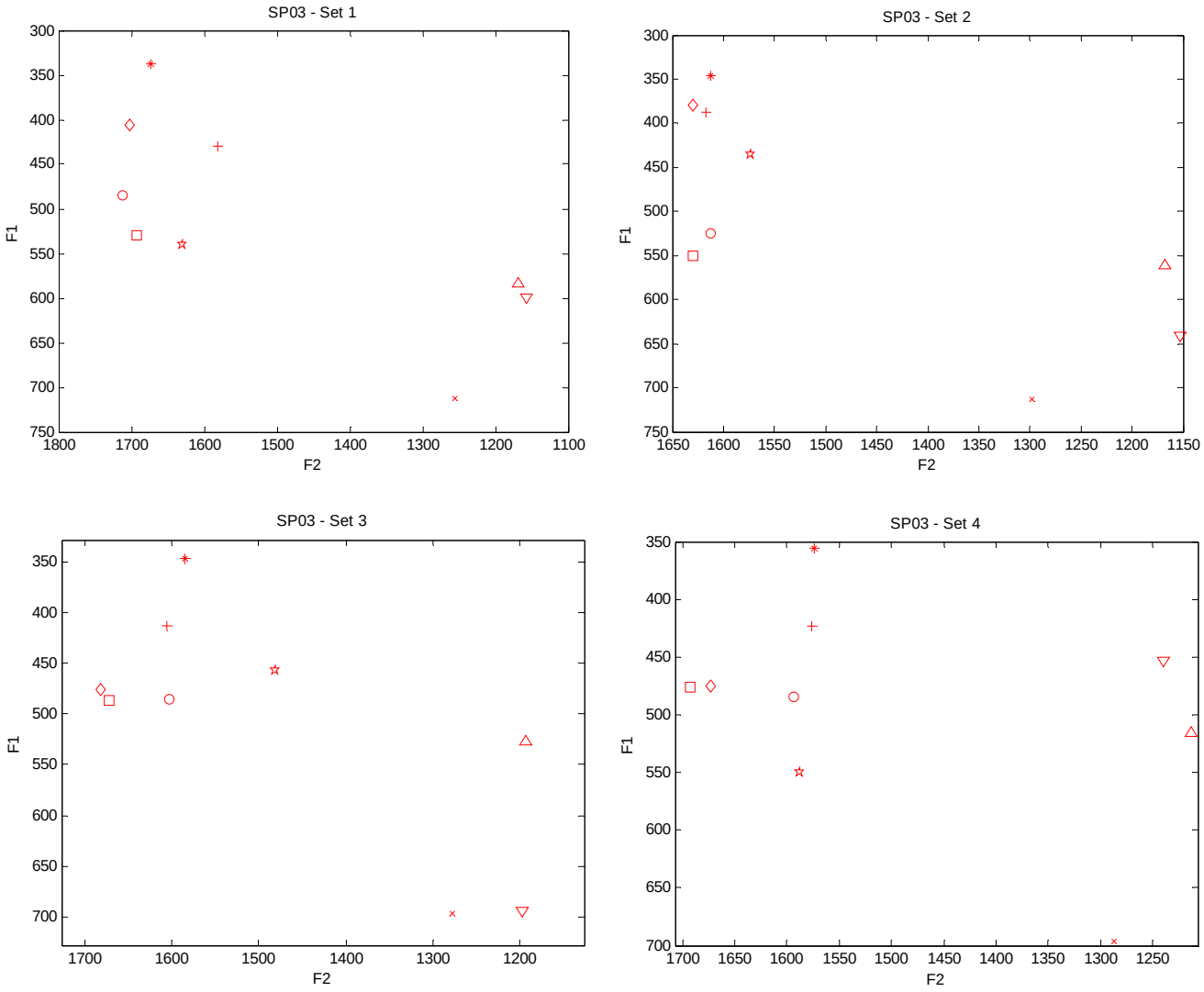


Figure B. 3: First resonance vs. second resonance plot of all vowels collected with AR method (SP03)

SP04

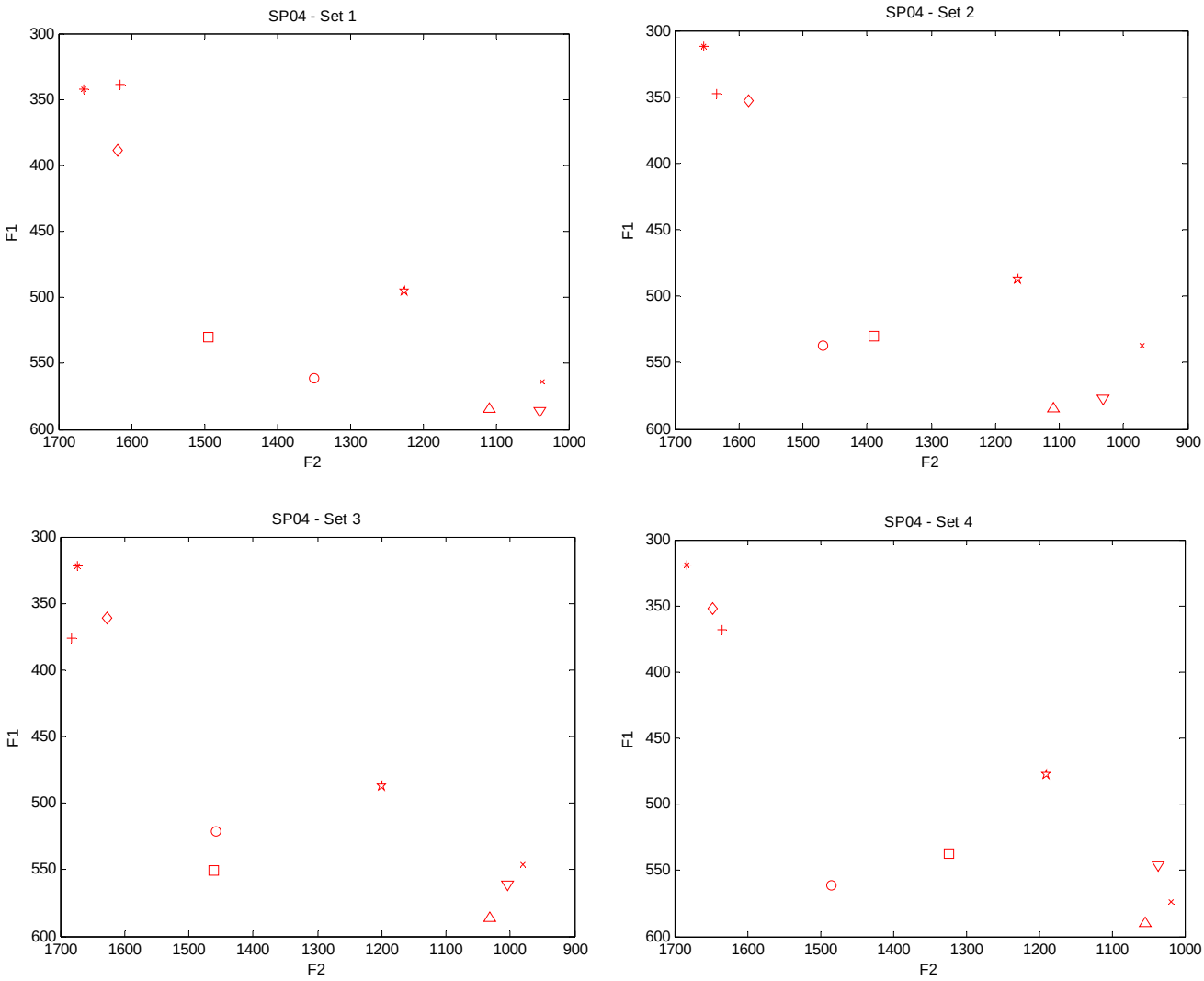


Figure B. 4: First resonance vs. second resonance plot of all vowels collected with AR method (SP04)

SP05

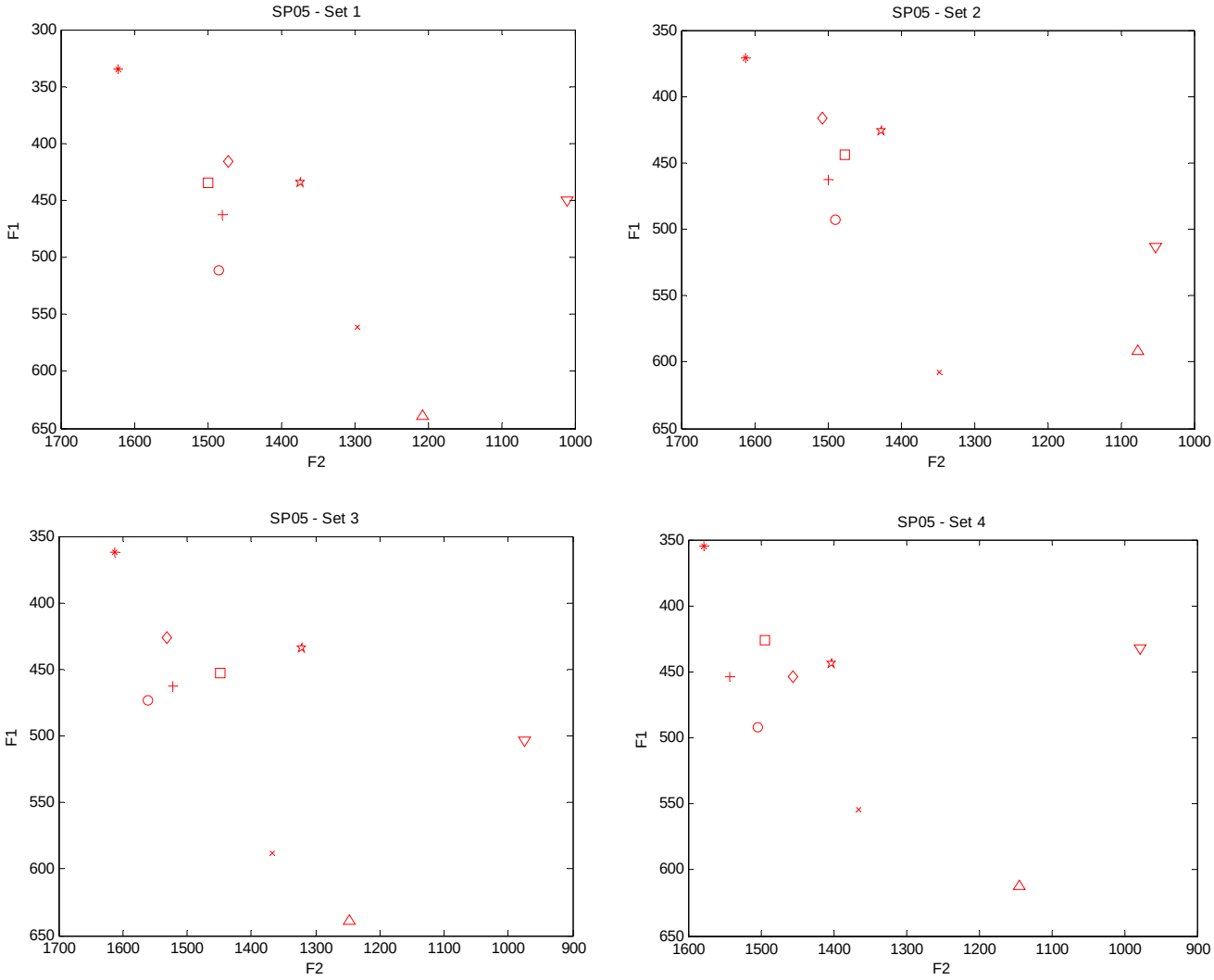


Figure B. 5: First resonance vs. second resonance plot of all vowels collected with AR method (SP05)

B.2. MRI data - R1 vs. R2 plot

SP01

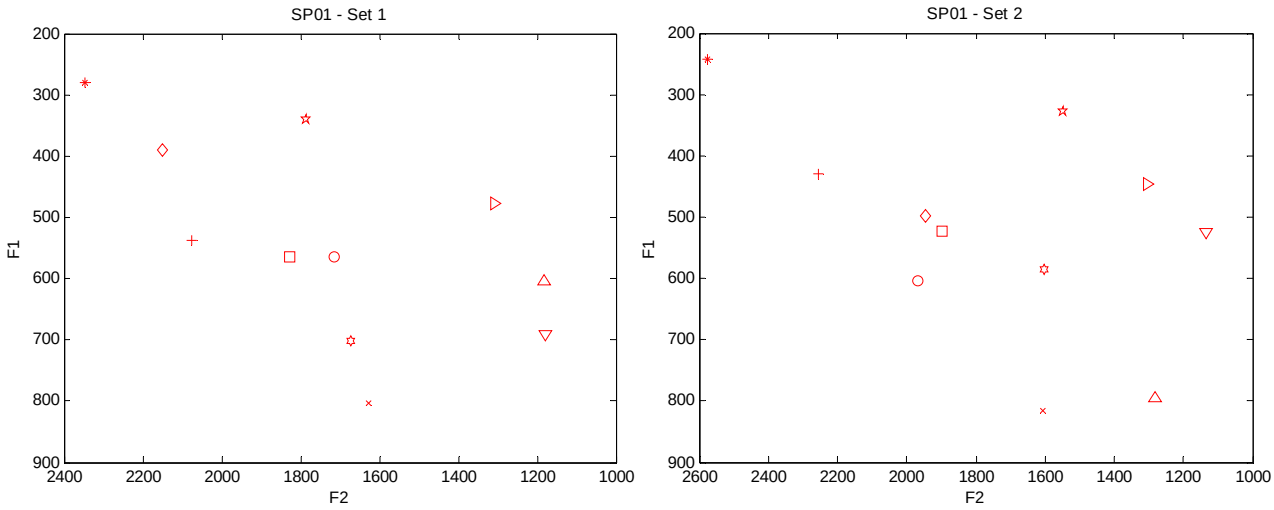


Figure B. 6: First resonance vs. second resonance plot of all vowels collected with MRI method (SP01)

SP02

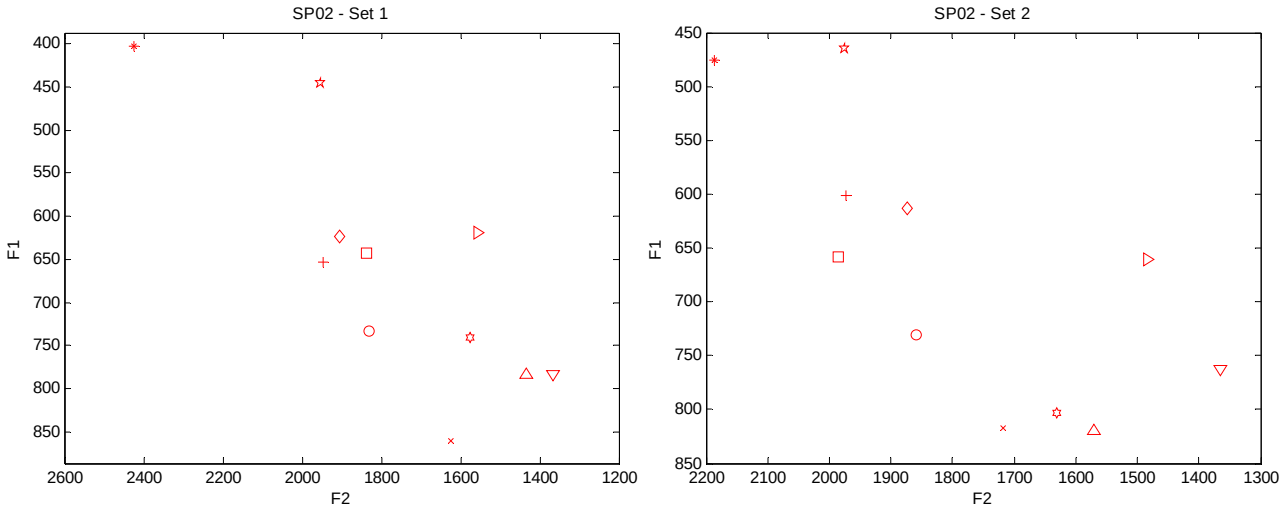


Figure B. 7: First resonance vs. second resonance plot of all vowels collected with MRI method (SP02)

SP03

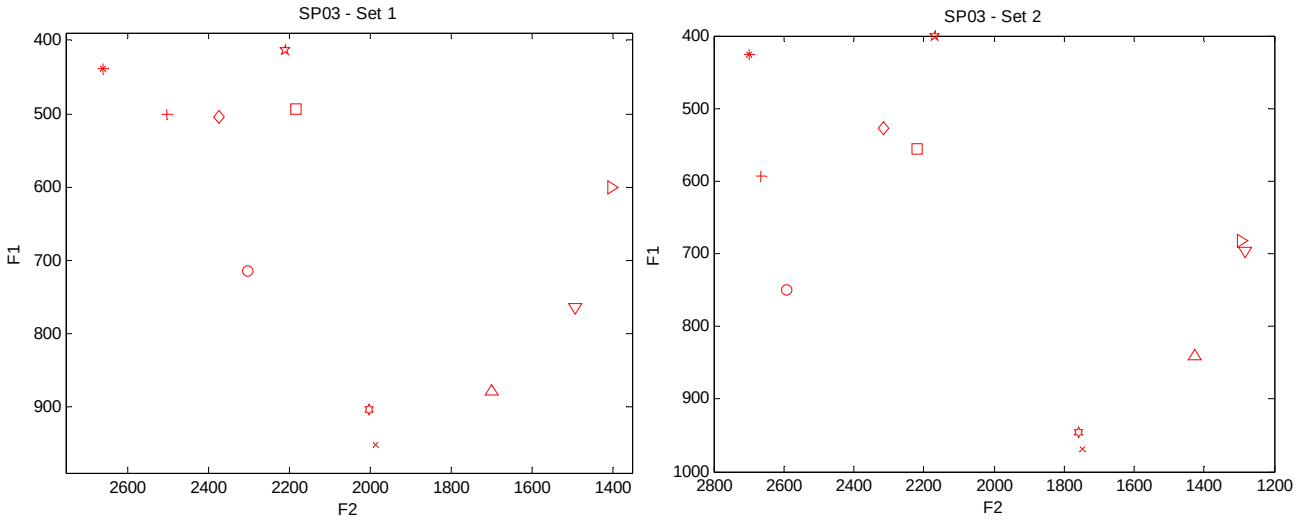


Figure B. 8: First resonance vs. second resonance plot of all vowels collected with MRI method (SP03)

SP04

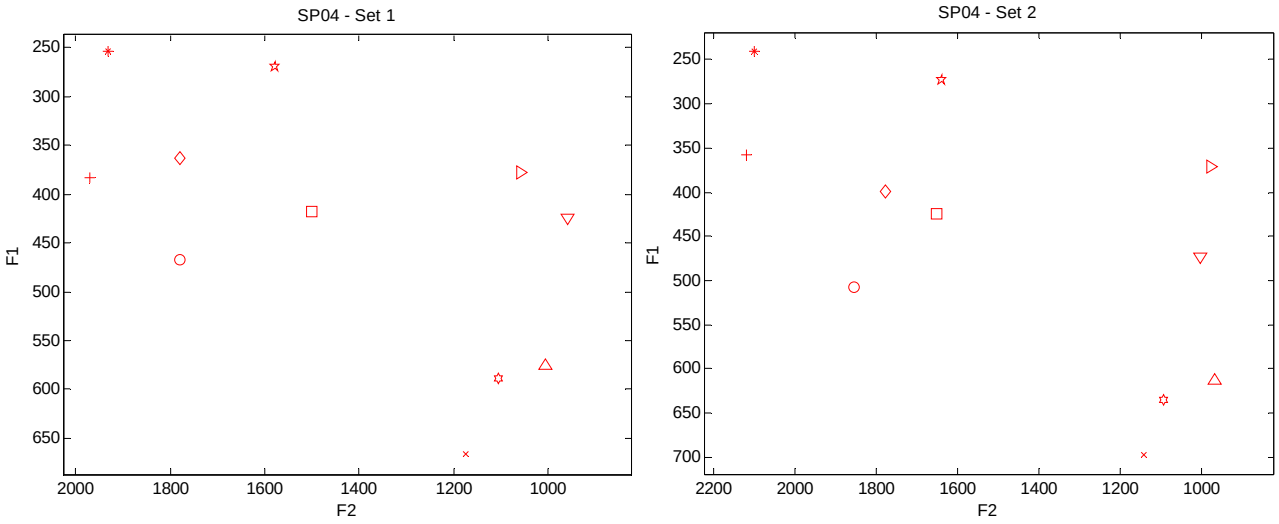


Figure B. 9: First resonance vs. second resonance plot of all vowels collected with MRI method (SP04)

SP05

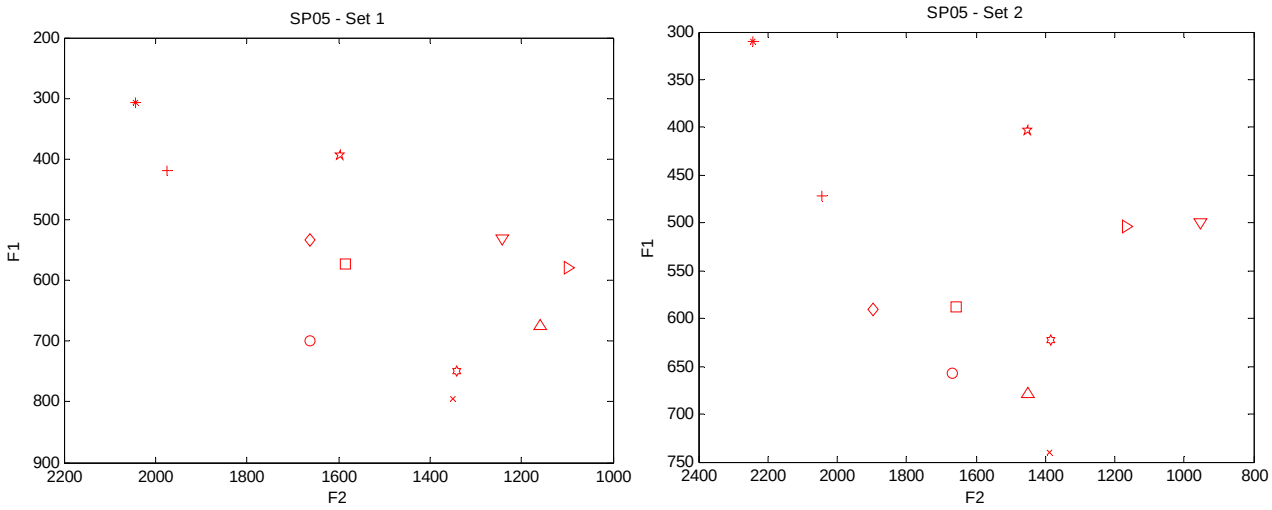


Figure B. 10: First resonance vs. second resonance plot of all vowels collected with MRI method (SP05)

B.3. AR resonance vs. Formants

SP01

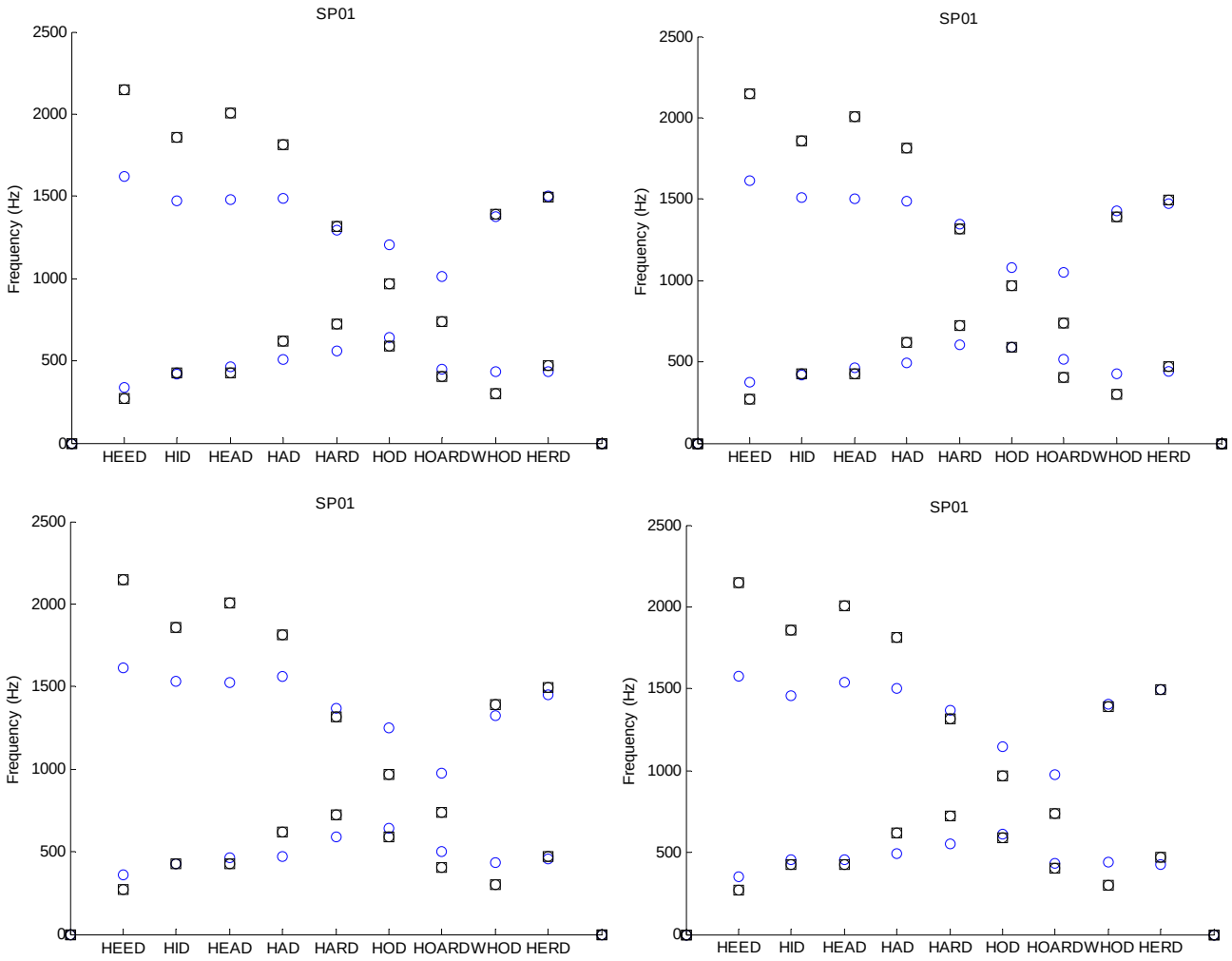


Figure B. 11: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP01)

SP02

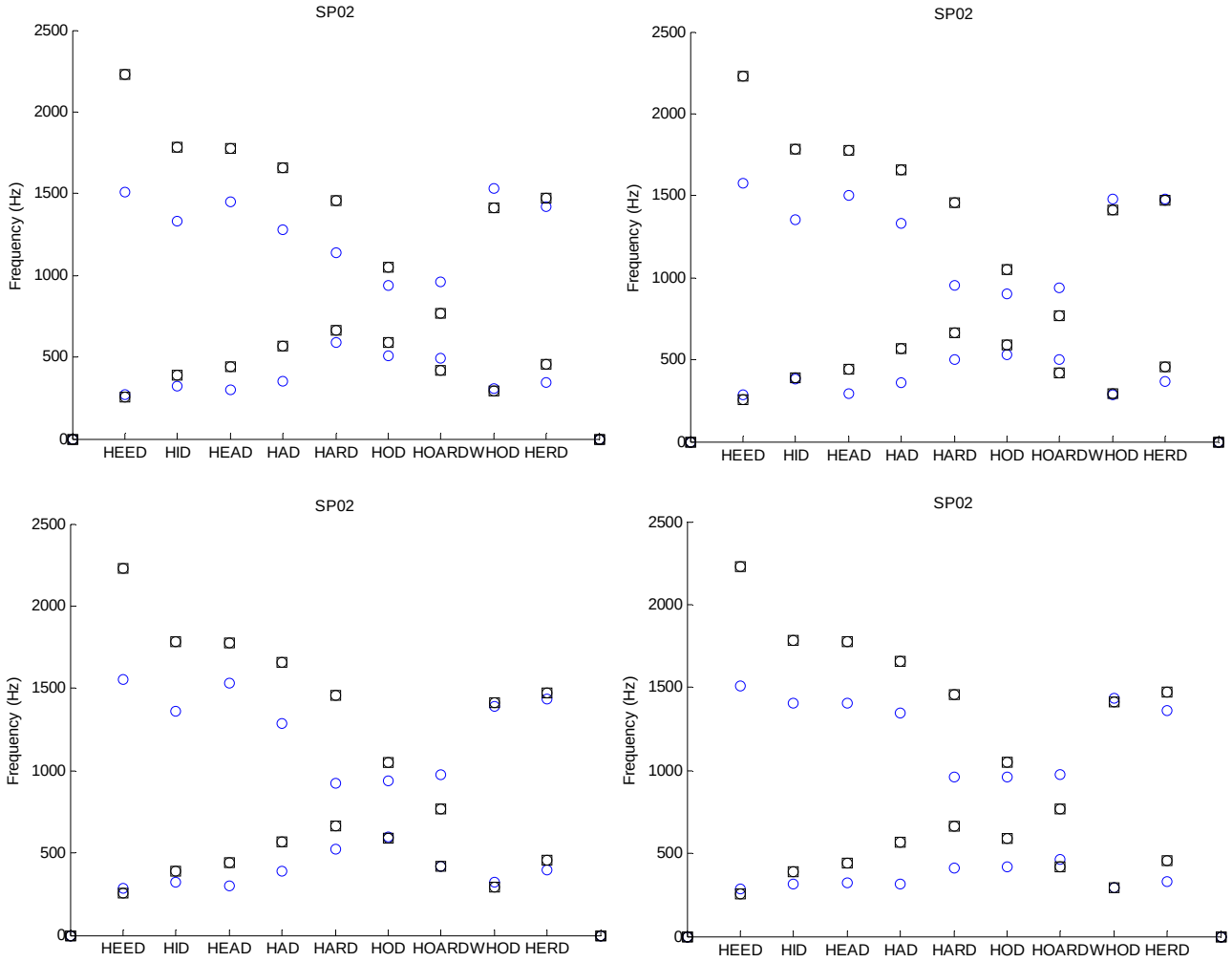


Figure B. 12: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP02)

SP03

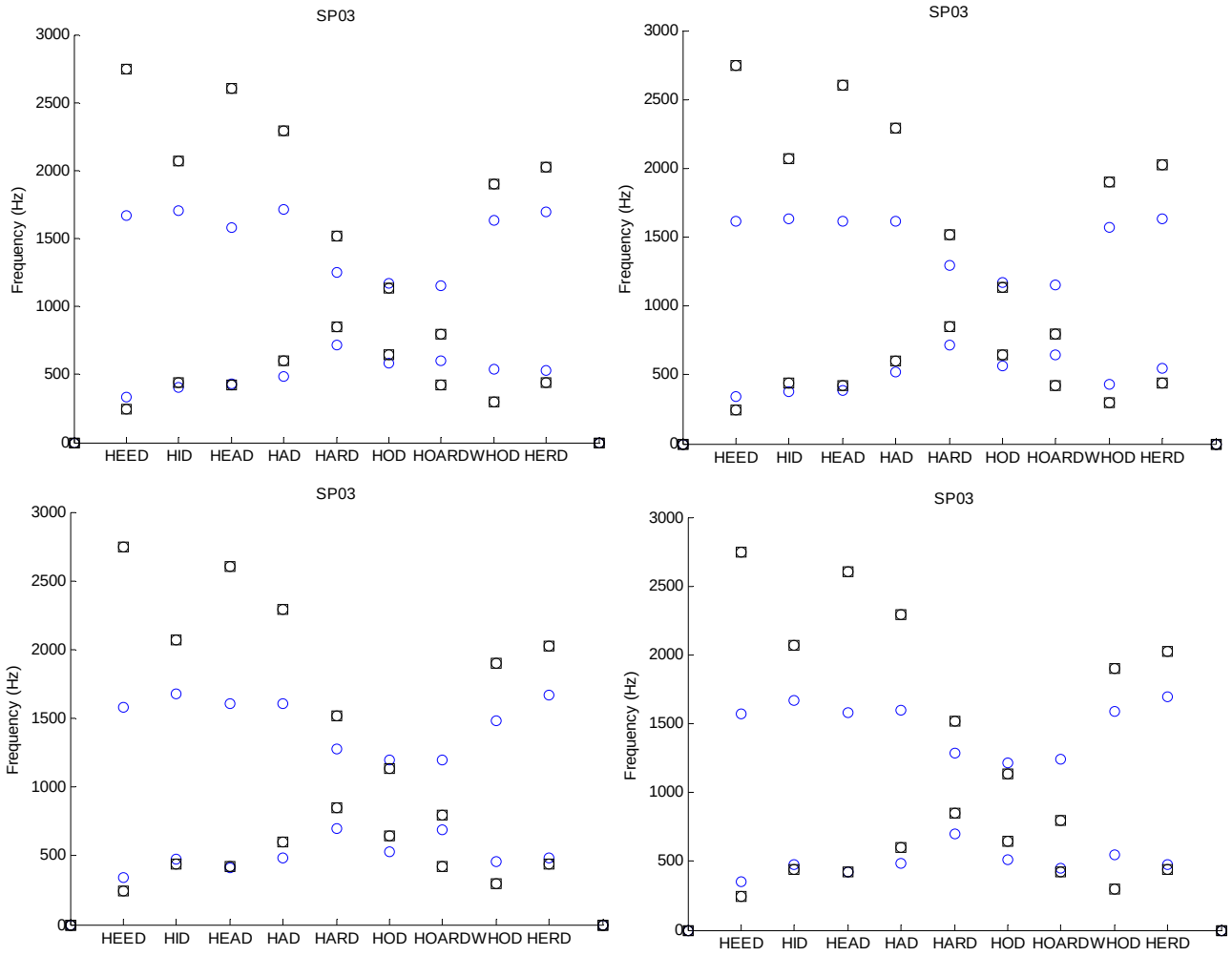


Figure B. 13: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP03)

SP04

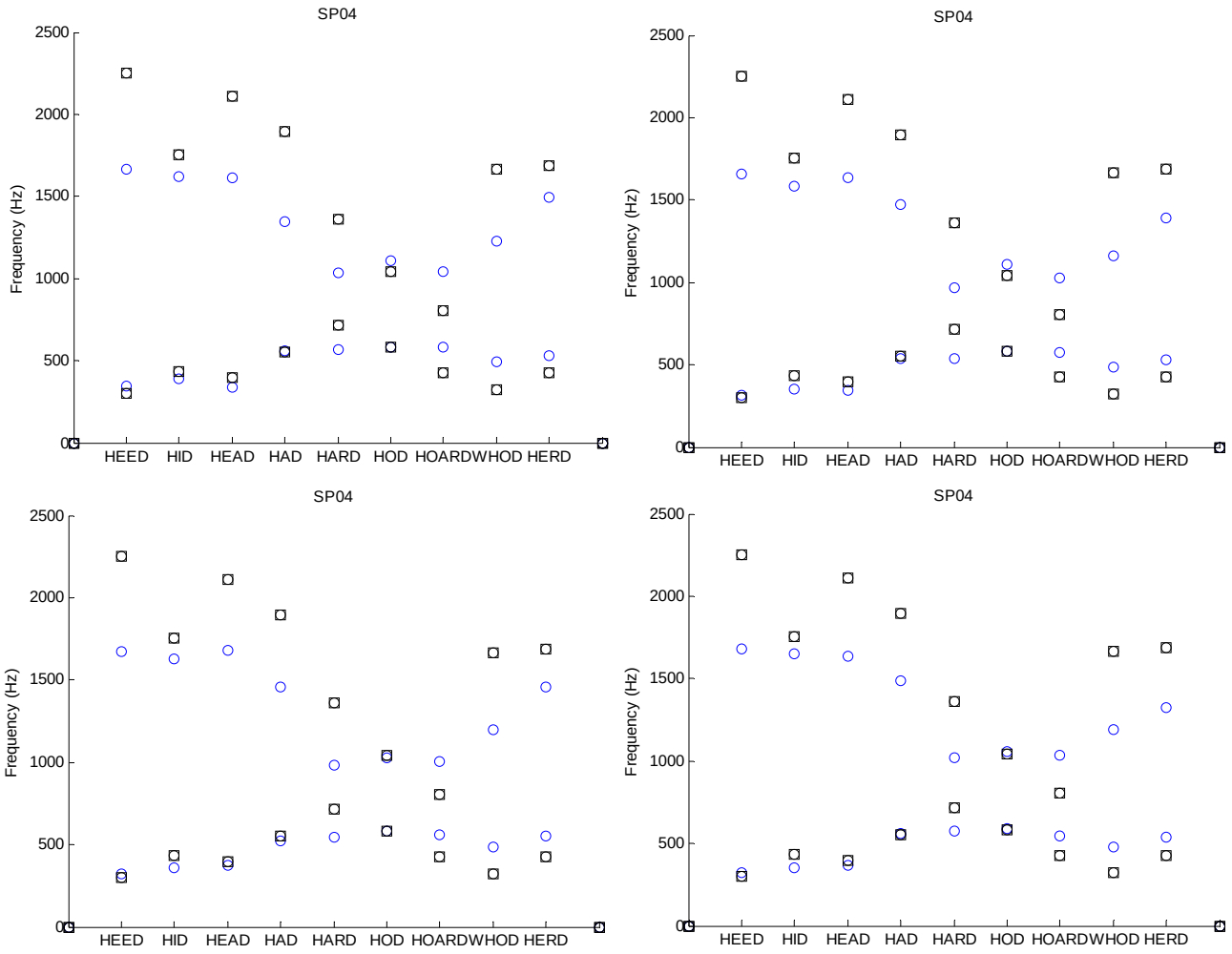


Figure B. 14: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP04)

SP05

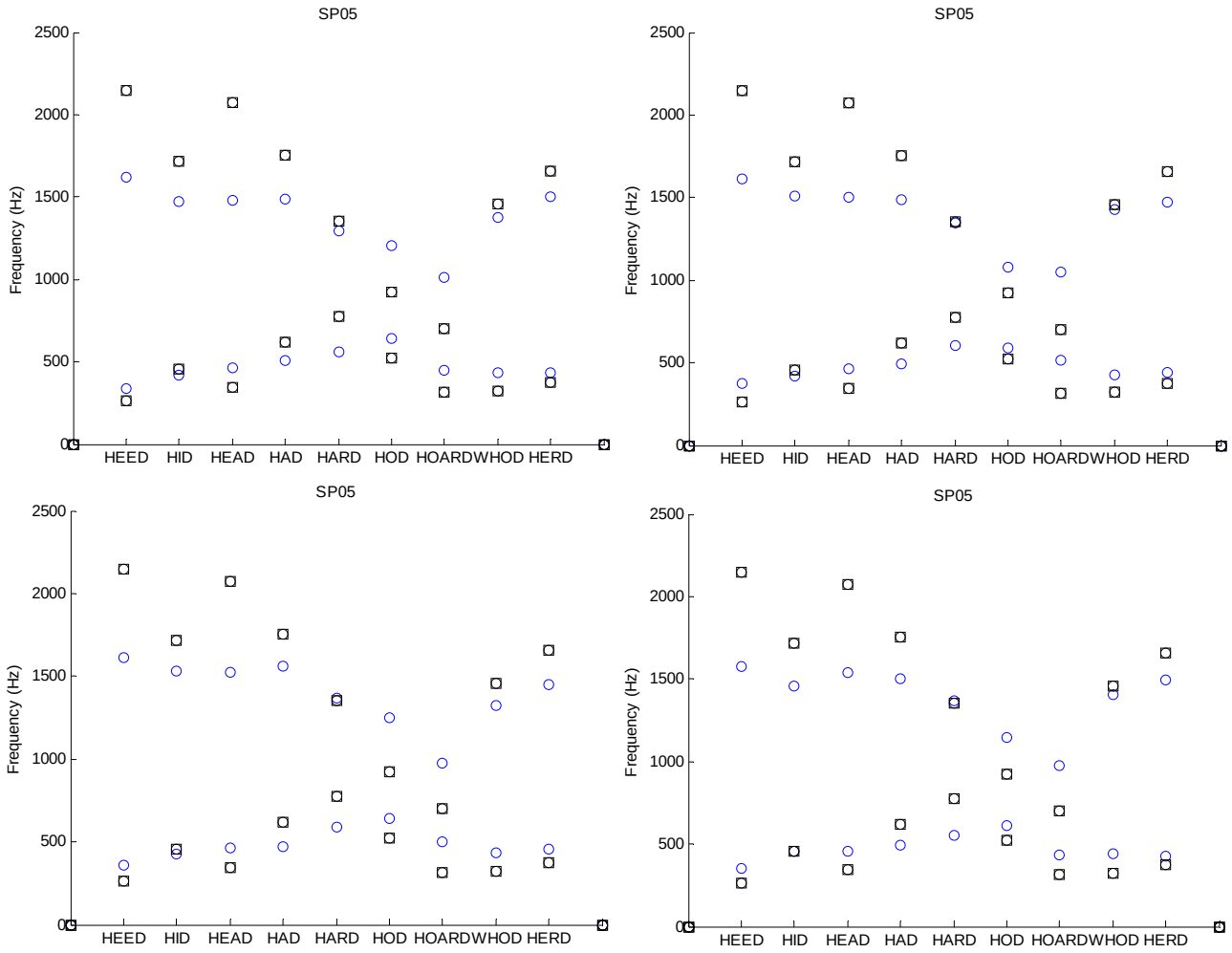


Figure B. 15: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP05)

B.4. MRI resonance vs. Formants

SP01

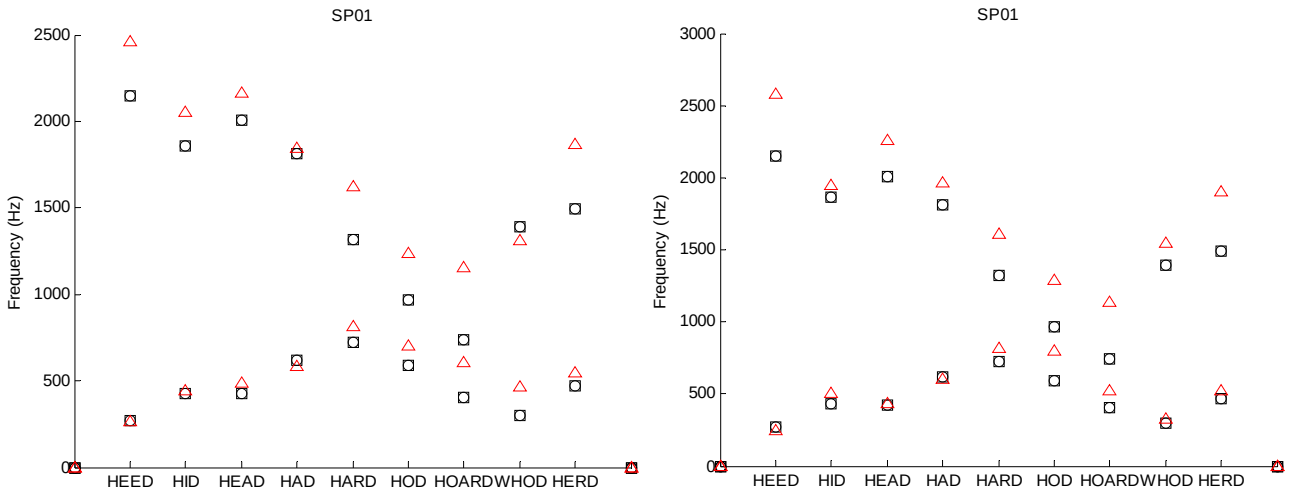


Figure B. 16: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP01 MRI)

SP02

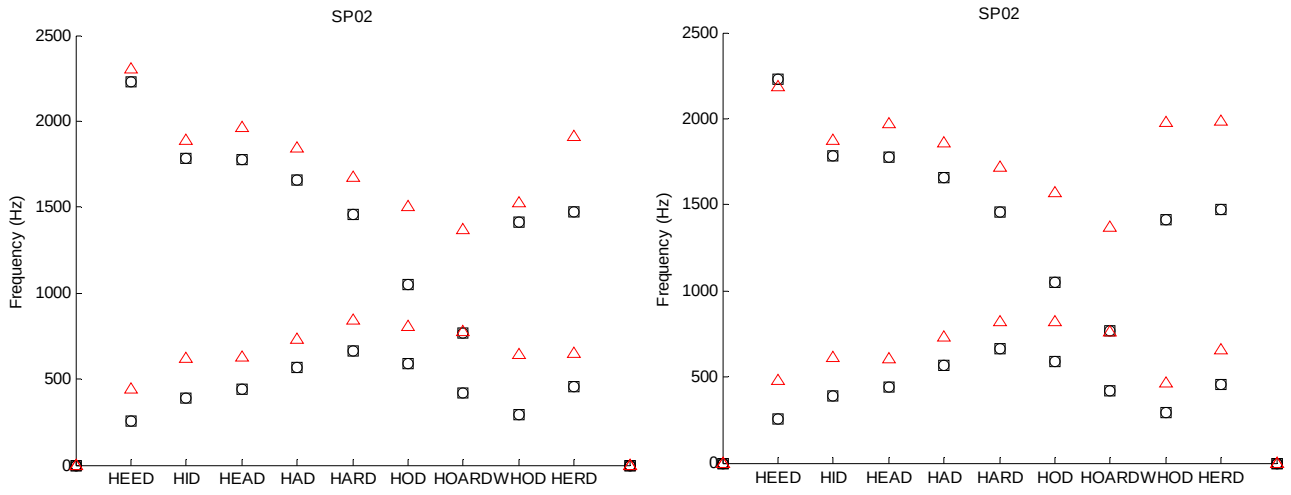


Figure B. 17: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP02 MRI)

SP03

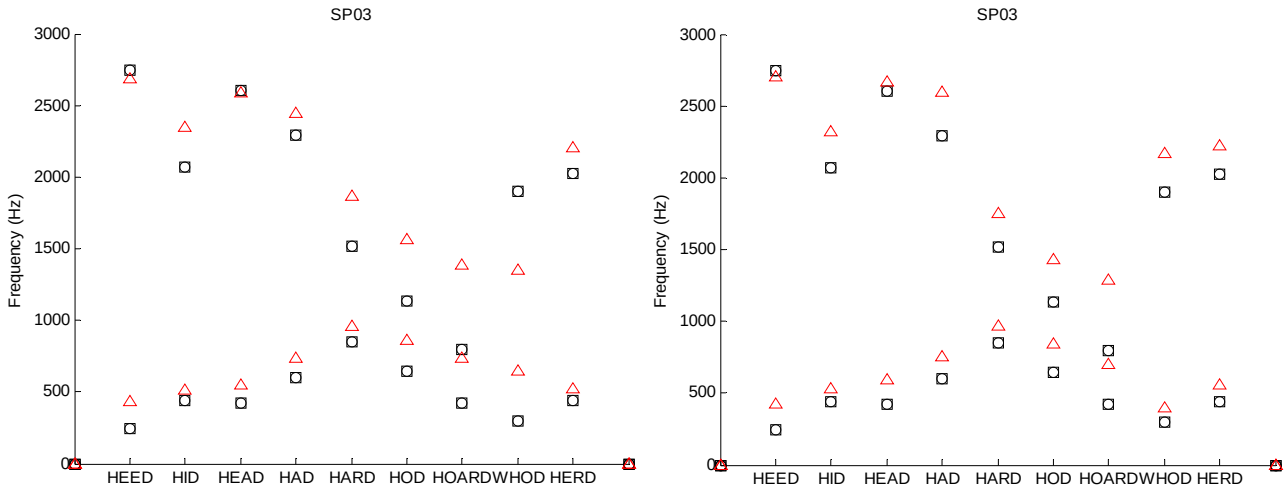


Figure B. 18: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP03 MRI)

SP04

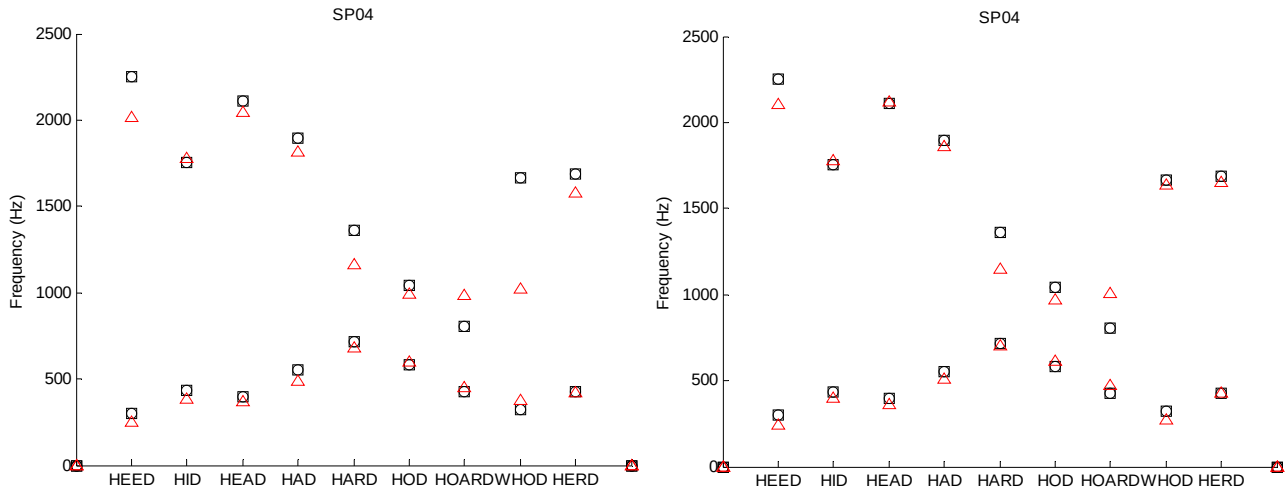


Figure B. 19: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP04 MRI)

SP05

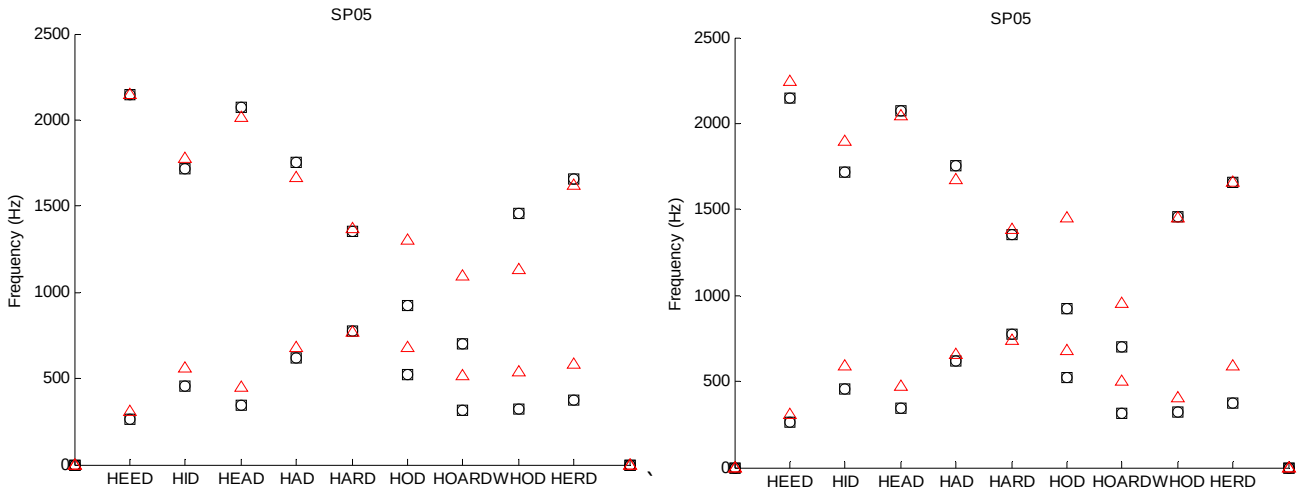


Figure B. 20: Plot of first and second resonances calculated from AR cross-sectional area function vs. first and second formants extracted from recorded speech. (SP05 MRI)

References

Baer, T., Gore, O. C., Gracco, L. C., & Nye, P. W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *The Journal of the Acoustical Society of America, 90*(2, Pt 1), 799-828.

Bier, P. (2003). *Modelling the Vocal Tract.* University of Auckland, Unpublished.

Clément, P., Hans, S., Hartl, D. M., Maeda, S., Vaissière, J., & Brasnu, D. (2007). Vocal Tract Area Function for Vowels Using Three-Dimensional Magnetic Resonance Imaging. A Preliminary Study. *J Voice, 21*(5), 522-530.

Dehqan, A., Scherer, R. C., Dashti, G., Ansari-Moghaddam, A., & Fanaie, S. (2013). The Effects of Aging on Acoustic Parameters of Voice. *Folia Phoniatrica et Logopaedica, 64*(6), 265-270.

Eason, D. (2009). *Practical Work Report*. Unpublished report. Electrical and Electronics Engineering. University of Auckland.

Fant, G. (1970). *Acoustic Theory of Speech Production*: Walter de Gruyter.

Gray, H. (1918). *Anatomy of the human body*. Philadelphia: Lea & Febiger.

Gregory, N. D., Chandran, S., Lurie, D., & Sataloff, R. T. (2012). Voice Disorders in the Elderly. *Journal of Voice, 26*(2), 254-258.

Harrington, J., & Cassidy, S. (1999). *Techniques in speech production*: Kluwer Academic Publishers.

Ladefoged, P. (1995). *Elements of Acoustic Phonetics*: University of Chicago Press.

Mullen, J., Howard, D. M., & Murphy, D. T. (2006). Waveguide Physical Modeling of Vocal Tract Acoustics: Flexible Formant Bandwidth Control From Increased Model Dimensionality. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, 14*(3).

Perrier, P., Boe, L. J., & Sock, R. (1992). Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modeling the transition with two sets of coefficients. *J Speech Hear Res, 35*(1), 53-67.

Rabiner, L. R., & Schafer, R. W. (1978). *Digital processing of speech signals*: Prentice-Hall.

Rosen, S., & Howell, P. (2010). *Signals and Systems for Speech and Hearing*: Emerald.

Sataloff, R. T., Caputo Rosen, D., Hawkshaw, M., & Spiegel, J. R. (1997). The aging adult voice. *Journal of Voice, 11*(2), 156-160. doi: http://dx.doi.org/10.1016/S0892-1997(97)80072-0

Stevens, K. N. (2000). *Acoustic Phonetics*: Mit Press.

Stone, M., & Davis, E. P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *J Acoust Soc Am, 98*(6), 3107-3112.

Stone, M., & Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *J Acoust Soc Am, 99*(6), 3728-3737.

Story, B. H. (2002). An overview of the physiology, physics and modeling of the sound source for vowels. *Acoust. Sci. & Tech, 23*(4), 195-206.

Story, B. H. (2008). Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *J Acoust Soc Am, 123*(1), 327-335. doi: 10.1121/1.2805683

Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *J Acoust Soc Am, 100*(1), 537-554.

Story, B. H., Titze, I. R., & Hoffman, E. A. (1998). Vocal tract area functions for an adult female speaker based on volumetric imaging. *J Acoust Soc Am, 104*(1), 471-487.

Story, B. H., Titze, I. R., & Hoffman, E. A. (2001). The relationship of vocal tract shape to three voice qualities. *J Acoust Soc Am, 109*(4), 1651-1667.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., & Fujimoto, I. (2006). Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. *J Acoust Soc Am, 119*(2), 1037-1049.

Tameem, H. Z., & Mehta, B. V. (2004). Solid modeling of human vocal tract using magnetic resonance imaging and acoustic pharyngometer. *Conf Proc IEEE Eng Med Biol Soc, 7*, 5115-5118. doi: 10.1109/iembs.2004.1404413

Titze, I. R. (1994). *Principles of voice production*: Prentice Hall.

Watson, C. I., Harrington, J., & Evans, Z. (1998). An acoustic comparison between New Zealand and Australian English vowels∗. *Australian Journal of Linguistics, 18*(2), 185-207. doi: 10.1080/07268609808599567

Watson, C. I., & Hui, C. J. (2010). *Two Short Studies in Vocal Tract Measurements.* Paper presented at the 13th Australasian International Conference on Speech Science and Technology

Watson, C. I., Thorpe, C. W., & Lu, X. B. (2009). A Comparison Of Two Techniques That Measure Vocal Tract Shape. *Acoustics Australia, 37*(1), 7-11.

Wismueller, A., Behrends, J., Hoole, P., Leinsinger, G. L., Reiser, M. F., & Westesson, P. L. (2008). Human vocal tract analysis by in vivo 3D MRI during phonation: a complete system for imaging, quantitative modeling, and speech synthesis. *Med Image Comput Comput Assist Interv, 11*(Pt 2), 306-312.

Xue, A. (1999). Age-related changes in human vocal tract configurations and the effects on speakers' vowel formant frequencies: a pilot study. *Logopedics Phoniatrics Vocology, 24*(3), 132-137. doi: doi:10.1080/140154399435084

Xue, S. A., Cheng, R. W., & Ng, L. M. (2010). Vocal tract dimensional development of adolescents: an acoustic reflection study. *Int J Pediatr Otorhinolaryngol, 74*(8), 907-912. doi: 10.1016/j.ijporl.2010.05.010

Xue, S. A., & Hao, G. J. (2003). Changes in the Human Vocal Tract Due to Aging and the Acoustic Correlates of Speech Production: A Pilot Study. *Journal of Speech, Language & Hearing Research, 46*(3), 689-701.

Xue, S. A., & Hao, J. G. (2006). Normative Standards for Vocal Tract Dimensions by Race as Measured by Acoustic Pharyngometry. *J Voice, 20*(3), 391-400.

Yang, C. S., & Kasuya, H. (1994). *Accurate measurement of vocal tract shapes from magnetic resonance images of child, female and male subjects.* Paper presented at the International Conference on Spoken Language Processing.