Acoustic Pulse Reflectometry for Measurement of the Vocal Tract with Application in Voice Synthesis

Calum David Gray

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy to the

University of Edinburgh

2005

Abstract

The measurement of human airway dimensions has been a frequent objective in the fields of respiratory medicine and speech research, but has proven difficult to achieve non-invasively due to the airway's function in breathing, swallowing and speaking. Acoustic pulse reflectometry (APR) has been employed in clinical studies of the vocal tract for several years, normally in the function of airway measurement. The focus of this work is to utilise APR in capturing vocal tract profiles during the phonation of vowel sounds, for the purposes of sound synthesis. By making an equivalent tube model of the vocal tract, the propagation of an acoustic wave can be readily calculated using techniques such as waveguide modelling, which will in turn allow us to synthesise sound and form the basis of a physical model of the voice. The attractions of this technique for vocal tract measurement are many: it is non-invasive, safe, repeatable and inexpensive.

In this thesis, the basic theory describing wave propagation in tubes of varying cross-section is outlined, together with a review of how the time domain technique of APR can be used to measure the input impulse response of a tubular object, such as the vocal tract, from which the bore profile can be calculated using the layer peeling algorithm.

Experimental measurements of the human vocal tract during the phonation (imitation) of five non-nasalised vowels [a, e, i, o, u] are presented, using recent enhancements to the APR technique (MLS excitation signals and virtual DC tube method) for a single subject, together with optimisation of the APR technique for vocal tract measurement and its application in a group study using adults and children.

To validate the results obtained using the APR technique, a comparative study with an accepted 'gold standard' imaging technique (Magnetic Resonance Imaging - MRI) is presented, using the same subject, a voice professional, in both studies. The results from this study show reasonable overall agreement between the APR and MRI data, with the limited resolution of the acoustic technique tending to broaden features and underestimate cross sectional areas, particularly in the region of the pharynx and glottis. Protocols and supplementary documentation required by scientific, clinical and ethical review bodies for the use of human volunteers in research trials are provided. From this study a data corpus of vocal tract measurements is gathered, using the techniques of APR and MRI, in adult males, adult females and children.

In conclusion, limitations of the APR technique for vocal tract measurement are discussed and potential improvements are proposed.

Declaration

I do hereby declare that this thesis was composed by myself and that the work described within is my own, except where explicitly stated otherwise.

Calum D. Gray

November 2005

Acknowledgements

A research study such as this would not have been possible without the help, expertise and guidance of others. I would like to use this opportunity to acknowledge and thank those who have contributed to the realisation of this thesis.

To my supervisors, Prof. Clive Greated and Prof. Murray Campbell from the School of Physics, and Dr Simon King from the Centre for Speech Technology Research (CSTR). My sincere thanks for your support and guidance which has allowed me to undertake and explore this exciting area of research.

To Dr Ian Marshall for your advice on airway measurement using APR in the early stages of my studies, and in particular, your assistance more latterly with the MRI study. Your support and interest in this work has been invaluable.

To the staff at the Brain Imaging Research Centre for Scotland (SBIRCS), especially Mr Lindsay Murray, for the support and guidance I received to gain approval from the ethical, medical and scientific bodies in relation to the MRI studies.

To Dr Peter Wraith from the Scottish National Sleep Centre who was instrumental in providing me with advice and equipment for the 1st prototype of the reflectometer.

My thanks to the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, for providing access to their Image Analysis Core and especially for the support and training I received from Dr Tom MacGillivray in analysing the data gathered through the MRI study.

To Dr Jonathan Kemp for your time and assistance in demystifying APR, and for fixing my reflectometer which I seemed to break on more than one occasion!

To Dr Maarten van Walstijn for your many chats about all things 'Physical Modelling'.

I would like to thank all my colleagues within the Acoustic and Fluid Dynamics Group and the CSTR, not only for providing a convivial working environment but also for being 'willing' volunteers in lending me their vocal tracts to perform trial runs of the APR.

To Ian, for your assistance and unfaltering good humour throughout many lengthy laboratory sessions.

The financial support from EPSRC is gratefully acknowledged

My good friends Peter, Keiron and Neil, for the welcomed distractions to academic study and links to the outside world.

Finally, to my family – Pam, Darren, Mum, Dad & Gran. Thank you for all your support and encouragement. None of this would have been possible without you.

Contents

Chapter 4

Chapter 5

List of Figures

List of Tables

Chapter 1

Introduction and outline of research programme

"In the beginning was the voice. Voice is sounding breath, the audible sign of life."

--Otto Jespersen [1]

1.1 Context

From the first cries of primeval man to the intelligent articulations of modern language, the human voice is a complex and wondrous instrument and has been used as a method of communication, expressing emotions, uttering danger signals and singing.

The voice is arguably the first and original musical instrument, and the art and development of singing, when musical tones are produced by vocal effort, is so basic to the human species that it was almost certainly a characteristic before language was developed.

Singing is part of all cultures and is associated with a diverse range of functions including religion, entertainment and other matters considered important to the individual. From time immemorial, mankind has used singing to praise gods, strike fear into enemies, woo suitors and recount history with ballads and stories.

1.1.1 Voice Synthesis

The interest of this thesis is in the area of physical modelling voice synthesis, and in particular, the synthesis of the singing voice for musical applications.

Over the last fifty years, research in the field of voice science and related technology has significantly intensified. This has led to a greater understanding of the human voice mechanisms, numerous new applications of computer speech, transmission and coding of speech, and voice recognition which have found a way into our daily life. From the mobile phones we use, to the daily announcements we hear, the products of voice science are all around.

One such area is that of voice synthesis. In the most general sense voice synthesis is the production of 'synthetic' or 'artificial' voice using a personal computer or other computing device[1]. This involves producing an electronic signal which when played through a transducing device, such as a loudspeaker, resembles the human voice enough for the human brain to interpret it as such. More technically, this means that the signal must contain a reasonable representation of the voicing and different harmonic resonances associated with the underlying formants in the vocal tract [2].

Several methods of voice synthesis exist, the most popular and commercially successful at present being concatenative speech synthesis. This method is based on the concatenation (or stringing together) of segments of recorded speech. Signal processing techniques are used to assemble the parts, which can vary in size from sentences or words down to syllables or phonemes, rejoining the pieces into suitable speech. Generally, concatenative synthesis gives the most natural sounding synthesised speech. However, natural variation in speech and automated techniques for segmenting the waveforms sometimes result in audible glitches in the output, detracting from the naturalness. A notable feature of this method is that no underlying principles of voice generation are modelled since, in basic terms, pre-recorded parts of real speech are played back from a speech database to produce the voice output.

---

[1] The reference here is to electronic voice synthesis by means of signal processing and negates earlier mechanical forms such as the bellows-operated 'Acoustic-Mechanical Speech Machine' by Wolfgang von Kempelen of Vienna, Austria, described in his 1791 paper *Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine* ("mechanism of human speech with description of his speaking machine", J.B. Degen, Wien).

It is generally considered that a plateau will be reached in terms of the optimal quality that can ever be achieved via this form of synthesis [3, 4], as there will always be a 'join cost' between each speech unit, no matter how large a speech database that units are selected from.

In contrast to this approach, physical modelling of the voice, or its counterpart in speech science termed 'articulatory synthesis', is the synthesis of voice sound using mathematical models of the human vocal tract and the articulation processes.

Physical models for voice synthesis have often been rejected because of their poor sound quality and are not currently sufficiently advanced or computationally efficient to be used in commercial systems. Therefore at present, this method of synthesis is mostly of academic interest. With a greater understanding of the phonatory system and the availability of fast computers, these trends are beginning to be reversed. The computer is able to solve iteratively complex systems of equations that describe the physics of flow generation and sound propagation in the human vocal folds and vocal tract, allowing high quality models to be realised in voice science research. By using equations and algorithms to describe and simulate the physical laws of voice generation, this method has the advantage of being able to interact with 'real' control parameters of voice physiology, giving it the potential to create a level of natural voice never before attained by any form of synthesis.

However, a model, by its definition, is only an approximation of the observed phenomenon. Therefore the challenge for the scientist is to encapsulate the important features of the real physical system into the model.

Speech generation is a very complex topic that extends from physical and medical science to the areas of linguistics and phonetics, therefore a multidisciplinary approach is required. A satisfactory physical model of the voice has not yet been realised, but the advantages of such an approach would be beneficial to each discipline. For example:

- the synthesis of individual voices could be based upon the actual physiology [5]
- speech transmission could use a set of physically-based parameters for increased naturalness of the speech
- applications in medical diagnostics – such as the modelling of pathologic voice caused by edema or fold nodules [5, 6]

Before such goals can be achieved, much work is required at a fundamental level. The first stage is the analysis and measurement of static and dynamic anatomical structures used in voice production which can be translated into future voice models.

The focus of this work is on developing a non-invasive method to measure the bore (shape) of the vocal tract for static targets of articulation, which can be used as 'real' geometric data for application in a physical model of the voice.

1.1.2 Airway Measurement

The measurement of the vocal tract is not only of interest to the voice science community. The anatomical structures used in voice production are shared with the other human processes of breathing and swallowing, therefore to analyse and understand their principle of operation is also of great importance to medical science. The human vocal tract and upper airway are one and the same, but we use them interchangeably to refer to the context of their application; the term 'vocal tract' when referring to the voice and the term 'airway' when referring to the respiratory system.

Marshall [7] showed that measurements of airway dimensions would be of clinical benefit in the investigation of obstructive sleep apnoea (which is caused by occlusion of the upper airway) and in the general anaesthetic management of patients. However, there were no techniques for measuring airways that met the requirements of being non-invasive, portable, inexpensive and simple to operate [8]. The standard clinical methods of X-ray computed tomography (CT) and Magnetic Resonance Imaging (MRI) were shown to produce limited data under certain specialised conditions, but were not suitable for routine measurement or monitoring. This was also true of vocal tract measurement for speech.

In vivo studies of the vocal tract are of limited use when investigating voice mechanisms as they are invasive and impinge on normal function. Therefore, for almost forty years, speech researchers have experimented with acoustic methods for estimating the dimensions of the vocal tract.

1.1.2.1 Acoustic Methods

In the late 1960s, Schroeder and Mermelstein [9, 10] measured the resonant frequencies of the vocal tract from which airway dimensions were then inferred (having assumed the vocal tract length). Several years later, Sondhi and colleagues [11] adapted a time domain reflection method which was originally developed as a seismological technique for the study of the earth's crust. The principle was that an explosion generated at the surface would cause an impulsive acoustic wave to travel down into the earth. At changes in density within the rocks, a portion of the incident wave would be reflected, and as there will be many layers of different density, the resulting reflections would form a complicated response when measured at the surface. The sequence of reflections measured at the surface was termed the input impulse response. Ware and Aki [12] were first to provide an algorithm which would calculate the reflection coefficients (densities) of the layers from this impulse response. However, their method did not compensate for losses experienced by the incident and reflected waves whilst propagating through the rock layers.

Sondhi et al. suggested that applying the same technique to propagation in air using an airborne pulse, the cross-sectional area of a tubular structure could be calculated from its reflections (input impulse response). For airway measurement, this meant applying a pulse of sound into the mouth of a subject and recording the resulting reflections at the lips. Using the Ware-Aki algorithm, the cross sectional area of the airway as a function of distance from the lips (area profile) could be determined. This technique did not require any assumption of vocal tract length. It is this technique, known as acoustic pulse reflectometry (APR), which forms the subject of the present work.

The history of this technique has been reviewed in detail in the thesis of Sharp [13] and only a summary is presented here.

1.1.2.2 APR Measurements on Human Patients

In the late seventies and early eighties, Jackson et al [14] were the first to report on the measurement of physiological structures by measuring the area profiles of excised dog tracheas and lungs using the APR technique with the Ware-Aki algorithm, and then subsequently human airway casts [15]. Their method of APR involved generating an acoustic impulse in a source tube and transmitting it into the airway. The signal underwent partial reflection and partial transmission at changes in impedance, caused by changes in cross-sectional area along the length of the airway, and created a reflection sequence. The returning sequence travelled back up the source tube without further reflection and was recorded by a microphone embedded in the wall of the source tube. The Ware-Aki algorithm used to reconstruct the cross-sectional profile did not take into account losses in the airway, however good area profiles were still achieved because losses had not proved significant when reconstructing a short object such as the airway.

The first clinical applications with measurements on human patients were carried out by Fredberg et al. [16], laying the groundwork from which a series of successful clinical trials followed [17-19]. Marshall [7, 20-23] continued the development of the reflectometer specifically for use in clinical applications and, in particular, obstructive sleep apnea. His work focussed on developing a reflectometer that was portable and simple to operate, whilst at the same time allowed subjects to breathe during measurements and display airway area profiles in ‘real-time’. Part of this involved shortening the source tube length to only a few centimetres and applying a mathematical treatment which removed the need for maintaining the time separation between the incident pulse and source reflections. The airway profiles calculated using the Ware-Aki algorithm were found to be reasonably accurate, but not as great as that achieved by the longer source tube reflectometer.

1.1.2.3 APR Measurements on Musical Wind Instruments

In parallel research in musical acoustics, much work had gone into developing the APR technique and applying it to the measurement of brass and wind instruments [24-26]. Compared with the human airway, many of these instruments are much longer in length. As a result, losses affect the input impulse response significantly and need to be accounted for in the bore reconstruction algorithm if accurate results are to be achieved. Such losses could not be included in the Ware-Aki algorithm, therefore an alternative algorithm was developed by Amir et al. which incorporated viscothermal losses [27, 28]. This lossy layer peeling algorithm was successfully used by Sharp, in conjunction with a method he proposed for removing the DC offset of the input impulse response (DC tube method), to provide accurate reconstructions of the internal profile of brass instruments [13, 29, 30]. As a variation on the DC tube method, Kemp [31] proposed a 'virtual DC tube' method which simulates the effect of the DC tube without the need for an intermediate physical coupling. This method was not only convenient, it also provided a perfect coupling between object and source tube.

The lossy layer-peeling algorithm and virtual DC tube method are used throughout this work and are described in more detail in Chapters 3 and 4.

1.2 Aims and outline of thesis

The aims of this thesis are:

1. to develop the APR technique as a non-invasive method for measuring the bore of the vocal tract for static targets of articulation.
2. to apply, and investigate the repeatability, of the APR technique for the measurement of the vocal tract in a group study of volunteers ranging from children to adults.
3. to implement a comparative study between APR and MRI for the measurement of the vocal tract during the phonation (imitation) of vowel sounds, and produce a framework for future investigations, which will include all supporting documentation to satisfy scientific, clinical & ethical bodies for the use of human volunteers in research trials.
4. to produce a corpus of data of vocal tract measurements from APR and MRI investigations, which can be used as 'real' geometric data for application in a physical model of the voice.

Chapter 2 begins with an introduction to voice production in relation to anatomy and physiology. Differences between speech and singing are summarised and the translation of physical principles into a physical model of the voice are presented.

In Chapter 3, the basic theory for plane wave propagation within tubular cavities is described together with a method for solving the inverse problem of determining the dimensions of an object from its input impulse response. A development to the standard APR technique using MLS excitation signals is also summarised.

A working reflectometer using enhancements to the standard technique together with modifications for its application to measuring the human vocal tract is described in Chapter 4. The operating procedure is discussed in detail and measurements on test objects are presented.

In Chapter 5, the optimisation of the APR technique for measurement of the human vocal tract using a single subject for the purpose of investigating repeatability of the technique under controlled conditions is described.

Chapter 6 describes the procedure and application of the APR technique for vocal tract measurement in a group of volunteers (ranging from children to adults).

A true validation of measurements can only be realised by comparing results with an accepted gold standard technique. Chapter 7 describes the design, process and results from a peer reviewed study entitled "*Comparative study between acoustic pulse reflectometry (APR) and magnetic resonance imaging (MRI) for volumetric measurement of the vocal tract during vowel production*".

In Chapter 8, a summary of the main conclusions of this work are presented. Suggested extensions of this work are then given.

Chapter 2

The human voice

2.1 Introduction

In this chapter, the basic principles of voice production are introduced and methods used to model the voice systems are presented.

2.2 The voice organ

The voice organ constitutes three different systems: the breathing apparatus, the vocal folds and the vocal tract. The principal features are shown in the midsagittal section through the head and neck sketched in Figure 2.1.

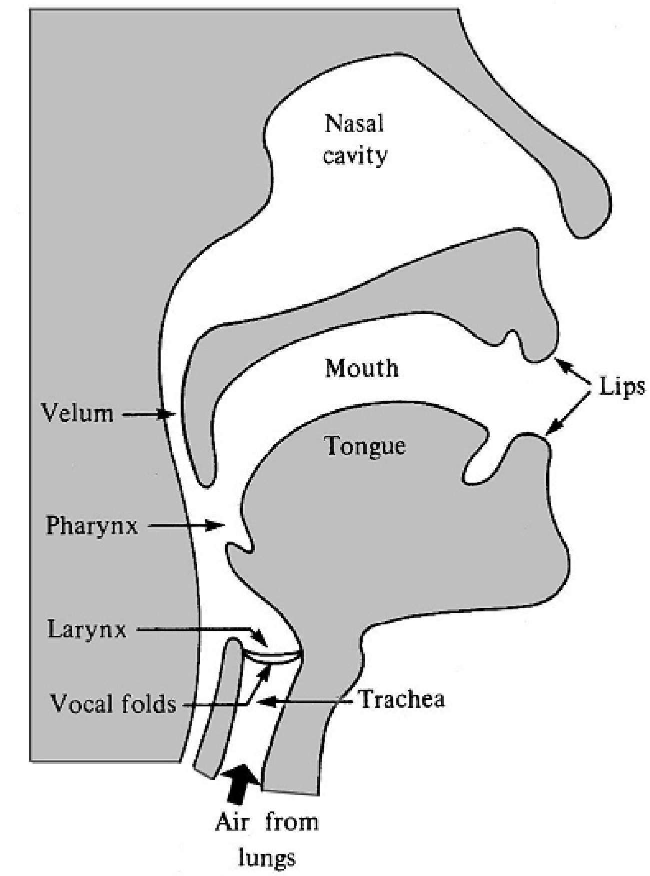


Figure 2.1: Mid-sagittal profile showing the principal features of the voice organ (adapted from [32] p.472)

Consider the physical system that gives rise to the acoustical signal of the voice. The acoustical behaviour of the organ begins by filling the lungs with air. By compressing this volume, an airstream is forced upwards through the trachea (windpipe). At the uppermost part of the trachea is the larynx which is a short tube composed of muscle and cartilage that acts as a valve, regulating the flow of air between the trachea and pharynx (the back of the throat). At the base of the larynx, the mucous membrane lining which is folded inward from each side forms what are collectively known as the vocal folds, and the slit-like opening between them is called the glottis. The separation and tension of the vocal folds are controlled by the muscles in the larynx and can be held apart to permit the free flow of air in breathing, or brought closely together to allow phonation. The forcing of an airstream through the glottis in this latter state sets the vocal folds into vibration, alternately opening and closing the glottal aperture. As a result, the flow of air from the lungs is modulated at the vibration frequency of the vocal folds, and it is this that determines the pitch of the speech or singing sound.

The sound wave generated within the larynx by the vibrating vocal folds travels through the pharynx, the mouth (oral cavity) and the nasal cavity before finally emerging from the lips and nose apertures. Each cavity has its own characteristic resonance frequencies, which play a vital role in determining the vocal output sound. Each cavity resonance gives rise to a formant region in the spectrum of the radiated sound. By adjusting the articulators (i.e. the opening of the lips, the position of the lower jaw, and the shape and position of the tongue), the singer or speaker is able to choose particular combinations of formant frequencies and produce what we perceive as voice sounds, such as vowels.

The human voice is able to produce a large variety of sounds; voiced, unvoiced, fricatives, plosives, nasals etc. However, in our discussion of the acoustics of the voice organ, we shall concentrate on the non-nasalised voiced sounds of the singing voice. That is, sounds that are produced by vibrating the vocal folds and resonate within the larynx, pharynx and mouth cavities, which constitute the vocal tract.

The pitch of a sung note is determined by the vibration frequency of the vocal folds. With very few exceptions, the motion of the folds is almost completely independent of the shape of the vocal tract, whereas the choice of a particular vowel sound is made by modifying the vocal tract to produce three or four characteristic formants; the frequencies of the formants corresponding to a given vowel are very nearly the same regardless of the pitch at which it is sung [32]. A convenient way of describing these acoustical features can be found in the 'source-filter theory' proposed by Fant in the early sixties [33]. The theory states that speech production can be divided into two independent parts:

- Source of acoustic energy
- Variable acoustic filter

In this model (see Figure 2.2), the source of acoustic energy is at the larynx (i.e. vibrating vocal folds), and the supralaryngeal vocal tract serves as a variable acoustic filter whose shape determines the phonetic quality of the sound.

For voiced sounds, the vibration of the vocal folds produces a periodic, harmonic-rich signal at the fundamental frequency. The signal is transmitted outside the mouth by the vocal tract, which has a frequency dependent gain. Resonances of the vocal tract produce peaks in the gain spectrum that in turn give rise to maxima in the envelope of the radiated sound spectrum. The broad peaks in the output sound spectrum are formants, which as discussed previously are characteristic features of voiced speech sounds. It is important to note however that the formant peaks in the resulting radiated sound spectrum will occur on the closest source harmonic peaks to the resonance peaks, therefore the resonant frequency and the formant frequency are not necessarily the same.

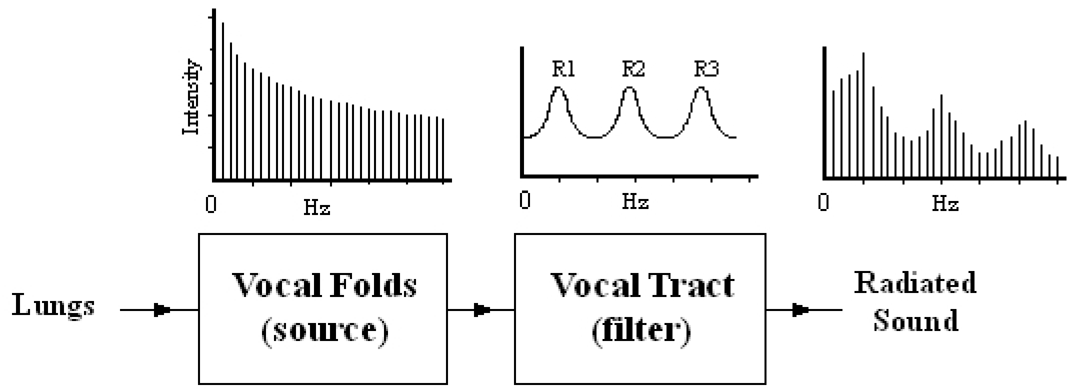


Figure 2.2: Source filter model of speech production

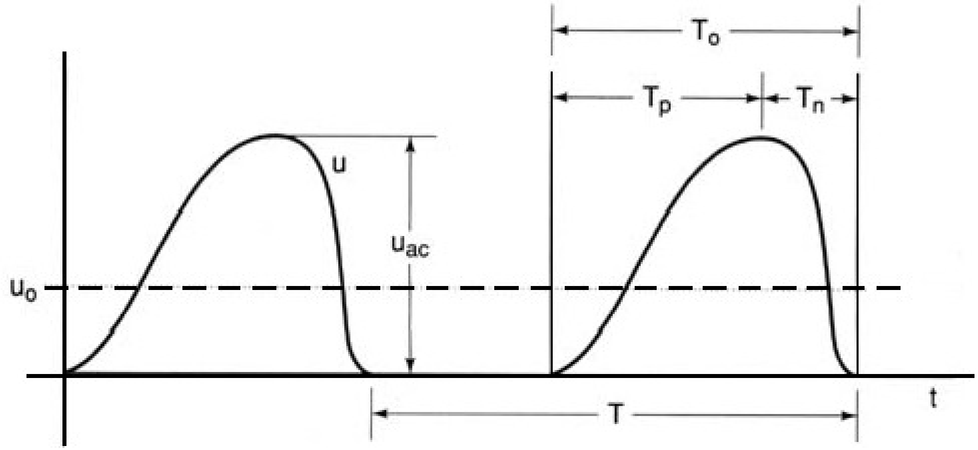
The following sections consider the acoustics of the voice organ in more detail by concentrating first on the sound source (the vocal folds), and then on the vocal tract resonances.

2.3 The vocal folds

The flow-induced oscillation of the vocal folds begins when a build-up of pressure behind the closed vocal folds eventually forces them apart, allowing air from the lungs to flow through. The Bernoulli force, which arises from the reduction in pressure due to the air flow, acts together with the natural myoelastic forces of the folds to pull them together again. As they close, the air flow is cut off, the Bernoulli force disappears and a new cycle begins. Each cycle produces a single small puff or air, with tens or hundreds of these small puffs of air being released every second and filtered by the vocal tract.

The vocal fold vibration must be self-sustaining to give rise to a continuous sound. When the folds are closing, the airflow through the glottis begins to decrease, but airflow above the glottis continues to move with its same speed due to inertia. This creates a region just above the vocal folds where the air pressure decreases as airflow is slower moving through the glottis than it is leaving above. Therefore the Bernoulli force boosts the closing of the folds before the airflow disappears at full closure. This is the regeneration condition which allows the vibration to draw energy from the air flow supplied by the lungs [32].

A typical glottal flow waveform is shown in Figure 2.3.



T_p	Increasing air flow rate
T_n	Decreasing air flow rate
T_o	Length of time during each cycle in which air is flowing (i.e. folds are open)
T	Total duration of each vibrational cycle
U_o	Average rate of air flow
U_{ac}	Maximum rate of flow

Figure 2.3: Two cycles of a typical glottal flow waveform (from [34] p.127)

In addition to the variables shown above, two other useful variables can be derived which define how the waveform is shaped [34]:

- Q_o - the open quotient, which is equal to T_o/T. This can be expressed simply as the percentage of time in each cycle during which the folds are open.
- Q_s - the skewing quotient, which is equal to T_p/T_n. This is the portion of time in each cycle during which the folds are moving outward, divided by the time in which they are moving inward. More simply, the skewing quotient is a number that explains how far from symmetric (how skewed) the waveform peak is.

The shape of the glottal waveform helps to determine the loudness of the sound generated, and also its timbre (sound quality). A 'jagged' waveform (with a large skewing quotient or small open quotient) which represents sudden changes in airflow will produce more high frequencies, and result in a 'brassy' timbre. A smoother curve, representative of gradual changes in airflow (open quotient and skewing quotient both high) will tend to produce a more 'breathy' sound [34].

A vocalist can control the open quotient (T_o) by moving the vocal processes (tips of arytenoid cartilages at one end of each vocal fold) closer together or further apart during phonation.

The fundamental repetition frequency of the waveform (T), which determines the pitch of the sung note, is controlled by the mass and tension of the vibrating areas of the vocal folds. Some of the characteristic differences between the male and female voice can be attributed to men in general having larger and more massive vocal folds than women, affecting the fundamental pitch in speech and the range of pitch in singing. At a low pitch the vocalist’s folds are relatively slack with almost the whole area vibrating. A rise in pitch is achieved by increasing the muscular tension on the folds, which become longer and thinner.

The movement of the vocal folds is more complex than that of a simple vibration. Not only do the folds move sideways, laryngoscopic studies using high-speed recordings and stroboscopic imaging have revealed a phase lag between movement of the upper and lower part of the folds in a wave-like motion (often called the mucosal wave). The folds have been shown to vibrate in many different normal modes in the same way as stretched strings or membranes of musical instruments [35]. However, a greater understanding of the mechanics and acoustics of the larynx is particularly difficult due to its inaccessibility and the non-uniformity and variability in time of its component parts [32].

The frequency spectrum of the glottal waveform shown in Figure 2.3 consists of a set of harmonic components whose amplitudes rapidly decrease with increasing frequency. Examples of spectral slope envelopes on such a spectrum are shown in Figure 2.4.

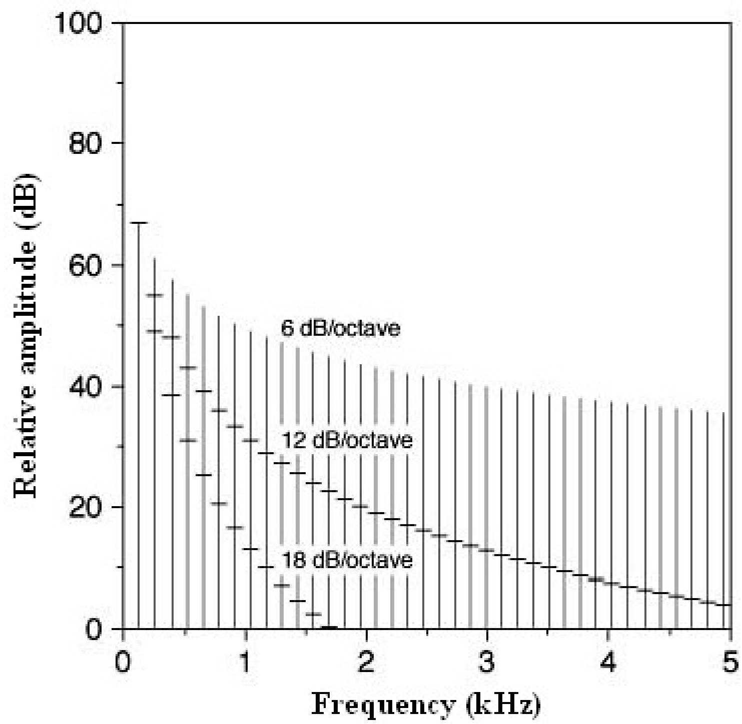


Figure 2.4: Examples of spectral slope envelopes on a harmonic (glottal) frequency spectrum (after Titze [34] p.130)

The influence of the glottal waveform shape on the timbre of the sound can be deduced from the spectral slope envelopes. In Figure 2.4, the middle slope depicts a fall of 12 dB/octave over the glottal spectrum which is considered the rate for normal vocal quality. At either side of this, a spectral slope of around 6 dB/octave, the least severe slope, results in stronger high frequencies and gives a more 'brassy' sound, whilst the most severe slope of 18 dB/octave would result in a more 'breathy' sound as it has stronger low frequencies compared to higher frequencies, which rapidly drop off in strength.

2.3.1 Vocal fold modelling

The most common approach to the modelling of vocal fold vibration behaviour treats the folds as mass-spring systems. The mass represents the body of the fold, whilst the spring portrays the tissue stiffness or restoring force in the vocal fold. More recently, damping constants have been introduced to represent the viscosity (energy absorption) of the tissue. Ishizaka and Flanagan [36] were the first to successfully capture the out-of-phase motions of the upper and lower surface of the folds using two-mass models of each fold, simulated by separate masses coupled by a spring (shown in Figure 2.5). Although the two-mass model can produce a glottal waveform for use in a voice model, it is still only a very basic approximation of the physiological system and is unable to model higher modes of oscillation.

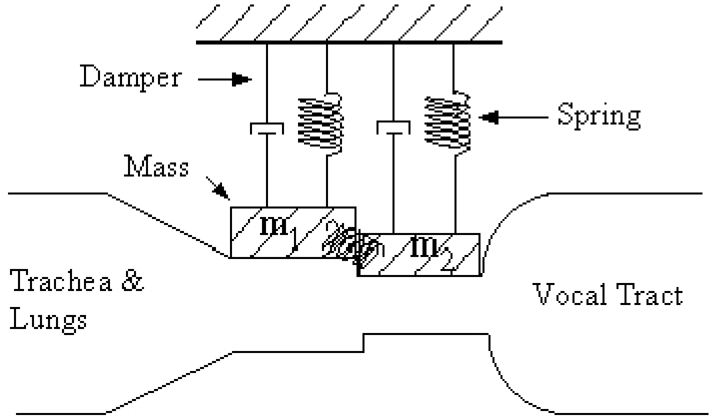


Figure 2.5: Two mass model of the vocal folds presented by Ishizaka and Flanagan [36]

Multiple mass coupled systems have been shown to simulate more complex vibration modes and are also able to mimic the disruption to the vibratory pattern of certain pathologies such as vocal fold nodules and edema [5, 37, 38]. However, the better a model imitates the physiological behaviour of the vocal folds, the more complex the calculations become and the number of parameters increase rapidly. Currently real-time calculations can only be achieved with simple models, therefore at present there is a compromise between the complexity of the model and its application in a synthesis system.

A comprehensive review of vocal fold models can be found in Kob [5].

2.4 Vocal tract resonances

Returning to the source-filter theory described previously, Fant [33] proposed that human vocal resonators act as an acoustical lens. As the sound wave leaves the vocal folds, it is filtered by the vocal tract resonance. This resonance imprints its own acoustical "personality" on the sound's harmonic spectrum. Based on the size, shape, and density of the walls of the enclosures, resonators can have one of four effects on the harmonic spectrum of the source sound. The filter can:

- Amplify
- Attenuate
- Nullify
- Pass the sound through unaltered

In the singing of most vowels, the velum closes off the entrance to the nasal cavity, therefore the vocal tract can be thought of as a long thin tube with a cross-section that varies considerably with its length (through the larynx, pharynx and mouth). The length of the vocal tract is typically 17.5cm in men and 15cm in women, although considerable variations occur. The resonances of such a tube depend strongly on the bore profile. If we simplify the vocal tract and consider it to be a cylindrical tube of constant cross-section closed at the larynx and open at the lips (referred to as a closed tube), we can calculate the resonances of the tube. The neutral schwa vowel [ə] matches closely this representation whereby the tongue is neither high nor low, nor is it placed to the front or the back, and in effect makes the cross section of the vocal tract approximately constant from glottis to lips.

Each resonance will have a standing wave pattern with a pressure antinode (a velocity node) at the closed end representing the larynx, and a pressure node (a velocity antinode) at the open end representing the lips. The formula for the resonance frequencies of a closed tube is [39]

$$F_n = \frac{c(2n-1)}{4L} \tag{2.1}$$

where n is the resonance number (an integer), c is the speed of sound in air, and L is the effective length of the vocal tract. Taking the vocal tract length of an adult male to be 17.5cm and approximating the speed of sound in air as $350ms^{-1}$, the first resonance frequency will be $F_1 = c/0.7 = 500\,\text{Hz}$. From equation (2.1), higher resonances will have frequencies which are odd integer multiples of F_1, therefore $F_2 = 1500\,\text{Hz}$, and $F_3 = 2500\,\text{Hz}$.

This is a simplified case and in general the relationship between the shape of the vocal tract and the formant pattern is a complex one [39]. Altering the shape, for example by moving the tongue, affects each resonance of the vocal tract in a different way. This will be explored in more detail in Chapter 5 where the acoustic and articulatory description of vowels is discussed.

2.4.2 Vocal tract modelling

The bore of the vocal tract is essentially tubular, therefore an accurate digital waveguide model [40] can be derived by approximating the bore profile using cylindrical segments. The classical approach for constructing an articulatory model of the vocal tract is the Kelly-Lochbaum model [41] (also referred to as a cylinder model or waveguide model). This involves approximating the bore of the vocal tract using cylindrical tubes of varying cross-sections from the glottis to the mouth or nose opening. Since the transitions between adjacent cylindrical segments represent discontinuities in the area function, the number of segments must be sufficiently high for an accurate approximation of the real vocal tract [5]. By using waveguides or digital filters, the propagation of an acoustic wave between cylindrical segments can be calculated. Figure 2.6 shows a cylindrical tube approximation of the vocal tract for a simulated vowel [u]. In Chapter 3, the propagation of plane waves in cylindrical tubes and the reflection and transmission of the waves at discontinuities is described. Cylindrical segments are used for ease of computation. However, refinements to the cylindrical modelling approach have been presented that use conical tube sections and fractional delay filtering techniques [42].

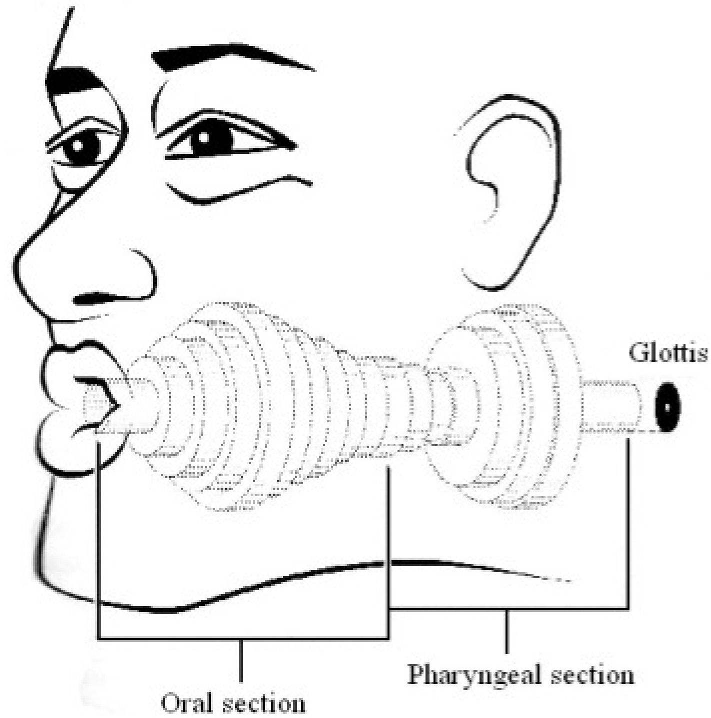


Figure 2.6: Cylindrical tube approximation of the vocal tract for a simulated vowel [u] (after Titze [34] p.151)

The underlying basis of all these modelling techniques requires ‘real’ geometric data of the vocal tract so that a discretised model can be implemented. As discussed in Chapter 1, airway and vocal tract imaging is inherently difficult. Traditional methods have involved calculating cross-sectional areas from X-rays, and more recently CT and MRI scans. However, access to such equipment is expensive and these methods have limitations on their applicability to vocal tract measurement. The focus of this thesis is in the development of a non-invasive acoustic technique that is able to measure the bore profile of the vocal tract for static targets of articulation (vowel sounds) which can then be used in discretised models for simulating the vocal tract in voice synthesis applications.

2.5 Differences between speech and singing

Speech and singing clearly have much in common, and the modern understanding of speech production is well surveyed by Fant [33] and Flanagan [43]. For example, in speech, and more especially in singing, there is an art of breathing. Ordinary inspiration and expiration necessary for the oxygenation of the blood is performed automatically and subconsciously. However, in singing, the respiratory apparatus is used like the bellows of a musical instrument, and it is controlled and directed by the will of the singer. The art of breathing properly is fundamental for the proper production of the singing voice and the speaking voice of the orator. It is necessary always to maintain in the lungs a sufficient reserve of air to finish a phrase, therefore when the opportunity arises it is desirable to take in as much air as possible through the nostrils, and without any apparent effort. The expenditure of the air in the lungs must be controlled and regulated by the vocalist in such a manner as to produce efficiency in loudness with economy of expenditure [44].

There are significant differences between speech and singing in terms of their production and perception by humans. In singing, for example, the intelligibility of the phonemic message is often secondary to the intonation and musical qualities of the voice. Vowels are often sustained much longer in singing than in speech, and precise, independent control of pitch and loudness over a large range is required [45]. The time ratio of voiced/unvoiced/silent phonation is approximately 60%/25%/15% in speech, compared to the nearly continuous 95% voice time of singing [46]. There is often an intentionally introduced deviation in the voice pitch in singing (singer's vibrato), and there can on occasions be the acoustical phenomenon of the singer's formant (grouping of the third, fourth, and sometimes fifth formants together for increased resonance [47, 48]).

A more detailed examination of the differences between speech and singing can be found in Cook [49].

The motivation to investigate the synthesis of the singing voice stems from the different priorities between it and that of speech synthesis. Speech involves complex articulations to form acoustical signals that we can recognise as words. Conversely, whilst singing can involve shaping the voice to form words, other types of voice instrumental music which use open sounds such as vowels or nonsense syllables ("vocables") also exist. Therefore, a synthesis model which negates complex articulations is still a valid model of the singing voice.

The work in this thesis involves measuring the vocal tract bore for static targets of articulation, such as the sung vowel sound. It is envisaged that the corpus of data collected here can be implemented into future voice and in particular, singing voice models.

2.6 Discussion

The basic principles of voice production and differences between speech and singing have been described in this chapter. The source-filter theory of speech production has been used to describe the acoustics of the voice and it has been shown how the functional components of the voice system can be treated as being independent of each other. The translation of physical principles and observed phenomena of the vocal fold and vocal tract systems into models which simulate their behaviour has been presented.

The focus of this work is to develop a non-invasive acoustic technique to measure vocal tract profiles in the phonation of vowel sounds. By making equivalent tube models of the vocal tract, the propagation of an acoustic wave through the cylindrical segments can be readily calculated. When used in combination with a vocal fold model, a physical model for voice synthesis can be realised.

In the next chapter, the nature of sound and the propagation of plane waves within varying cross-section tubular objects is discussed.

Chapter 3

Acoustic Pulse Reflectometry

3.1 Introduction

Acoustic pulse reflectometry (APR) was introduced in Chapter 1 as a tool for non-invasive measurement of the input impulse response of a tubular object, from which the internal bore dimensions can be calculated. Before proceeding with experimental measurements of test objects it is first necessary to understand the basic principles of sound propagation and how the input impulse response is created through acoustic wave reflection and transmission within an object. The following will examine how a sound wave is reflected if it meets a single discontinuity compared to multiple discontinuities (varying cross-sectional areas, as in biological cavities such as the vocal tract) and how this input impulse response can be used, together with the layer-peeling algorithm, to reconstruct the bore profile of the object or cavity.

3.2 The nature of sound

Sound is the propagation of energy by a mechanical wave through matter, which obeys Newtonian mechanics and requires a mass containing medium to support its transmission [50]. The structure of the medium will determine both the velocity and wave characteristics of the transmitted wave. A sound wave incident upon this medium will cause a displacement of the particles of the medium, so that they move back and forth in relation to one another. This longitudinal motion causes periodic alternate crowding and un-crowding of adjacent particles, creating regions where particles are compressed together (compressions) and other regions where the particles are spread apart (rarefactions). The distance between two bands of compression or rarefaction of the particles determines the wavelength of the mechanical wave in a particular supporting medium.

Sound waves are introduced into a medium by displacing particles from their rest state (e.g. by the vibration of an object). The disturbance then travels from particle to particle through the medium, transporting energy as it moves. The amount of energy which is transported past a given area of the medium per unit of time is known as the intensity of the sound wave, therefore the greater the amplitude of vibrations of the particles of the medium, the greater the rate at which energy is transported through it, producing a more intense sound wave.

The intensity of a sound wave constantly decreases as it travels through a medium such as air or tissue. This decrease in intensity is called attenuation and is due to the following three factors:

1. **Divergence of the sound wave;**

As a sound wave diverges, the energy of the sound wave is being distributed over a larger cross-sectional area. i.e. as energy is conserved and the area through which this energy is transported increases, the intensity must decrease. Therefore intensity is proportional to energy per unit area and decreases in proportion to the divergence of the sound wave (obeying an inverse square relationship).

2. **Absorption of sound energy by the medium (e.g. air or tissue);**

Absorption is the transfer of energy from the sound wave to the medium. The energy is spent overcoming the internal frictional forces of the particles within the medium and ultimately degrades into heat production.

3. **Deflection of sound out of the wave;**

The greater the sound frequency for a given amplitude, the more rapidly the particles within the medium move and the greater the energy expended in overcoming friction. Thus absorption is proportional to frequency. In addition to frequency, the amount of absorption depends on the viscosity of the medium through which the sound travels. In general, the more rigid the medium, the greater the absorption.

In addition, the path of sound waves depends upon the surface it meets. Sound waves are reflected by smooth surfaces and scattered by rough surfaces. A surface is "smooth" if the size of irregularities is small relative to the wavelength and "rough" otherwise [51].

The simplest mode of propagation of sound is in the form of plane waves. The fundamental (plane wave) mode propagates at all frequencies and travels in an axial direction with wavefronts that are uniform across the cavity cross-section. Higher order modes reflect off the cavity as they travel along its length and consequently have non-uniform pressure distributions across the cavity. Each mode has an associated cut-off frequency below which it rapidly attenuates, so does not propagate. Above its cut-off frequency, a mode will propagate. The cut-off frequency corresponding to the first nonplanar mode is $\omega_c = 1.84c/r$ for an air-filled cylindrical cavity, where c is the speed of sound and r is the radius of the tube. Expressed in Hertz, this is approximately $f_c = 100/r$ [52]. Therefore, if the bandwidth of the input signal is greater than the lowest cut-off frequency of the cavity (which may vary along its length), a mixture of plane waves and higher order modes will propagate.

In the case of the vocal tract, higher order modes do propagate but, at present, the bore reconstruction algorithm can only deal with plane waves. It is acknowledged therefore that the bore reconstructions may not be completely accurate due to the fact that energy is lost to higher order modes which could result in an underprediction of the cavity cross-section. Research to include higher order modes in the bore reconstruction algorithm is currently ongoing [31].

To satisfy the conditions of the bore reconstruction algorithm, assumptions being made in the present work are plane wave propagation and rigid airway walls.

3.3 Sound wave propagation in air

Acoustic waves in air are longitudinal in nature, meaning that the motion of the air particles transmitting the wave is parallel to the direction of propagation of the wave. For one-dimensional propagation in the *x*-direction, the wave equation in terms of the pressure *p* is expressed as [53]

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{3.1}$$

where *c* is the wave velocity (≈343 m/s) and *t* is the time. A solution of this equation has the form

$$p(x,t) = p^+(x,t) + p^-(x,t) \tag{3.2}$$

where

$$p^+(x,t) = \alpha e^{i(\omega t - kx)} \tag{3.3}$$

$$p^{-}(x,t) = \beta e^{i(\omega t + kx)} \tag{3.4}$$

where $p^{+}(x,t)$ is the forward travelling pressure wave with amplitude α, and $p^{-}(x,t)$ is the backward travelling pressure wave with amplitude β. The angular frequency, ω, is given by $\omega=2\pi f$, where f is the frequency of the wave in Hertz, and $k=\omega/c$ is the wave number. The equation of wave motion for air relating the particle velocity u to the acoustic pressure p is

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho}\frac{\partial p}{\partial x} \tag{3.5}$$

where ρ is air density. Solving this equation, the particle velocity is

$$u(x,t) = \frac{1}{\rho c}\left[\alpha e^{i(\omega t - kx)} - \beta e^{i(\omega t + kx)}\right] \tag{3.6}$$

In terms of forward and backward components from equations (3.3) and (3.4),

$$u^{+}(x,t) = \frac{p^{+}(x,t)}{\rho c} \tag{3.7}$$

$$u^{-}(x,t) = -\frac{p^{-}(x,t)}{\rho c} \tag{3.8}$$

The ratio of the acoustic pressure p to the velocity u is known as the specific acoustic impedance Z_s, which is

$$Z_s = \frac{p}{u} \tag{3.9}$$

Combining the forward and backward components of particle velocity in equations (3.7) and (3.8), with equation (3.9), for plane wave propagation

$$Z_s = \frac{p}{u} = \pm \rho c \tag{3.10}$$

with the positive value applying to waves travelling in the +*x* direction and the negative value applying to waves travelling in the –*x* direction. The product on the right hand side of equation (3.10) is defined as the characteristic impedance of the medium (in this case air), and is given as

$$Z_c = \rho c \tag{3.11}$$

3.4 Sound wave propagation in tubular structures

In the previous section we described one-dimensional plane waves propagating in open space where there was no confinement of the wave in the plane perpendicular to the direction of propagation. Here we describe acoustic wave propagation in a tube.

The wavefronts of a sound wave are restricted by the diameter of the tube, therefore it is necessary to take into account the cross-sectional area of propagation. Here we introduce the volume velocity $U = Au$, where A is the cross-sectional area of the tube. The acoustic impedance at any cross-section in the tube is defined as the ratio of the pressure and the volume velocity

$$Z = \frac{p}{U} \tag{3.12}$$

Combining equations (3.10) and (3.12), the acoustic impedance at a cross-section of area A is given by

$$Z = \frac{p}{Au} = \pm\frac{\rho c}{A} \tag{3.13}$$

with dimensions $kg\, s^{-1}\, m^{-4}$. The impedance for forward travelling waves is $+\rho c / A$ and for backward travelling waves is $-\rho c / A$ where $\rho = 1.21\, kg\, m^{-3}$ is the equilibrium density of air. This concept of wave impedance can be used to discuss reflections of sound from interfaces among various media.

3.4.1 Reflection from a single discontinuity

A sound pressure wave does not disappear when it reaches the end of its medium, or is interrupted by an obstacle. It will perform one of the three behaviours: reflection, diffraction, or transmission. The process of wave reflection can be defined as the return of all or part of a sound wave when it encounters a change in impedance. This may be due to a change of medium or due to a change in area. Therefore, when a planar acoustic wave propagating in an air-filled tube encounters a change in cross-sectional area, the associated change in characteristic impedance causes partial reflection and partial transmission of the incident wave.

In order to understand how acoustic reflection is used to obtain an area distance function, we can examine the simple case of a straight tube with a single change in cross-sectional area. Figure 3.1 shows a junction between two discontinuously joined cylindrical tubes (termed a scattering junction), one with cross-sectional area A_1 and the other with cross-sectional area A_2. The incident and reflected waves in the first tube are represented by $p_1^+(x,t)$ and $p_1^-(x,t)$ respectively while $p_2^+(x,t)$ is the transmitted wave in the second tube. Both tubes are assumed to be semi-infinite so there will be no reflections returning from either ends of the tube.

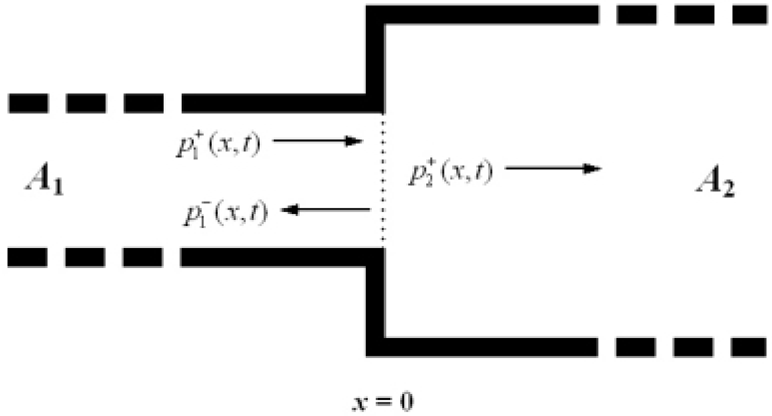


Figure 3.1: Scattering junction – reflection from single boundary between two discontinuously joined cylindrical tubes (adapted from Sharp [13]).

The incident and reflected waves in the first cylindrical tube are

$$p_1^+(x,t) = \alpha_1 e^{i(\omega t - kx)} \tag{3.14}$$

$$p_1^-(x,t) = \beta_1 e^{i(\omega t + kx)} \tag{3.15}$$

while the transmitted wave in the second cylindrical tube can be written as

$$p_2^+(x,t) = \alpha_2 e^{i(\omega t - kx)} \tag{3.16}$$

The pressure and volume velocity are continuous at the junction between the tubes, therefore, at $x = 0$, the following continuity equations apply

$$p_1^+(x,t) + p_1^-(x,t) = p_2^+(x,t) \tag{3.17}$$

$$\frac{p_1^+(x,t)}{Z_{c1}} - \frac{p_1^-(x,t)}{Z_{c1}} = \frac{p_2^+(x,t)}{Z_{c2}} \tag{3.18}$$

where $Z_{c1} = \dfrac{\rho c}{A_1}$ is the characteristic impedance of the first cylindrical tube and $Z_{c2} = \dfrac{\rho c}{A_2}$ is the characteristic impedance of the second cylindrical tube. Combining equations (3.17) and (3.18),

$$\frac{p_1^-(x,t)}{p_1^+(x,t)} = \frac{Z_{c2} - Z_{c1}}{Z_{c2} + Z_{c1}} = \frac{A_1 - A_2}{A_1 + A_2} \tag{3.19}$$

Examination of equations (3.14) and (3.15), reveals that, at $x = 0$, the ratio of the instantaneous pressures of the reflected and incident waves can be simplified as the ratio of their amplitudes

$$\frac{\mathrm{p}_1^-(x,t)}{\mathrm{p}_1^+(x,t)} = \frac{\beta_1}{\alpha_1} \tag{3.20}$$

Therefore, combining equations (3.19) and (3.20) yields the reflection coefficient $r_{1,2}$ (the ratio of the pressure amplitude of the reflected wave to that of the incident wave) for the junction boundary between the first and second cylindrical tubes

$$r_{1,2} = \frac{\beta_1}{\alpha_1} = \frac{p_1^-(0,t)}{p_1^+(0,t)} = \frac{A_1 - A_2}{A_1 + A_2} \tag{3.21}$$

An increase in area will give a negative reflection coefficient, a decrease in area will give a positive reflection coefficient, and there is no reflection when $A_1 = A_2$. The transmission coefficient $t_{1,2}$ (the ratio of the pressure amplitude of the transmitted wave to that of the incident wave) can be expressed in terms of the reflection coefficient

$$t_{1,2} = 1 + \frac{p_1^-(0,t)}{p_1^+(0,t)} = 1 + r_{1,2} = \frac{2A_1}{A_1 + A_2} \tag{3.22}$$

Thus, the reflection and transmission coefficients between adjacent tubes depend only on the change of cross-sectional area.

3.4.2 Multiple reflections from multiple discontinuities

The previous section detailed the simple case of a pressure wave incident on a single discontinuity created by a change in the cross-sectional area of a cylindrical tube. In most practical applications however, the object under investigation will usually involve varying cross-section (multiple discontinuities), therefore an impulse incident on the object will give rise to multiple reflections. It is this case that will be considered here. Plane wave propagation is again assumed.

From the discussions on vocal tract modelling in Chapter 2, we have seen that any tubular object whose cross-sectional area varies with its length (axial distance) can be approximated (discretised) by a series of discontinuously joined cylindrical segments. This is the case with the vocal tract for which any complicated vocal tract shape can be modelled using cylindrical tubes. Figure 3.2 shows the first few sections of a discretised model connected to a source tube.

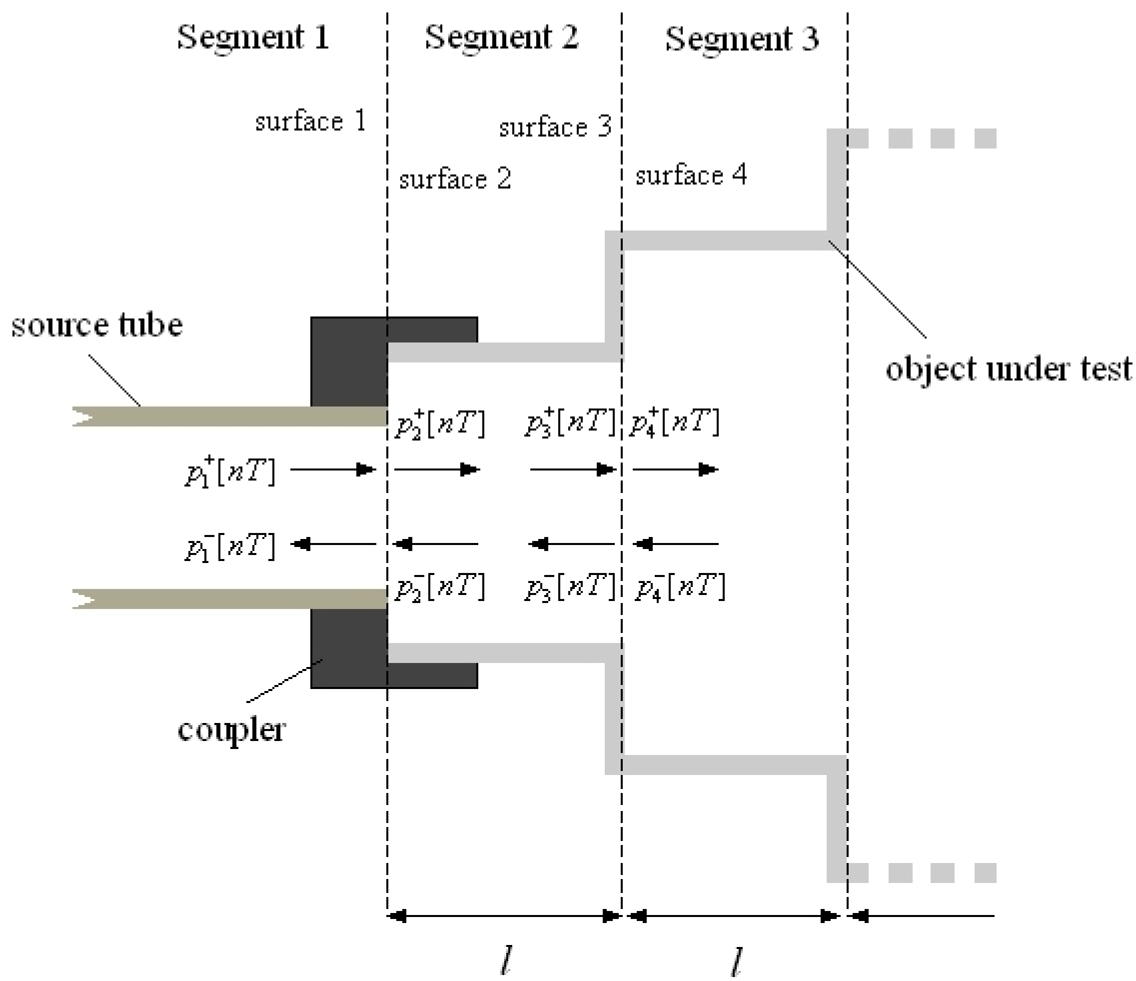


Figure 3.2: Pressure waves at boundaries between multiple segments of a cylindrically segmented object (adapted from [13, 31]).

The length of the cylindrical sections is $l = cT/2$ [23] where $T = 1/F_s$ is the sample period and F_s is the sample rate such that the time for travel from the left hand side of a cylinder to the right hand side, reflection off the discontinuity and return to the left hand side will correspond to one sample in the time domain [31].

At the junctions between segments, pressure waves will experience partial reflection and partial transmission. If we designate surface 1 as the plane at the end of the source tube and adopt the same notation as before, the forward and backward going waves at this point are labelled p_1^+ and p_1^- respectively. Surface 2 is then the plane on the other side of the discontinuity (change in cross-sectional area). Surface 3 is a distance l away on the right of the first cylindrical segment with surface 4 the plane immediately on the other side of the next discontinuity, and so on.

The input impulse response (*iir*) is defined as the sequence of reflections which return from the object under test when an ideal delta function impulse (δ) is fed into the input. By taking $t = 0$ as the time of arrival of the input pulse at the input plane, the forward going wave on surface 1 is then an impulse

$$p_1^+[nT] = \delta[nT] = \begin{cases} 1 & : \quad n = 0, \\ 0 & : \quad n \geq 1 \end{cases} \tag{3.23}$$

n is an integer running from 0 to N-1 where N is the total number of samples taken. The signal which returns from the object is the input impulse response

$$p_1^-[nT] = iir[nT] \tag{3.24}$$

Therefore, ignoring the effect of losses experienced by the signal whilst propagating through the cylindrical segments, the input impulse response $iir[nT]$ of the object is composed of a series of returning impulses, spaced a time T apart.

3.5 Bore reconstruction

The essence of the APR technique relies on solving the 'Inverse Problem', which involves calculating an object's dimensions (reconstructing the bore profile) from the measured input impulse response. Solving the inverse problem for bore reconstruction was first performed using the algorithm developed by Ware and Aki [12] and has been reviewed in detail in [7, 13]. The Ware-Aki algorithm however, does not compensate for the effect of losses when reconstructing bore profiles, therefore the lossy reconstruction (layer-peeling) algorithm developed by Amir et al. [27], as used by Sharp, Kemp and Li [13, 29, 31, 54], has been adopted in this present study.

3.5.1 Layer-peeling

Returning to the case shown in Figure 3.2 where the source tube of a reflectometer is attached to an object. If the incident wave injected into the duct is an acoustic impulse ($p_1^+[nT] = \delta[nT]$), then the reflected wave is the input impulse response ($p_1^-[nT] = iir[nT]$). At time $t = 0$ (defined as being the instant that the incident wave arrives at the entrance to the object – the boundary between the first and second segments), there cannot be any backward travelling waves in the second segment incident on surface 2. That is, $p_2^-[nT]$, $p_3^-[nT]$ and $p_4^-[nT]$ are all zero. This is a consequence of the property of causality which states that no backward travelling wave can be present in a segment before a forward travelling wave has reached that segment [13]. Consequently, the backward travelling wave in the first segment is simply the reflection of the forward travelling wave in that segment. From equation (3.21) the reflection coefficient at the boundary between segments 1 and 2 is

$$r_{1,2} = \frac{p_1^-[0T]}{p_1^+[0T]} = iir[0T] \tag{3.25}$$

where $p_1^+[0T]$ and $p_1^-[0T]$ are the incident and reflected waves and $iir[0T]$ is the input impulse response, all at $t = 0$.

By rearranging equation (3.21), the cross-sectional area A_2 can be expressed in terms of the area A_1 and the reflection coefficient $r_{1,2}$

$$A_2 = A_1 \left(\frac{1 - r_{1,2}}{1 + r_{1,2}} \right) \tag{3.26}$$

Therefore, knowing the cross-sectional area A_1 of the first cylindrical segment (the source tube), the cross-sectional area A_2 of the second segment can be found from the reflection coefficient $r_{1,2}$ using equation (3.26). From here we can proceed to calculate the resulting forward and backward travelling waves at this point. The forward travelling wave emerging from surface 2 is equal to the sum of the transmission of the forward travelling wave incident on surface 1 and the reflection of the backward travelling wave incident on segment 2. Similarly, the backward travelling wave on segment 1 is equal to the sum of the transmission of the backward travelling wave on segment 2 and the reflection of the forward travelling wave on segment 1. This can be expressed in matrix notation as follows

$$\begin{pmatrix} p_2^+[nT] \\ p_1^-[nT] \end{pmatrix} = \begin{pmatrix} t_{1,2} & r_{2,1} \\ r_{1,2} & t_{2,1} \end{pmatrix} \begin{pmatrix} p_1^+[nT] \\ p_2^-[nT] \end{pmatrix} \tag{3.27}$$

$r_{2,1}$ represents the reflection coefficient from the discontinuity between segments 2 and 1 when waves are incident from segment 2 only. This can be found by interchanging the subscripts in equation (3.21) giving

$$r_{2,1} = \frac{A_2 - A_1}{A_1 + A_2} = -r_{1,2} \tag{3.28}$$

$t_{1,2}$ represents the transmission coefficient from the discontinuity between segments 1 and 2 when waves are incident from segment 1 only. From equation (3.22) this was shown as $t_{1,2} = 1 + r_{1,2}$. By interchanging the subscripts of this equation, $t_{2,1}$ is the transmission coefficient from the discontinuity between segments 2 and 1 when waves are incident from segment 2 only

$$t_{2,1} = \frac{2A_2}{A_1 + A_2} = 1 + r_{2,1} = 1 - r_{1,2} \tag{3.29}$$

The forward and backward travelling pressure waves at the left side of the second segment (to the right of the discontinuity), $p_2^+[nT]$ and $p_2^-[nT]$, can be obtained from the forward and backward travelling waves on the right side of the first segment (to the left of the discontinuity) by rearranging the simultaneous equations in equation (3.27), to give

$$\begin{pmatrix} p_2^+[nT] \\ p_2^-[nT] \end{pmatrix} = \frac{1}{1-r_{1,2}} \begin{pmatrix} 1 & -r_{1,2} \\ -r_{1,2} & 1 \end{pmatrix} \begin{pmatrix} p_1^+[nT] \\ p_1^-[nT] \end{pmatrix} \tag{3.30}$$

This equation is performed for all values of n from 0 to N-1.

Following on, the pressure waves at the right side of the second segment must be determined. The forward travelling wave at the right side of the second segment ($p_3^+[nT]$) is found simply by adding a delay of $T/2$ to the forward travelling wave on the left of the segment ($p_2^+[nT]$), to account for the time taken to travel a distance of l.

$$p_3^+\left[\left(n+\frac{1}{2}\right)T\right] = p_2^+[nT] \tag{3.31}$$

Similarly, the backward travelling wave at the right side of the second segment ($p_3^-[nT]$) is found by subtracting a delay of $T/2$ from the backward travelling wave on the left of the segment ($p_2^-[nT]$).

$$p_3^-\left[\left(n-\frac{1}{2}\right)T\right] = p_2^-[nT] \tag{3.32}$$

At the boundary between the segments 2 and 3, when $t=T/2$, there is no backward travelling wave in segment 3. Therefore, the reflection coefficient $r_{3,4}$ (the discontinuity between segments 2 and 3) is given by the ratio of the backward and forward travelling waves at the right side of segment 2.

$$r_{3,4} = \frac{p_3^-[T/2]}{p_3^+[T/2]} \tag{3.33}$$

Using the general form of equation (3.26), (expressing the (*j+1*)th segment in terms of the *j*th segment), the cross-sectional area of the third segment A_3 can be obtained from the previously calculated cross-sectional area of the second segment A_2 and the reflection coefficient $r_{3,4}$. This layer-peeling process is carried out recursively[2], calculating the cross-sectional area of each segment for the entire length of the object so that a bore profile can be produced.

3.5.2 Losses

The previous section represented the effect of sound waves propagating between each cylindrical segment as simply a delay of $T/2$. No account was taken of viscous thermal losses which can be significant within the bore of long objects, causing frequency dependent attenuation of the propagating wave through each segment. Experiments by Watson and Bowsher [24-26] on brass instruments in the 1980s using the APR technique and the Ware-Aki bore reconstruction algorithm (which does not take into account losses) showed a tendency to increasingly under predict the radius with the distance

[2] Numerical implementation issues of the algorithm are treated in detail in [31]

along the bore. Attempts to prevent this under prediction using manual adjustments of the input impulse response did not provide a rigorous method of compensating for the losses experienced by the object reflections. A way of including losses in the layer-peeling algorithm has been presented by Amir et al. [27] and is adopted here. Although the lossy layer-peeling algorithm is more effective when applied to objects of long bore, such as brass instruments which may be several metres in length, its application here to the case of the much shorter vocal tract is still beneficial. Also, due to the computing power afforded by modern PC's, the additional load of including losses during the bore reconstruction calculation is not significant and has little detriment to the overall speed of the calculation.

A derivation of the frequency domain formula for losses associated with propagation of plane acoustic waves down a tube of length L can be found in [55] and is given as

$$H(\omega) = e^{-\Gamma L} \tag{3.34}$$

where Γ is a complex wavenumber. Amir et al. [27, 28] gives the numerical implementation of the equivalent digital frequency domain lossy filter, and by inverse Fourier Transforming this, the digital filter $h_j[nT]$ is found. To represent the losses in each segment, the digital filter $h_j[nT]$ (which is dependent upon the length and radius of the segment) is applied in addition to the $T/2$ delay to each segment in the layer-peeling algorithm. For forward travelling wave propagation (from the left side to the right side of the segment), the wave is simply passed through the filter. Meanwhile for backward travelling waves (from the right side to the left side of the segment), the wave is passed through the inverse filter. The passage of the forward and backward travelling waves through the filters are represented by the following equations

$$p_j^+[nT] = p_j^+[nT] \otimes h_j[nT] \tag{3.35}$$

$$p_j^-[nT] = p_j^-[nT] \otimes^{-1} h_j[nT] \tag{3.36}$$

The operators $\otimes$ and $\otimes^{-1}$ represent convolution and deconvolution, and $h_j[nT]$ is the digital lossy filter in the *j*th segment. To include the effect of losses, these equations are inserted before equations (3.31) and (3.32) in the reconstruction procedure.

3.6 MLS Signals

The standard method for measuring experimentally the impulse response of a system involves injecting a pulse of short duration (usually microseconds) into the source tube and measuring the reflections from the object. These reflections are recorded by a microphone embedded in the wall of the source tube and are sampled by the computer's data acquisition board at a set sampling rate. The procedure is repeated 1000 times, storing the samples from each successive reflection sequence to the computer's hard drive. A delay is included between each repetition to ensure all the signal from the previous pulse no longer exists. The sampled reflections are then averaged to improve the signal-to-noise ratio by a factor of $\sqrt{1000}$ [13].

A development to the standard APR technique presented by Kemp [31] is to use an input signal that continues over a longer time interval than that of a single pulse. This enables more energy to be injected into the system, improving the signal to noise ratio and removing the need to average over a large number of measurements. The solution proposes the use of a pseudo-random binary signal called ‘maximum length sequences’ (MLS). An MLS signal consists of an apparently random sequence of 0’s and 1’s that has a flat frequency spectrum for all frequencies up to the Nyquist frequency with the exception of the DC value. Unlike other random noise excitation signals such as white noise (with which MLS shares similar spectral properties), an MLS signal can be generated deterministically from an m^{th} order primitive polynomial [56] (of maximum length 2^m-1), and is therefore repeatable.

The MLS is proved to correspond more closely than a single experimental pulse to the ideal Dirac (δ) impulse excitation [57], but in order to extract the input impulse response of the system, a cross-correlation procedure between the MLS input signal and the system output signal is required.

With these properties, coupled with the fact that they are a computationally efficient to implement, MLS signals have been employed frequently in a variety of fields, including building acoustics (measuring the input impulse response of rooms for reverberation measurement [58-60]) and duct leak detection [61].

An in-depth review of generating MLS signals and their application in APR is given in [31]. A summary is presented here.

3.6.1 Generating MLS Signals

MLS signals are generated using feedback binary shift registers [59, 62], each register consisting of a group of m binary elements in a line. At each time unit the numbers held in the memory elements are shifted one step to the right and the vacated element on the left is generated by a recursion relation which depends on the number of memory elements, m [31]. Cycling through this process, the shift registers reproduce every possible value for a binary number with m digits, with the exception of the zero DC term. The number of possible values for a binary number with m digits is 2^m therefore the maximum length for an MLS signal will be 2^m-1.

With the exception of the zero term, the initial state of the registers is unimportant to the properties of the MLS signal since the sequence is periodic and will cover all non-zero states. For generating an MLS, a list of primitive polynomials up to m = 168 is given by Stahnke [56] and are shown in recursion relation form up to m = 20 in [31].

As an example, taking a polynomial of m = 19, the length of the sequence is 2^{19}-1 = 524,287 binary samples, which corresponds to over 11.9 seconds of sound when played as an acoustic signal at 44.1 kHz. Compared with the standard APR method, the input energy of the MLS signal is more than 10^3 times greater than for a single pulse, therefore the need for signal averaging can be eliminated.

3.6.2 Autocorrelation property of MLS

Correlation determines the degree of similarity between two signals. If the signals are identical, then the correlation coefficient is 1; if they are totally different, the correlation coefficient is 0, and if they are identical except that the phase is shifted by exactly 180° (i.e. mirrored), then the correlation coefficient is -1. When two independent signals are compared, the procedure is known as cross-correlation, and when the same signal is compared to phase shifted copies of itself, the procedure is known as autocorrelation.

Autocorrelation is a mathematical tool used frequently in signal processing for analysing functions or series of values, such as time domain signals. It is useful for finding repeating patterns in a signal, such as in the present case of determining the presence of a periodic signal which has been buried under noise.

The autocorrelation function, ρ, is defined as [62]

$$\rho(i) = \frac{1}{n}\sum_{j=0}^{n-1} s_j s_{i+j}^{*} \qquad (3.37)$$

where $n = 2^m - 1$ is the length of the sequence. If the subscript $i + j$ exceeds n, the subscript is taken modulo n (i.e. n is subtracted so that s_{i+j} is a circularly shifted version of s). The symbol * denotes complex conjugation which can be dropped in the current application since all entries in the sequence are real, therefore the autocorrelation for MLS becomes

$$\rho(i) = \frac{A - D}{n} \tag{3.38}$$

where A is the number of times the elements s_j and s_{i+j} agree and D is the number of times they disagree. The first term in ρ is given by $i = 0$ so s_j and s_{i+j} agree for all values of j giving $\rho(0) = 1$. When $1 \le i \le n-1$, the calculation is of the agreement between the signal and a circularly shifted version. As the signals appear random, it can be perceived that the agreement and disagreement should be almost equal. The proof in [62] shows that $A - D = -1$ for $1 \le i \le n-1$. The autocorrelation of the MLS is therefore

$$\rho(i) = \begin{cases} 1 & : \quad i = 0 \\ -1/n & : \quad 1 \le i \le n-1 \end{cases} \tag{3.39}$$

which is distinguished from a perfect digital impulse by the presence of the small non-zero value when $i \neq 0$. While the frequency spectrum of an ideal digital impulse is equal for all frequencies, the frequency spectrum of ρ is the same for all frequencies except for the zero frequency component.

3.6.3 Extraction of system impulse response from MLS measurement

The frequency content of the signal recorded at the microphone will contain information on the frequency response of the system under test [31]. In order to extract the impulse response of the system, we must use the autocorrelation property of the MLS signal, since the measured signal is the convolution of the MLS and the system impulse response.

Consider an MLS signal $s(t)$, which is inherently periodic with a period of m^2-1, applied to an acoustic system with an input impulse response $h(t)$, which is not periodic. The result is that the output pressure signal $y(t)$ measured at the microphone is the periodic response of the system to continuous excitation by the periodic MLS. The time and frequency input/output relationships are [63]

$$y(t) = s(t) \otimes h(t) \tag{3.40}$$

$$Y(\omega) = S(\omega)H(\omega) \tag{3.41}$$

where $\otimes$ denotes convolution. It is important at this stage to realise that the system impulse response of a reflectometer is not simply the input impulse response of the object on the output end of the reflectometer; it includes the impulse response of the loudspeaker, the losses in the source tube, the input pulse passing the microphone on its way to the loudspeaker, the source reflection and so on [31]. In practice, the system impulse response, $h(t)$, that we have calculated is sliced to isolate the object reflections.

Performing correlation with respect to *s* on both sides of equation (3.40) gives [64]

$$\phi_{sy} = \phi_{ss} \otimes h = \rho \otimes h \tag{3.42}$$

where ϕ_{ab} is the notation for the correlation of *a* and *b*. Note that convolution in the time domain corresponds to multiplication in the frequency domain, so the fact that the frequency spectrum of ρ is flat, except for the zero frequency component, means that *h* is left unchanged by convolution with ρ except for a small DC offset of the order of $1/n$. The impulse response of the system can therefore be extracted from the measurement of the system response by correlation with the MLS input [31].

It can also be shown that cross-correlation of the two signals in the frequency domain is given by [31]

$$\mathrm{fft}(\phi_{\mathrm{sy}}) = \frac{1}{\mathrm{n}}(\mathrm{fft(s)})^{*} \times (\mathrm{fft(y)}) \tag{3.43}$$

Some precautions must be taken when using the MLS method:

- The MLS signal length must be longer than the impulse response of the system under test or have the same length. If these conditions are not satisfied, some parts of the computed impulse response will overlap (time-aliasing).

- The MLS only has perfect autocorrelation properties when it is periodic, so in a measurement system it is necessary to play two sequences, one to excite the object being measured and the second from which to make the calculation. This is referred to as the periodicity condition.
- The system under test must be time-invariant, at least during the measurement interval. This is important in the case of the human vocal tract which may experience articulator drift or involuntary muscle movement during the static imitation of vowels, therefore the measurement length must be kept suitably short for the subject to remain still and refrain from breathing. This issue will be discussed in detail in the following chapters.

3.7 Conclusions

In this chapter, the nature of sound has been described and the reflection and transmission of plane waves within varying cross-section tubular objects was discussed. The basic theory for solving the inverse problem of determining the dimensions of an object from its input impulse response has been presented. Using this reflection sequence, it has been shown how the layer-peeling algorithm can be used to reconstruct the bore of the object. A development to the standard APR technique proposed by Kemp [31] using MLS excitation signals has also been summarised.

In the next chapter, an experimental reflectometer is developed, using the enhancements to the standard APR technique, together with modifications for its application to measuring the human vocal tract.

Chapter 4

APR measurement using MLS excitation signals

4.1 Introduction

The fundamental theory of acoustic pulse reflectometry for measuring the internal bore of tubular objects was introduced in Chapter 3. In this chapter, the APR experimental apparatus developed by Sharp et al. [13] using the MLS excitation signal employed by Kemp [31] is described, together with modifications for its application to measuring the human vocal tract.

4.2 Apparatus

The principle of operation can be summarised as follows: An audible sound signal is generated in a source tube and transmitted into the object under investigation, in this case the vocal tract. The signal undergoes partial reflection and partial transmission at changes in impedance, caused by changes in cross-sectional area along the length of the cavity, creating a reflection sequence. The returning sequence travels back up the source tube without further reflection and is recorded by a microphone embedded in the wall of the source tube. Applying this reflection sequence (which at the input end to the cavity is termed the input impulse response) to suitable algorithms, the bore of the object can be reconstructed.

Implemented in the experimental reflectometer, the operation sequence is as follows: An MLS signal is generated by a 200MHz Pentium PC with a soundcard and is output to a Pioneer A-110 stereo amplifier which is connected to a JBL loudspeaker (2426J with 16Ω impedance). The emitted signal from the loudspeaker travels along a coupled 4.6m source tube into the object being measured. A Knowles microphone connected to the PC is embedded part of the way along the source tube and records the reflections from the test object. The signal from the microphone is sampled through the PC soundcard input using Matlab data acquisition software for Windows with a sample rate (F_s) of 44.1 kHz. A schematic diagram of the reflectometer used in this study is shown in Figure 4.1. A photograph of the apparatus is shown in Figure 4.2.

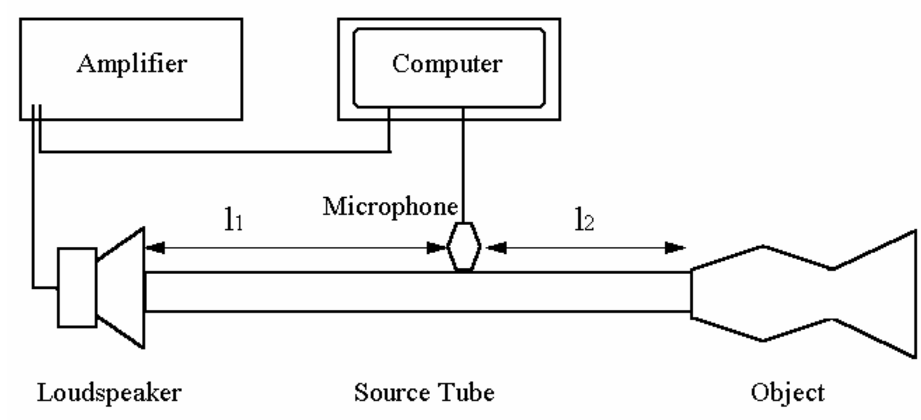


Figure 4.1: Schematic diagram of the APR apparatus

Figure 4.2: Photo of the APR source tube and loudspeaker

4.2.1 Loudspeaker/Source Tube Coupling

A tapered aluminium coupler was fabricated to couple the loudspeaker with the source tube (Figure 4.3).

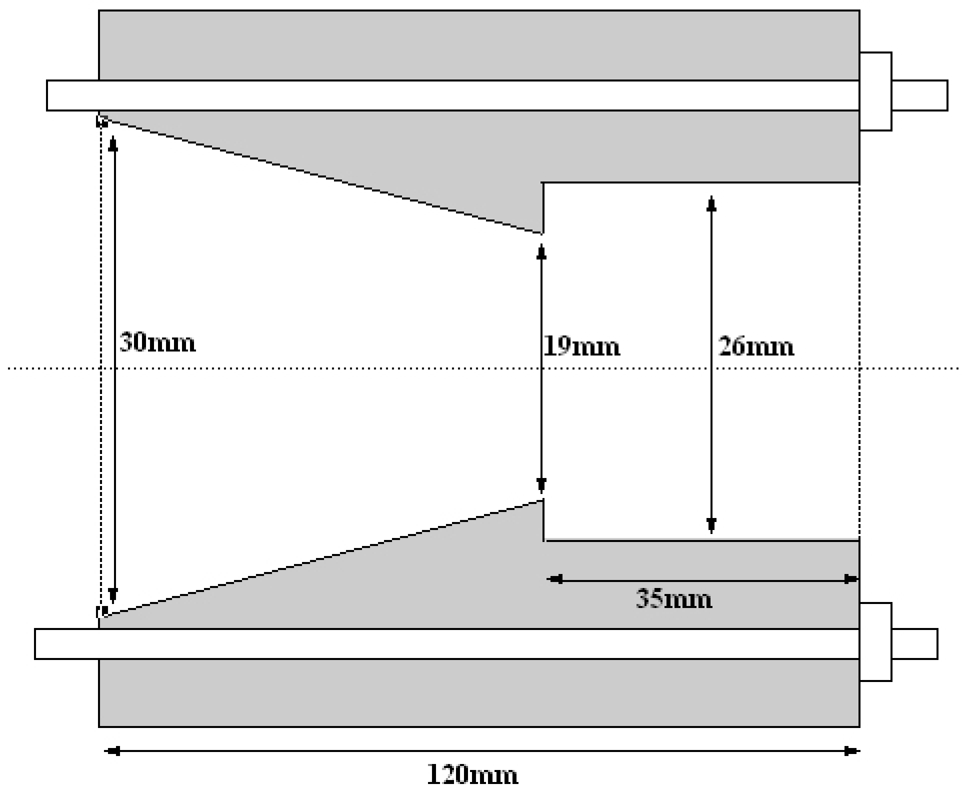


Figure 4.3: Schematic diagram of the coupler connecting the loudspeaker and source tube

The tapered end of the coupler fits flush with the internal bore of the source tube which negates a large discontinuity at the join and reduces any ‘ringing’ of the input signal. Long bolts ensure the coupler and loudspeaker are held together tightly.

4.2.2 Source Tube

The diameter of the source tube (although not critical) was chosen to achieve an approximate impedance match with a relaxed mouth opening. Commercial grade plastic tubing (BS6572 Medium Density Polyethylene - MDPE) with an internal diameter of 19.5mm and 3mm wall thickness was used throughout this work. A discussion on different source tube materials can be found in [7].

It is necessary to choose a suitable value for the source tube section l_2 to ensure that the input pulse has completely passed the microphone before the first reflections return from the object to the microphone. After the object reflections pass the microphone they are further reflected by the loudspeaker. The source tube section l_1 is needed to separate the object reflections from these source reflections. It ensures that once the object reflections reach the microphone, they can be recorded for up to $(2l_1)/c$ seconds (the time taken to travel the distance from the microphone to the loudspeaker and back) before the source reflections return and interfere with the signal.

In practice, several experiments were conducted to establish suitable lengths for l_1 and l_2. During these experiments, it became apparent that, with the initial placement of the microphone embedded in the pipe wall, the input pulse did not fully pass the microphone before the object reflections returned as l_2 (the distance between the microphone to the end of the source tube) was too short. The 'pulse width' (i.e. the width of the pulse that results once the correlation between the microphone signal and the MLS input signal has been carried out) was re-calculated and found to be longer than at first thought. The consequence of this meant that the original source tube was not long enough to accommodate the required lengths for l_1 and l_2. A new longer source tube with the same properties was required. By measuring the ‘pulse width’ of the correlated microphone signal and MLS input sequence, placement of the embedded microphone a distance l_1 = 2.60m ±0.01m away from the loudspeaker and a distance l_2 = 2.0m ±0.01m away from the end of the source tube/object coupling, was found to achieve good separation of the input signal and reflections. This resulted in a source tube length of 4.6m.

Due to the nature of the source tube material and wall thickness, it became apparent that although the pipe was flexible enough to be coiled into a large circle without internal bore deformation, it could not be easily manipulated. The elastic properties of the pipe also meant that it needed to be fixed to a supporting stand so that test objects and human volunteers could readily access the end of the source tube. Since the choice of source tube material proved challenging to work with, an acoustic measurement of the tube length was made to verify the manually measured length.

The open end of the source tube was blocked off by a cap and a MLS signal m = 18 was directed into the tube whilst the PC recorded the signal from the microphone (shown in Figure 4.4(a)). Following the procedure of correlating the microphone signal with the MLS input signal (discussed in section 3.6.3), the time difference between the ‘pulse peaks’ (i.e. the input signal recorded as it passes the microphone and its reflections off the cap and loudspeaker), could be measured (shown in Figure 4.4b). From this, the distance travelled (i.e. source tube length) can be calculated.

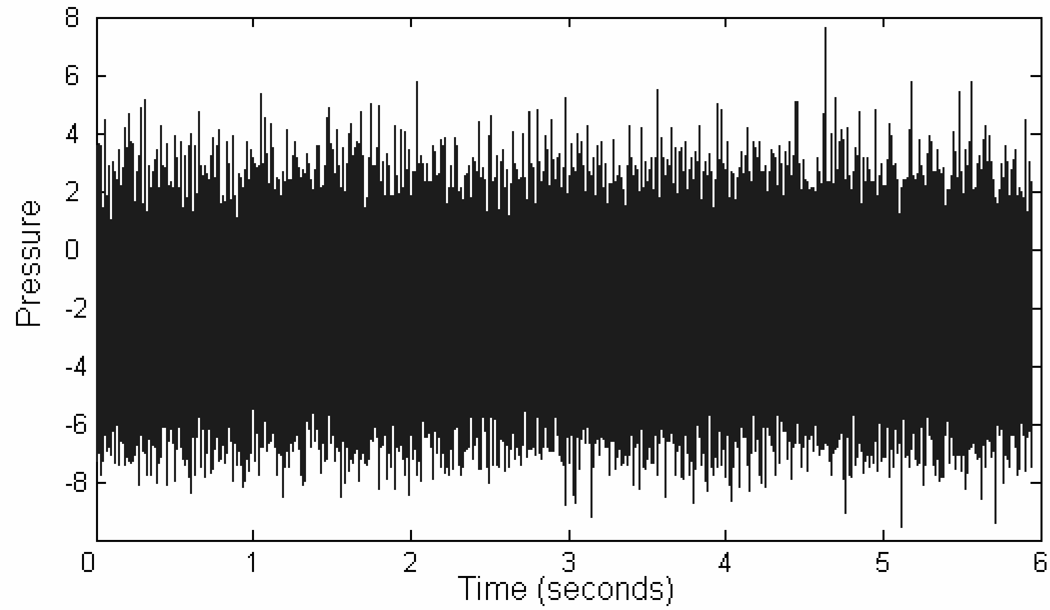


Figure 4.4(a): Microphone signal for excitation with a MLS m = 18 (single repetition)

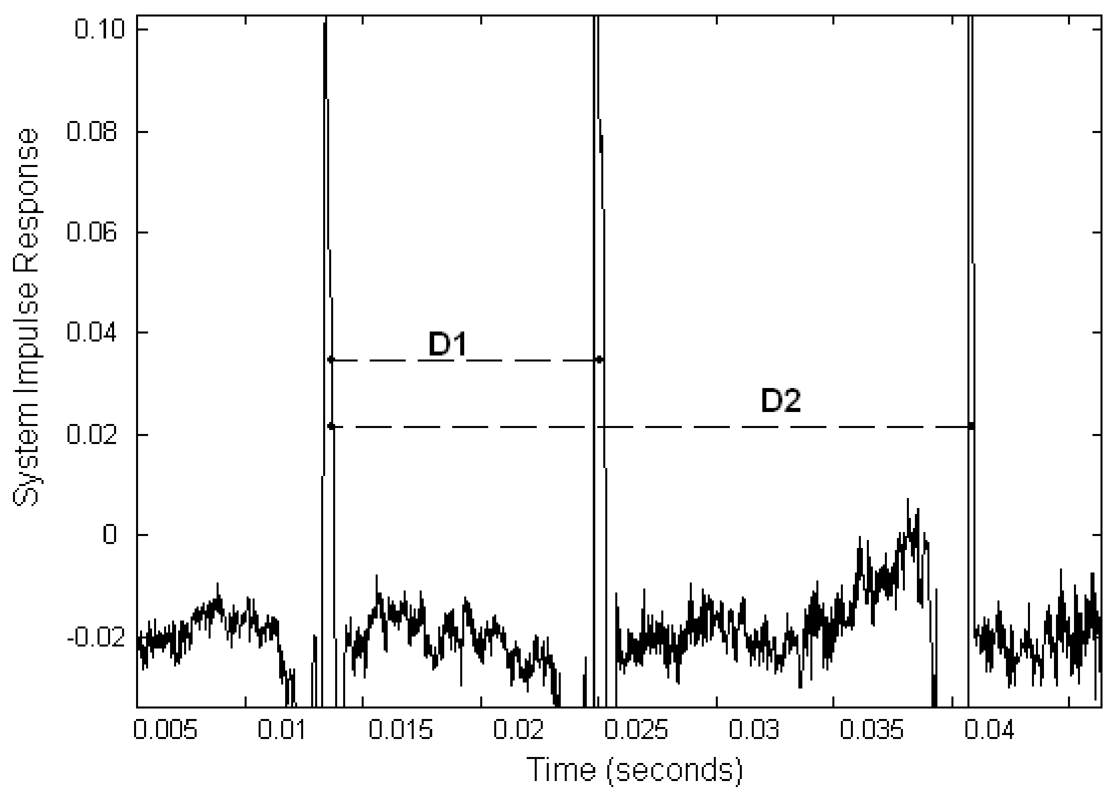


Figure 4.4(b): Acoustic measurement of a closed source tube using a MLS m = 18

The time interval D1 is the time taken for the pulse to pass the microphone, reflect off the cap and return to the microphone. The distance travelled is equal to $2l_2$ in a time of 0.012 seconds, therefore l_2 is measured as 2.056m with a reading error ±0.05m. The time interval D2 is the time interval D1 plus the time taken for the pulse to travel back down the source tube, reflect off the loudspeaker and return to the microphone. The distance travelled is twice the length of the source tube (i.e. $2(l_1 + l_2)$) in a time of 0.027 seconds. Therefore the length of the source tube is measured as 4.63m where l_1 = 2.575m. Both measurements are within 3% of the manually measured lengths. This provides sufficient accuracy to window the end section of the source tube and determine the sample point from which to begin the reconstruction algorithm.

4.2.3 Source Tube Cap

A cap was fabricated from black nylon, shown in Figure 4.5, in order to block the open end of the source tube during the calibration phase to deconvolve the source tube measurement from the object reflection measurement. The cap was designed to provide a tight fit to the source tube to prevent acoustic leakage. From previous work by Marshall and Sharp et al. [7, 13], we can conclude that the black nylon cap provides a good approximation to a perfect reflecting surface for measuring the reflection of the input pulse shape.

Figure 4.5: Source tube cap

4.2.4 Mouthpiece

A mouthpiece to connect the source tube to the test object was fabricated from black nylon and is shown in Figure 4.6. The mouthpiece was designed to be an extension of the source tube (matching cross sectional area) and to allow a comfortable coupling between human subjects and source tube. In the interests of hygiene, the outer diameter of the mouthpiece was matched with commercially available sterile sleeves (normally used in pulmonary function tests). Each participant used a new sleeve when performing measurements which meant only a quick wipe of the mouthpiece with disposable sterile wipes was required between each sitting. This proved to be a clean and efficient method of readying the reflectometer for multiple test subjects.

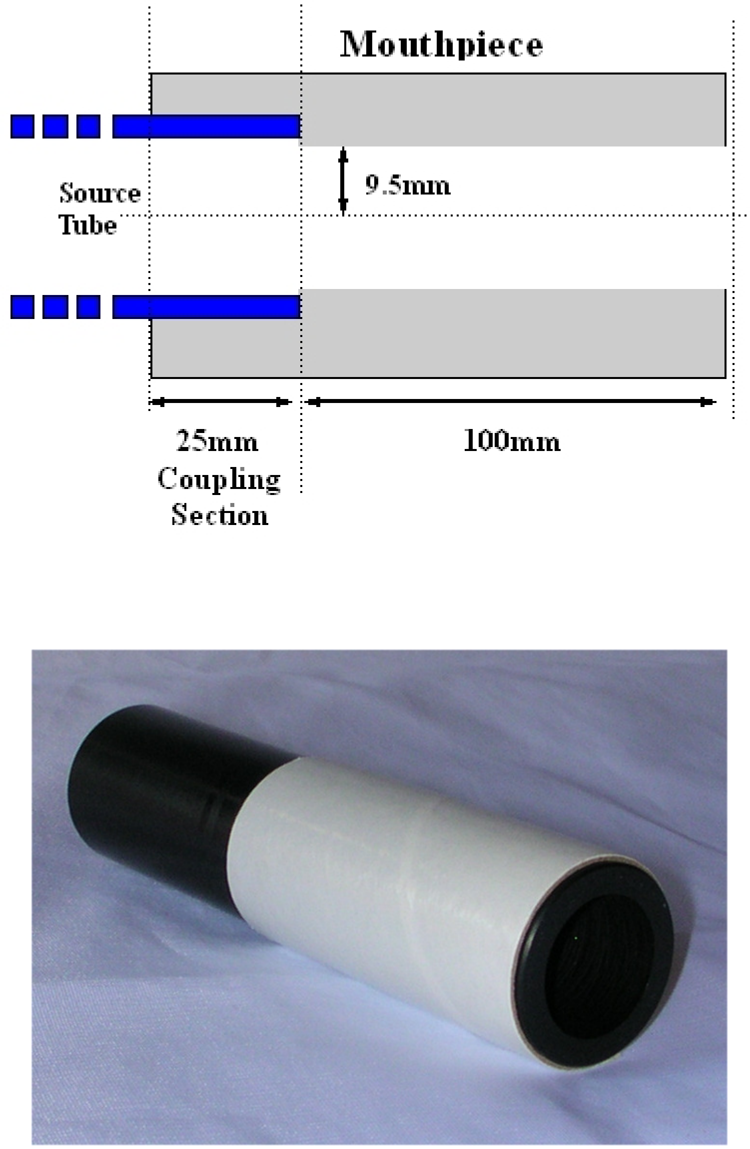


Figure 4.6: Schematic diagram and photo of mouthpiece for coupling with APR source tube

4.2.5 Nose Clips

During the initial stages of human trials with the acoustic pulse reflectometer for volumetric measurement of the vocal tract, subjects were asked to wear nose clips (shown in Figure 4.7) when a measurement phase was triggered.

Figure 4.7: Nose clips

Even though the vowels being investigated in this study are non-nasalised, the idea to use nose clips stemmed from an initial observation that a raised or lowered velum (a muscular flap that closes off the nasopharynx during swallowing or speaking) would be sampled as a branching pipe into the nasal tract and therefore be shown as acoustic leakage in the cross sectional area reconstruction of the vocal tract.

By placing nose clips on the nostrils it was thought that the subject would not open the velum to facilitate nasal breathing whilst their mouth was attached to the mouthpiece. This was to encourage the participant to hold their breath for the duration of the measurement (a vital requirement for the experiment). In effect, the use of nose clips proved to be a counter productive addition. By not only asking the participants to hold their breath, this move also blocked both orifices for breathing. The result was that some participants were more inclined to create a small positive pressure against the mouthpiece and hence the source tube, which was picked up by the pressure sensor (microphone) embedded in the wall of the pipe. This resistance against the mouthpiece and source tube seemed to be an instinctive reaction to blow away the pipe to enable breathing (rather than a negative pressure which would have suctioned onto the participants lips). Similar results were found in [18].

From these observations, it was decided to forgo the use of nose clips in future experiments.

4.3 Sample Rate & Axial Resolution

The method employs the use of a PC which digitally samples the analogue acoustic signal. Information is received on the reflections from the object once every sample (i.e. once every $T = 1/F_s$ seconds where F_s is the sample rate). The result is that information is obtained on the change in internal profile at discrete points along the object's bore. The reconstructed bore will therefore be approximated by a series of cylinders whose length is such that the primary reflections from successive cylinders occur at the sample rate [31]. The length of each cylindrical segment is $l = cT/2$ (the speed of sound in air is taken to be $c = 331.6\sqrt{1+\tau/273}\,\text{m/s}$, where τ is the air temperature in ºC). All measurements were made at τ = 20ºC with a sampling rate F_s = 44.1 kHz, hence, the length of the cylindrical segments making up the reconstruction is:

$$l = \frac{cT}{2} = \frac{\left(331.6\sqrt{1+\frac{20}{273}}\right)\times\left(\frac{1}{44100}\right)}{2} = 0.0039\text{m} = 3.9\text{mm} \tag{4.1}$$

4.4 Virtual DC Tube

The bore reconstruction is very sensitive to the DC offset which is present in the calibration pulse and object reflections, presented as a small constant positive value. The DC offset is therefore also present after deconvolution in the input impulse response. During the bore reconstruction, this is interpreted as a small reflection coefficient that causes the reconstruction to expand or contract too rapidly. For accurate reconstruction, the offset must be separated from the signal before applying the reconstruction algorithm. A solution proposed by Sharp et al. [29] placed a 40.3cm long cylindrical connector of the same radius as the reflectometer source tube between the source tube and the object. The connector, referred to as the DC tube, meant that object reflections would arrive at the microphone 2ms later because of the additional time for the sound wave to travel in the tube. There would also be no reflected signal from the DC tube as there were no expansions or contractions along its length, therefore the average value of the input impulse response within the first 2ms could be subtracted from the whole input impulse response, removing the DC offset. In practice, the join between the source tube and DC tube was not perfectly smooth, giving rise to a small reflection at the start of the input impulse response. The solution was to replace the response between 0ms and 1ms with zeros and calculate the average value of the input impulse response between 1ms and 2ms and subtract this from the whole input impulse response.

In a development to this method, Kemp et al. [31, 65] recognised that is was inconvenient to have to physically couple the source tube and object under test and noted that only half the available 2ms in the input impulse response was used in the DC offset calculation.

The new method proposed the application of a ‘virtual DC tube’ which simulates the effect of the DC tube without the need for an intermediate physical coupling. This is accomplished by initiating recording of reflections from an object 2ms earlier and using a digital filter to add the losses that would have occurred if the sound had travelled across a DC tube. By removing the physical join between DC tube and source tube, the virtual DC tube can be perfectly joined thus eliminating the small reflection and the need for a method with which to correct it. It is the virtual DC tube method that has been adopted throughout this study.

4.5 Operation of the Reflectometer

In section 3.4.2, the input impulse response (*iir*) was defined as the sequence of reflections returning from an object under test when an ideal delta function (δ) is fed into the input. In theory, the mathematical construction representing a perfect impulse has infinite amplitude, an infinite extent in frequency, and zero width (infinitely short in time duration) [63]. In practice, the acoustic pulse that is produced is not an ideal delta function (impulse) because of its finite duration. However, using signal processing techniques, an approximation to the input impulse response can be obtained by deconvolving the reflections which return to the input plane of the object with the input pulse. Therefore, for the experimental measurement of the input impulse response which is required for the bore reconstruction algorithm, two separate measurements are made: i) a calibration pulse measurement and ii) a measurement of reflections from the object.

4.5.1 Calibration Pulse Measurement

The input pulse is found by terminating the source tube using a cap with a flat rigid surface (see 4.2.3) and recording the reflected pulse at the microphone [66]. This ensures that both the object reflections and the input pulse travel the same length l_2 of source tube and hence experience the same source tube losses. It also ensures that the deconvolution yields just the input impulse response of the object, without including the section l_2 of source tube.

Figure 4.8 shows a measurement of the calibration pulse resulting from the correlation between the microphone signal and a m = 18 MLS input signal, the procedure for which is discussed in section 3.6.3.

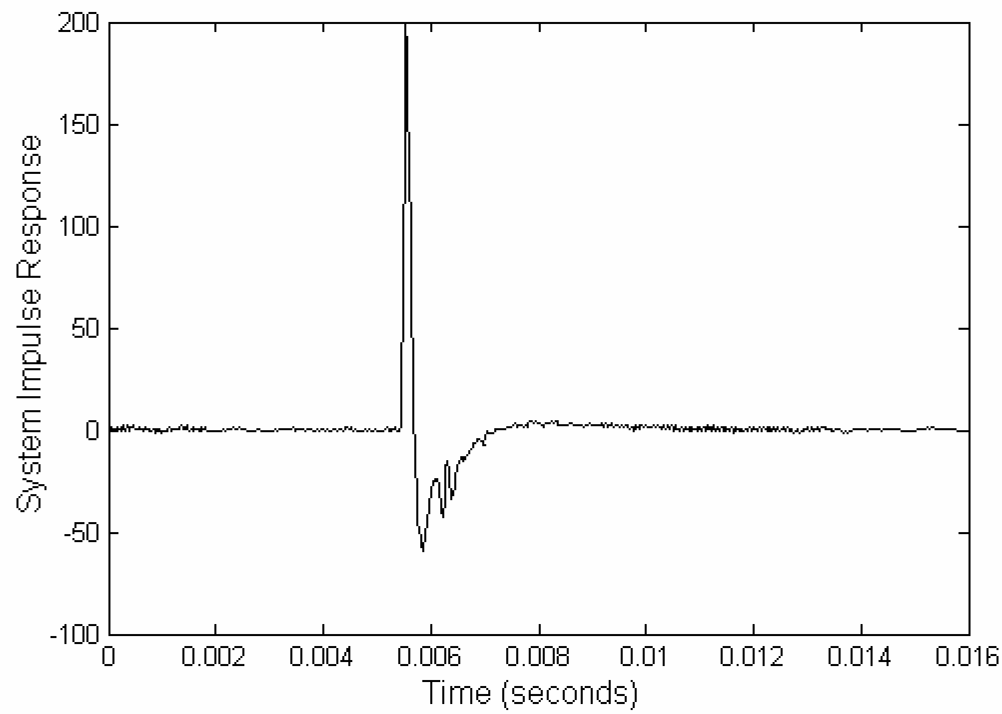


Figure 4.8: Calibration Pulse

4.5.2 Object Reflections Measurement

Figure 4.9 shows the reflections from correlating the MLS input signal and signal measured at the microphone from a test object consisting of stepped cylinders (fabricated from black nylon). A schematic diagram of the stepped tube attached to the end of the source tube is shown in Figure 4.10. Both Figures are labelled and correlate each reflection to its associated change in cross-sectional area.

The first section of the stepped tube acts as a coupler to the source tube and contributes a 50mm long section with a radius of 9.5mm (an approximate match to the radius of the source tube) to the bore of the object. The second section of the stepped tube object is 40mm long with an internal radius of 8mm; the third section is 60mm long with an internal radius of 3mm; the fourth section is 50mm long with an internal radius of 8mm; the end section is 30mm long with an internal radius of 5.5mm and is open, allowing sound to radiate. The stepped cylinders were measured with callipers with a reading error of ±0.5mm.

The labels referred to in Figures 4.9 and 4.10 are explained as follows:

A. Since the recording is triggered to start immediately before the reflections from the test object arrive at the microphone, the first few milliseconds consist only of background noise.

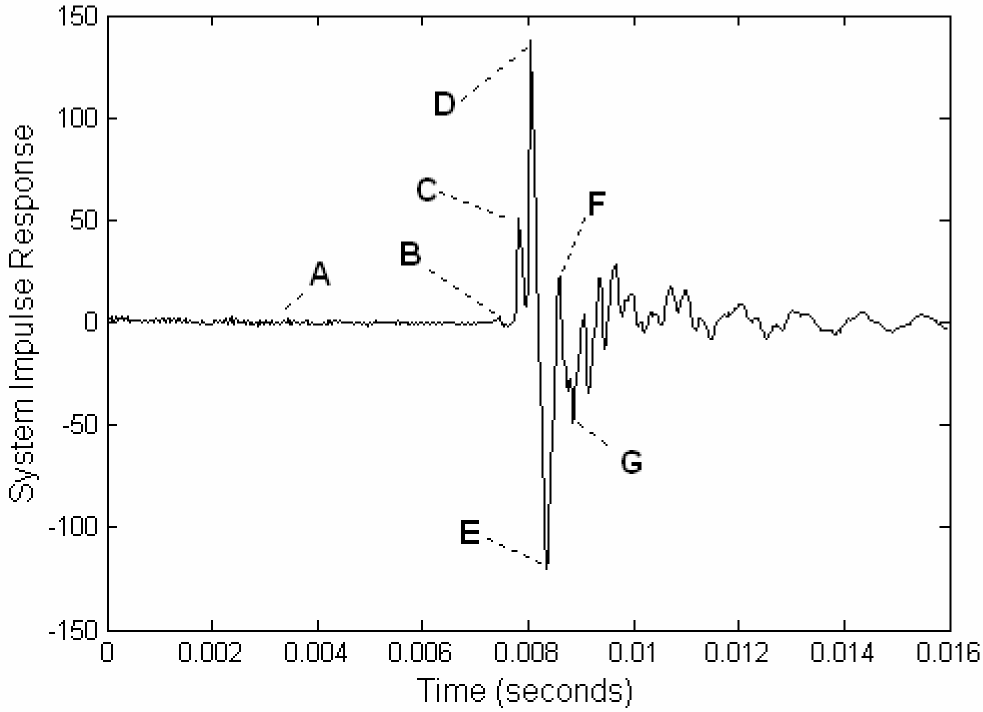


Figure 4.9: Reflections measured from a stepped tube

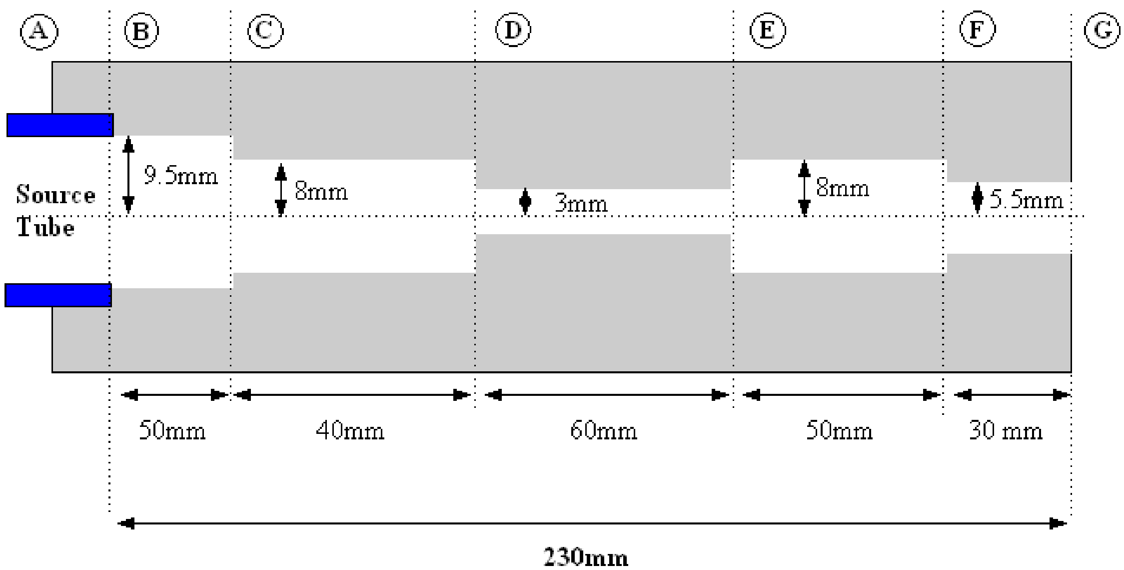


Figure 4.10: Schematic diagram of a stepped tube connected to source tube

B. As there is a slight difference (~0.5mm) between the radii of the cylindrical source tube and the first ‘coupler’ section of the stepped tube test object, there is a small reflection of sound at this join.

C. There is then a positive spike corresponding to the reflection of the input pulse from the contraction at the first step.

D. The second positive spike corresponds to sound transmitted into the second section of the test object, where there is a further contraction of the test tube radius.

E. The first negative spike is when sound is being transmitted into the third section of the test object. The sign change is due to the reflection from the expansion because transmission is always positive.

F. The third positive spike corresponds to sound transmitted into the fourth section of the test object, where there is a contraction of the test tube radius.

G. The second negative spike is when the component of the input pulse is transmitted to the open end of the stepped tube test object where it receives a negative reflection and is transmitted back to the microphone.

These are the primary reflections, where the input pulse experiences only one cycle of reflections before reaching the microphone. The remaining parts of the recorded pressure signal are multiple reflections where the pulse experiences at least one reflection back down towards the open end of the stepped tube test object before returning to the microphone.

4.5.3 Deconvolution

The deconvolution is carried out by performing a FFT on both the object reflections and the input pulse (calibration measurement). A complex division is then carried out in the frequency domain:

$$IIR(\omega) = \frac{R(\omega)}{I(\omega)} \tag{4.2}$$

where ω is the discretised angle frequency, $IIR(\omega)$ is the Fourier Transform of the input impulse response (which is then inverse FFT'ed to give the input impulse response in the time domain, $iir[nT]$), $R(\omega)$ is the Fourier Transform of the object reflections, and $I(\omega)$ is the Fourier Transform of the input pulse given by the calibration measurement. In practice however, a constrained deconvolution is used

$$IIR(\omega) = \frac{R(\omega)I^*(\omega)}{I(\omega)I^*(\omega)+q} \tag{4.3}$$

where $I^*(\omega)$ denotes the complex conjugate of $I(\omega)$. A 'constraining factor', q, is introduced to the denominator to prevent division by zero which would otherwise occur since the calibration pulse measurement consists only of background noise at high frequencies. In practice, it low-pass filters the input impulse response, removing high frequency noise [7]. It is set by trial and error to be small, relative to the in-bandwidth spectral components, but large relative to the noise floor, since then $IIR(\omega)$ tends to zero [57].

4.6 Choice of MLS Signal

Maximum Length Sequences (MLS) were introduced in Chapter 3. Here we discuss the application of MLS excitation in acoustic pulse reflectometry.

Previous methods used for measuring the impulse response of a system involved injecting a pulse into the source tube and measuring the reflections. By using a signal that continues over a longer time interval (such as MLS) we may put more energy into the system, improving the signal to noise ratio and removing the need to average over hundreds of measurements [31]. One of the challenges of using MLS excitation is choosing an appropriate value of signal (‘*m*’ value) and therefore time length. Too small a value can lead to deterioration in the signal to noise ratio and too high a value would increase measurement time and put a strain on available computer processing resources.

A limiting factor on this choice as far as measurements involving the vocal tract are concerned is the length of time that a human subject is able to hold a phonatory position (vocal tract and articulators) for imitating a vowel sound whilst also refraining from breathing. The time averaged MLS technique is normally applied to linear time invariant (LTI) systems, but in the case of the vocal tract, relatively small shifts in vocal articulators can sometimes produce major differences in resonation characteristics between vowels which will cause errors in the measurement.

Before proceeding with any measurements, a group of volunteers were asked to hold a phonatory position for different vowels and refrain from breathing. Using a video camera, each participant was recorded using a framed shot of his or her opened mouth and neck region. By analysing the recordings for visible motion of the vocal tract and articulators, a better understanding of the time duration an individual is able to remain stationary was gained. The results showed an average maximum duration of less than 15 seconds for a human to remain comfortably immobile under the conditions of the experiment. As described in Chapter 3, the MLS signal must be played end to end twice, ignoring the response during the first run through in order to make sure that the response being measured conforms to the periodicity condition. This translates into an MLS signal no greater than m = 18 (where $(2^m - 1)/F_s \approx 5.94$ seconds for one cycle of the signal).

Experiments were performed using MLS signals in the range $14 \le m \le 18$ to ascertain the best values to be used on test objects and human subjects. Two experiments were devised to test the effectiveness of the MLS signals; the measurement and reconstruction of an open and closed source tube. Figure 4.11 shows the response of the system (closed source tube 'calibration measurement' – see section 4.5.1) to a range of MLS signals $14 \le m \le 18$. From this we can see that higher order MLS signals result in a greater amount of energy being input into the system as they continue over a longer time period.

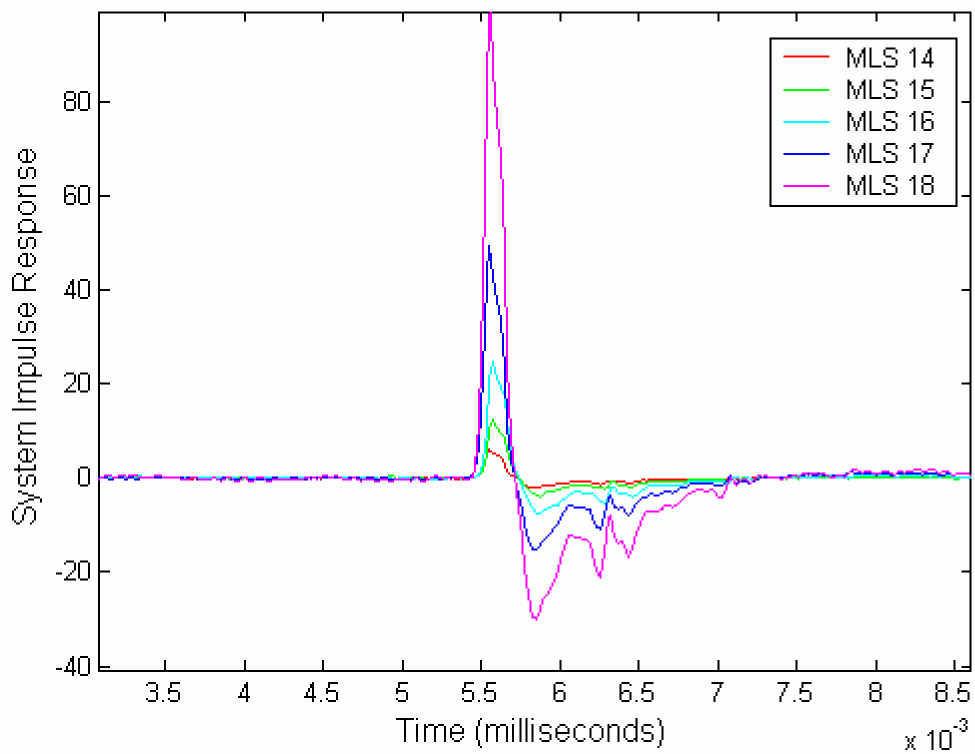


Figure 4.11: Response of system to a range of MLS signals

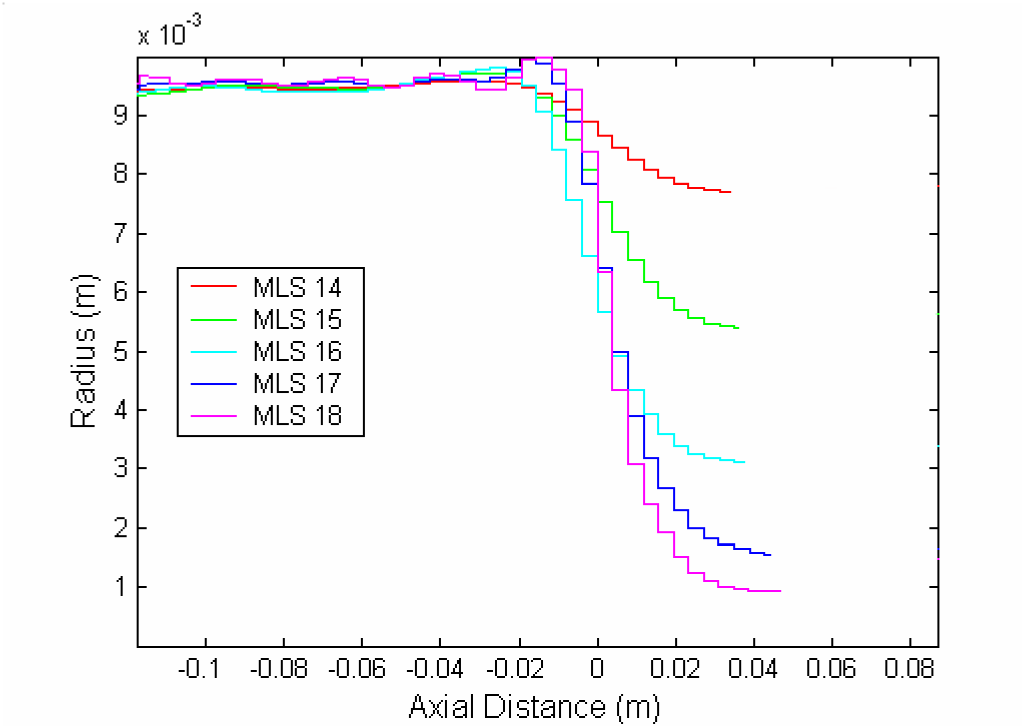


Figure 4.12: Reconstruction of a blocked source tube using a range of MLS signals

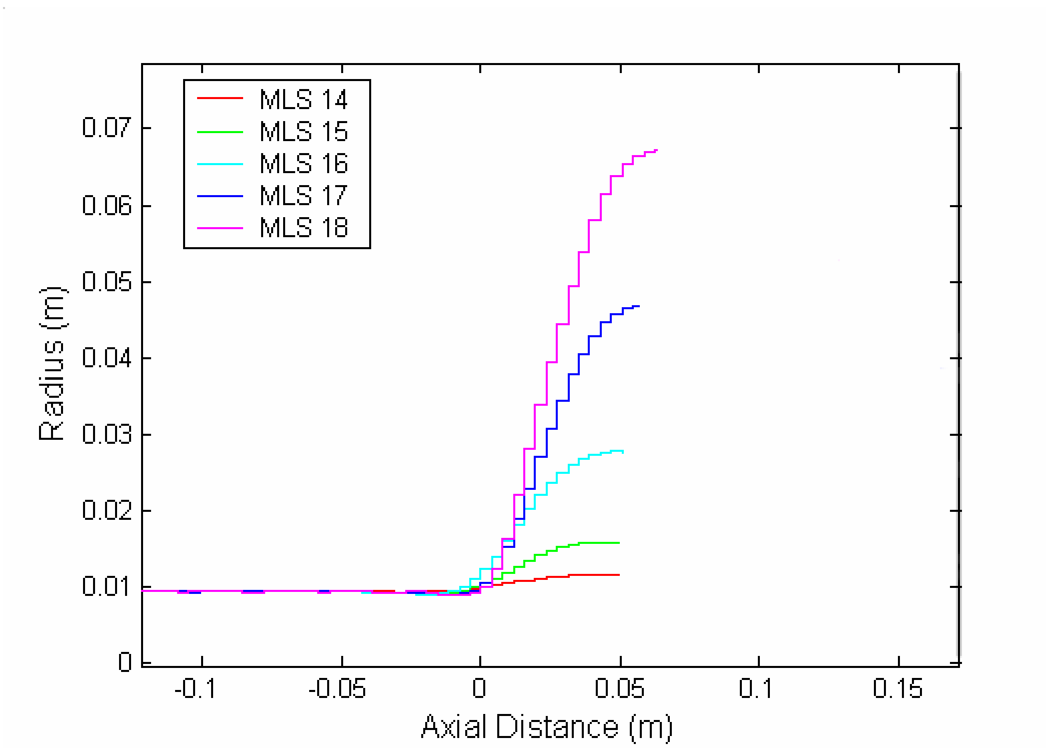


Figure 4.13: Reconstruction of an open source tube using a range of MLS signals

Examination of Figures 4.12 and 4.13 shows that bore reconstructions using MLS signals $m<16$ perform poorly at extreme changes in cross sectional area (open and closed source tube end). A closed pipe (Figure 4.12) would in theory be an abrupt change in cross sectional area down to zero radius. An open pipe (Figure 4.13) can no longer be treated as a tubular structure and would in theory, show a bore reconstruction increasing to infinite cross sectional area. The reflectometer is designed to measure cross sectional areas along the axial plane that are not wildly different to the source tube. Therefore, both of these cases are outwith the realms of the practical reflectometer. However, the experiments show that the lower order MLS signals may not input enough energy into the system for accurate reconstruction of the varying cross sectional profile of the vocal tract.

Due to the important biomechanical functions used in breathing, swallowing and speaking, it is important that experiments on the vocal tract are non-invasive and can be carried out speedily. Whilst it is clear that lower order MLS signals do provide shorter measurement times, the results show this advantage to be lost in lower signal to noise ratios, leading to poor bore reconstructions.

From these results, suitable values for MLS signals throughout this work will have a lower bound of $m=16$.

4.7 Measurement of Test Objects

To validate the performance of the reflectometer, measurements were made of two test object using an MLS signal of $16 \le m \le 18$. The test objects were constructed from black nylon and both have a cylindrical symmetry but differ in cross sectional area:

- Mouthpiece model - a smooth bore (no internal area changes) matching the internal diameter of the source tube, (shown in Figure 4.6)
- Stepped Tube model - a tube with discontinuous area changes (shown in Figure 4.10)

4.7.1 Mouthpiece Reconstruction

The mouthpiece was first introduced in section 4.2.4 and is used to couple the source tube with a human subject. It also acts as a useful calibration tool for measuring objects of unknown bore. With its known dimensions, the mouthpiece acts as an intermediate stage for the reconstruction (which begins at the end of the source tube after the virtual DC tube) when measuring objects of unknown bore. As such, it enables any anomalies in the object coupling to be highlighted. For example, when first fabricated the mouthpiece was believed to be a tight fit with the source tube. Unfortunately an error occurred during machining where the overhang of the mouthpiece coupling section had a slight flare, preventing a flush coupling against the end of the source tube. This was not immediately obvious as the internal coupling section was not visible. During test measurements, however, instead of matching the internal radius of the source tube with a flush join, there was a small expansion (step) at the beginning of the mouthpiece reconstruction. Figure 4.14 shows a schematic diagram and a 3D area equivalent (cylindrical tube) model of the discontinuous coupling between the source tube and mouthpiece.

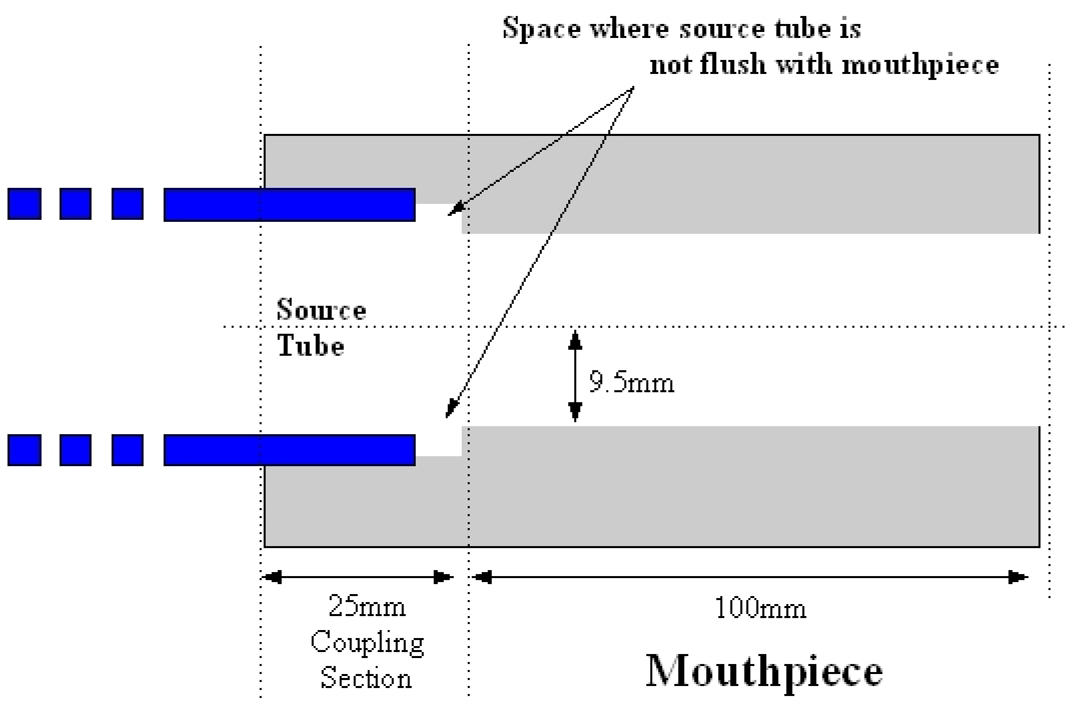


Figure 4.14: i) Schematic diagram of the discontinuous coupling between source tube and mouthpiece.

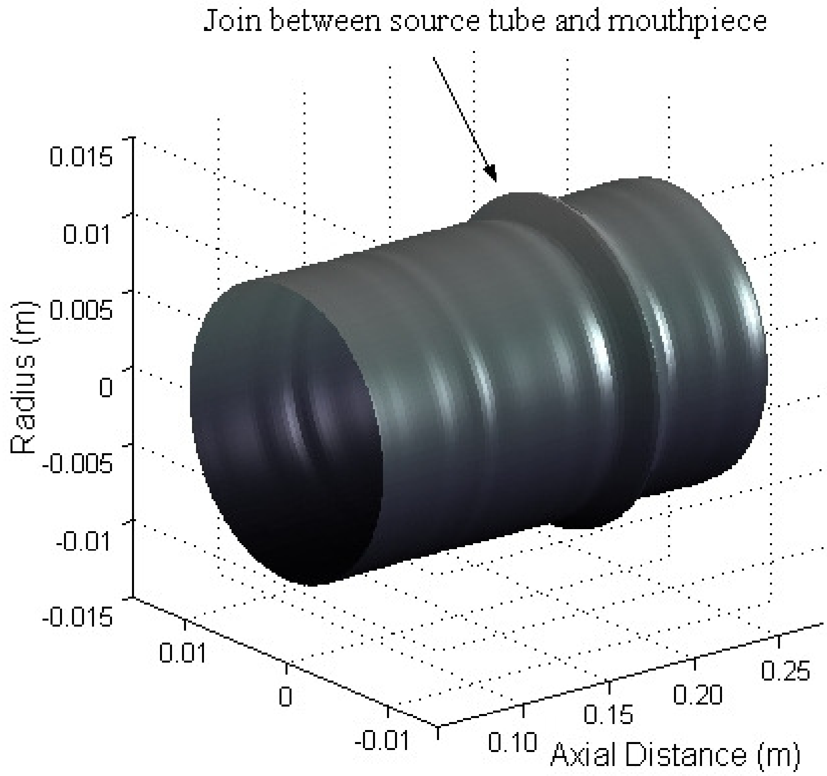


Figure 4.14: ii) 3D area equivalent (cylindrical tube) model of the discontinuous coupling between the source tube and mouthpiece

From these results, the mouthpiece was re-machined to enable a flush coupling with the source tube. Figure 4.15 shows the reconstructions of the mouthpiece using MLS excitation with a range $16 \leq m \leq 18$.

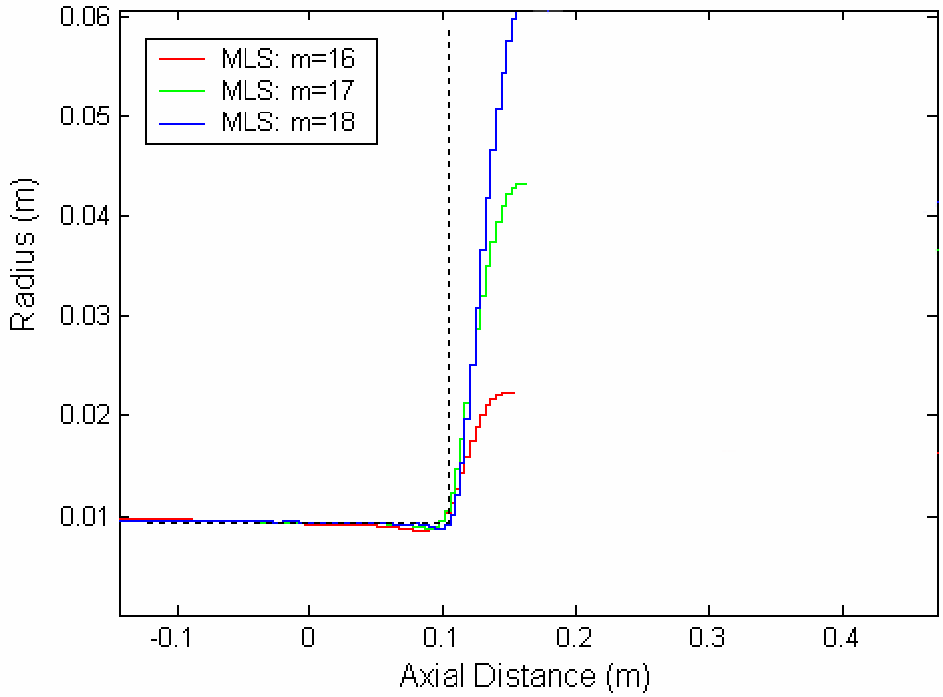


Figure 4.15: Reconstructions of the mouthpiece using different orders of MLS excitation

The reconstructed profiles for each of the MLS signals show good agreement with the measured mouthpiece. At 70mm along the 100mm mouthpiece the m = 16 signal deviates from the measured value by an average -0.3mm, but within an error bound of ±0.5mm. The reconstructions continue past the open end of the mouthpiece (infinite expansion) for a distance of up to 50mm. The reconstructions using the MLS signals m = 16, 17, 18 were able to measure an expansion of 220%, 450% and 600% respectively before stopping.

4.7.2 Stepped Tube Reconstruction

As described in Chapter 2, reasonably detailed models of articulation can be developed by using a small number of tubes (stepped cylinders) to model the vocal tract. The stepped tube object (Figure 4.10) was modelled on a 4-tube model of the vocal tract described by Fant [33], with the addition of a coupling section to attach the object to the source tube. This object consists of 4 parts representing a simplified model of the vocal tract:

- one tube for the area behind the tongue
- one tube for the constriction created by the tongue during articulation
- one tube for the area in front of the tongue
- one tube for the lip aperture

The stepped tube object is an acoustic model of the vocal tract for the placement of formants F_1 and F_2 in vowel space and is not representative of the actual profile of the vocal tract which has a bend at the back of the mouth (the oropharyngeal junction) and does not have cylindrical symmetry. It is nevertheless useful in investigating the performance of the reflectometer to discontinuous area changes. Figure 4.16 shows the reconstructions of the stepped tube using MLS excitation with a range $16 \leq m \leq 18$.

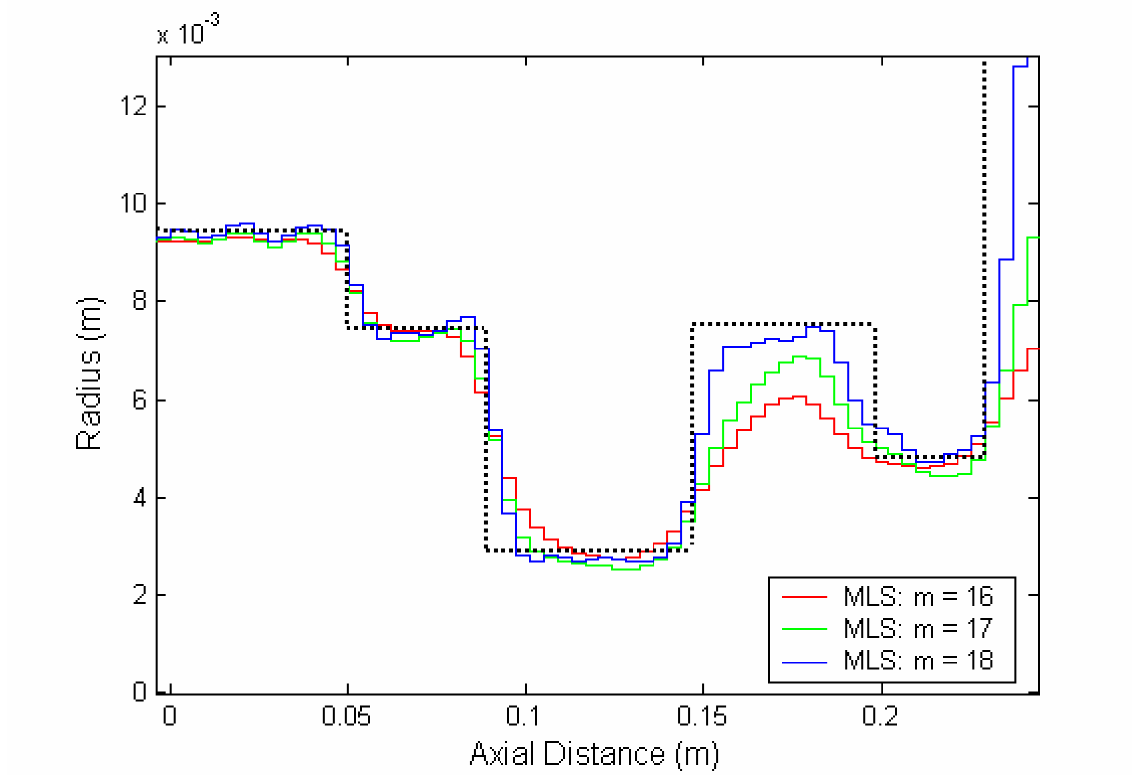


Figure 4.16: Reconstructions of a stepped tube using different orders of MLS excitation

Examination of Figure 4.16 reveals that the reconstructed profiles all agree with the measured radii to within a reading error of ± 0.5mm for the first two steps of the stepped tube object (shown as a dotted line).

At the third cylindrical section (changing from 8mm ± 0.5mm to 3mm ± 0.5mm), the reconstructed profiles take a longer axial distance before reaching values within the error bounds of the measured radius; when m = 17 and18, this axial distance is approximately 10mm whereas when m = 16, the distance is approximately 20mm.

At the fourth cylindrical section (changing from 3mm ± 0.5mm to 8mm ± 0.5mm), there is a marked difference between each of the reconstructed profiles. Only when

m = 18 does the reconstruction reach a value within the error bounds of the measured radius (this occurs at a distance of 20mm from the junction). When m = 16 and m = 17, the reconstructions reach maximum radii of 6mm and 7mm (outwith the error bounds of the measured radius) and these maxima occur 35mm from the junctions.

At the final cylindrical section (changing from 8mm ± 0.5mm to 5.5mm ± 0.5mm), each of the reconstructed profiles agree with the measured radius to within the error bounds within 5mm of the junction.

Finally, to the far right of the graph, the reconstructed profiles of the open end of the stepped tube object can be seen, with the higher order MLS profile more accurately reflecting the cross sectional radius change within a shorter axial distance of the junction.

From these results, the following observations can be made:

- From the large steps in radius exhibited by the reconstructed profile with m = 18, it is evident that an increased axial resolution would provide a more accurate transition between discontinuous area changes in short pipes.
- The levels of oscillations around the measured radial value are slightly reduced by the increase in order (particularly evident in the third cylindrical section – the main constriction in the stepped tube). This implies that the noise level at high frequency is reduced as a result of the fact that we have added more energy to the system.

- The fact that the m = 18 reconstruction of the second last cylinder has an average value closer to the correct radius is evidence of greater accuracy at low frequencies, as was shown by Kemp [31].

4.8 Conclusions

As expected, the most accurate reconstructions of both the stepped tube and mouthpiece objects used the highest order MLS excitation signal (m = 18). However, several important differences remain between the measurement of test objects and the human vocal tract. For example:

i. The test objects are static and consist of the same material throughout their structure. As described in Chapter 2, the human vocal tract is malleable with various fleshy tissues, muscles and teeth within a varying bone structure.

ii. The human vocal tract exhibits greater changes in cross sectional area than the mouthpiece test object, but is more smoothly varying than the discontinuous area changes of the stepped tube object.

iii. The important biomechanical functions of the vocal tract (breathing, speaking and swallowing) limit the time a human subject can comfortably remain immobile under the conditions of the experiment. The highest order MLS excitation signal used in these experiments (m = 18) is at the upper end of this time limit (see section 4.6). Therefore, although the higher order MLS signals have been shown to provide greater accuracy in the reconstruction of bore profile, this advantage may be lost to errors introduced by a non-stationary vocal tract.

In the following chapter we will apply acoustic pulse reflectometry to vocal tract measurement in human subjects using MLS excitation signals in the range $16 \leq m \leq 18$.

Chapter 5

APR for vocal tract measurement with a human subject

5.1 Introduction

This chapter describes the application of acoustic pulse reflectometry to vocal tract measurement with a human subject using MLS excitation signals. The results from this chapter formed the basis of a paper that was presented at the Stockholm Musical Acoustics Conference 2003 [67].

Chapter 2 introduced the voice and the differences between speech and singing. Here we focus on vowels as static targets of articulation, in which the cross sectional profile of the vocal tract is measured using APR.

5.2 Vowels

The word vowel comes from the Latin word vocalis, meaning “uttering voice” or “speaking”. A vowel is a sound produced by an oral configuration during phonation without occluding, diverting, or obstructing the flow of air from the lungs (sometimes referred to as an open configuration of the vocal tract). This is in contrast to consonants, which are characterised by a constriction or closure at one or more points along the vocal tract.

The purpose of using vowels within this study is that they are static targets of articulation in speech (i.e. once the vocal tract has assumed a set shape through the use of the articulators, the resonating air column is fixed to filter the complex glottal waveform into what we perceive as vowel sounds). As discussed in Chapter 2, singing places greater emphasis on the use of vowels than speech as the bias is towards the phonation of sustained pitched sounds (both as sung text and also vocal texture). The exaggeration of the mouth in singing to satisfy aesthetic and acoustic requirements such as voice projection (which Titze describes as the *megaphone effect* [68]) is particularly evident in the phonation of vowel sounds. The jaw is lowered to open the mouth as wide as possible in an attempt to match the acoustic impedance of free space. The lips are projected forward to prevent the spread of this gesture, which would otherwise shorten the effective length of the vocal tract. The forward movement of the lips and inward movement at the corners of the lips produce an exaggerated mouth shape, widening and lengthening the vocal tract to create the megaphone effect.

The prevalent use of vowels in singing, combined with the static nature of the vocal tract and open mouth during vowel production, lends itself to be a more readily accessible quantity of the voice to be measured using APR.

5.3 Acoustic and articulatory descriptions of vowels

Here we will discuss the complex interrelationship of the vocal articulators and the causal effect on the perceived acoustical signal.

> *"Each vowel sound is associated with a specific articulatory profile, producing a specific area function that in turn gives a specific combination of formant frequencies"*
>
> -- Johan Sundberg ([69] p.23)

The glottal waveform generated by the vibrating vocal folds propagates through the various differently sized areas in the vocal tract. Some of the frequencies within the spectrum will resonate more than others, depending on the size of the resonant cavity within the tract. Larger cavities will resonate at lower frequencies, while smaller cavities resonate at higher frequencies.

The two largest cavities within the vocal tract are the pharynx and the mouth, which therefore produce the two lowest resonant frequencies, or formants. These formants are designated as F_1 and F_2. The movement of the articulators for singing or speech shape the resonant cavities and control the formants.

5.3.1 The F_1- F_2 vowel chart

In 1952, Peterson and Barney [70] demonstrated that vowels are perceived on the basis of the two lowest formant frequencies of the vocal tract. They showed that the formant frequencies (F_1 and F_2) belonging to a vowel category can occupy an entire range of values if speaker differences, ages, and genders are included in a group of speakers. A convenient way of displaying this range is to transform data from a spectrum to an F_1- F_2 vowel chart (Figure 5.1).

While there are some differences in oral and pharyngeal resonance shape created by jaw opening and lip protrusion, the tongue is the prime mediator in the production of vowels. These phonemes are aurally recognisable because each vowel has its own distinctive F-pattern that is produced by a very precise configuration of the tongue. Most vowel charts (often referred to as a vowel triangle, or more precisely, a vowel quadrilateral [71]) only reveal the placement of formants within their set frequency region. This chart adds the general placement of the tongue in the mouth to assist with the understanding of the movement of the tongue in two dimensions (vertical and longitudinal) that create vowel sounds.

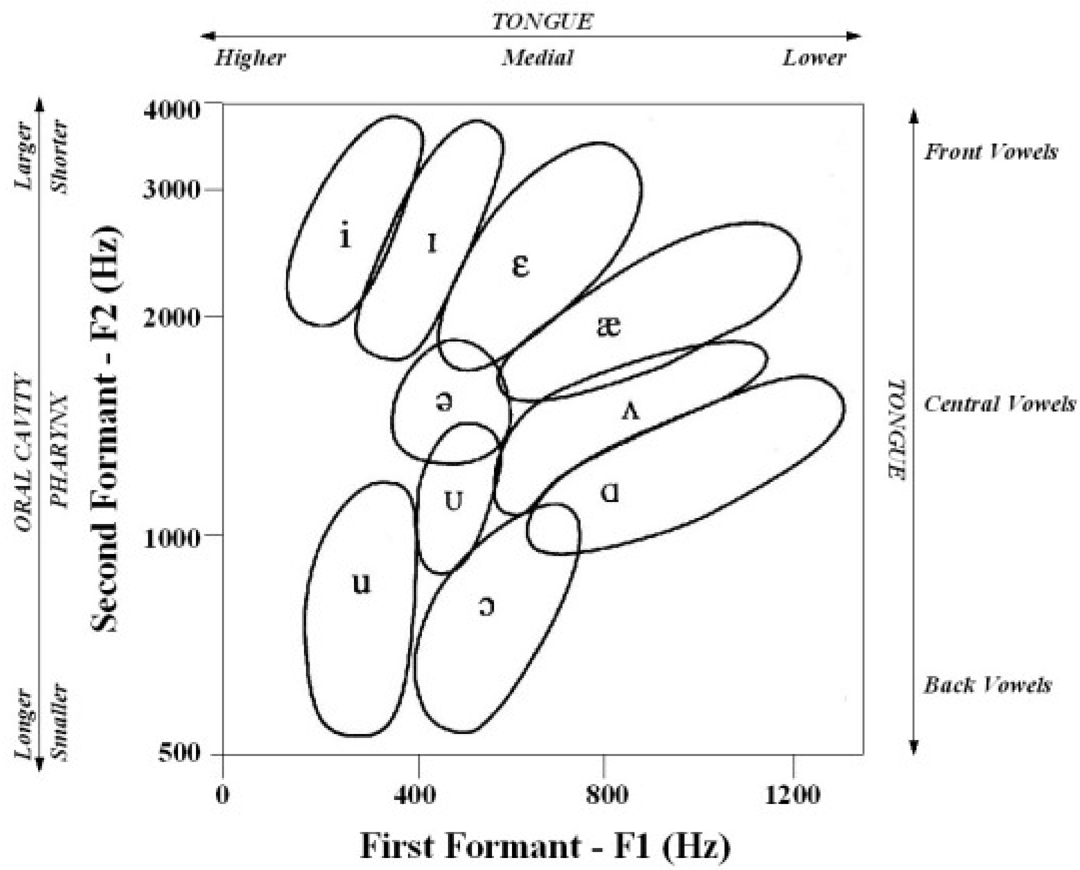


Figure 5.1: F_1/ F_2 vowel chart including tongue position

5.3.2 Tongue placement

In traditional articulatory phonetics, vowels are described on the basis of a high-low and a front-back placement of the tongue [39].

5.3.2.1 Vertical tongue movement in the oral cavity

As the tongue moves in the vertical plane, it has a combined effect on both the oral and pharyngeal cavities:

- Higher tongue placement creates a larger pharyngeal cavity as the tongue centres its mass whilst extending upwards
- Lower tongue placement causes a spreading of the tongue mass and increases the size of the oral cavity while the size of the pharyngeal cavity decreases.

The phonetic descriptions to indicate the three areas of production on the vertical plane (tongue height) are:

- Low
- Middle (often referred to as medial)
- High

5.3.2.2 Longitudinal tongue movement in the oral cavity

In speech science, the movement along the posterior/anterior axis is called tongue advancement. As in the vertical plane, tongue movement causes shifts in the ratio of oral/pharyngeal cavities:

- Frontal placement along the longitudinal axis increases pharyngeal space. As this happens, there is a corresponding decrease in oral space.
- As the tongue moves farther back, an increase in oral space is experienced with a concurrent reduction of pharyngeal space.

The phonetic descriptions to indicate the three areas of production on the longitudinal plane (tongue advancement) are:

- Front
- Central
- Back

5.3.3 Tongue shape

Any tongue configuration features muscular shaping on both planes simultaneously and will form a two-dimensional entity that offers a wide range of shapes and ratios of oro/pharyngeal resonance. Constant reshaping of the tongue occurs during voice production (in both speech or singing).

There is a third dimension in tongue movement. When viewed from the front, the tongue is also capable of executing a variety of compound lateral curved shapes. The outside edges of the tongue can be moved upward, while the area of the midline groove (running on the posterior/anterior axis, called the median lingual sulcus) remains depressed. As viewed in X-ray images in Appelman [72], all vowels employ this groove to a greater or lesser extent; the groove may extend for virtually the entire length of the tongue, from only the middle of the tongue to the back, or from only the middle of the tongue to the tip in the front. Many of the vowel configurations in speech require these outside raised edges of the tongue to make contact with the upper teeth, occluding an area varying anywhere from the third molar forward to the first premolar. Even if linguadental contact is not required, all vowels feature some degree of this compound curve. Figure 5.2 shows the vowel [i] both in sagittal and palatogram views. The grey area on the palatogram indicates tongue contact with the palate.

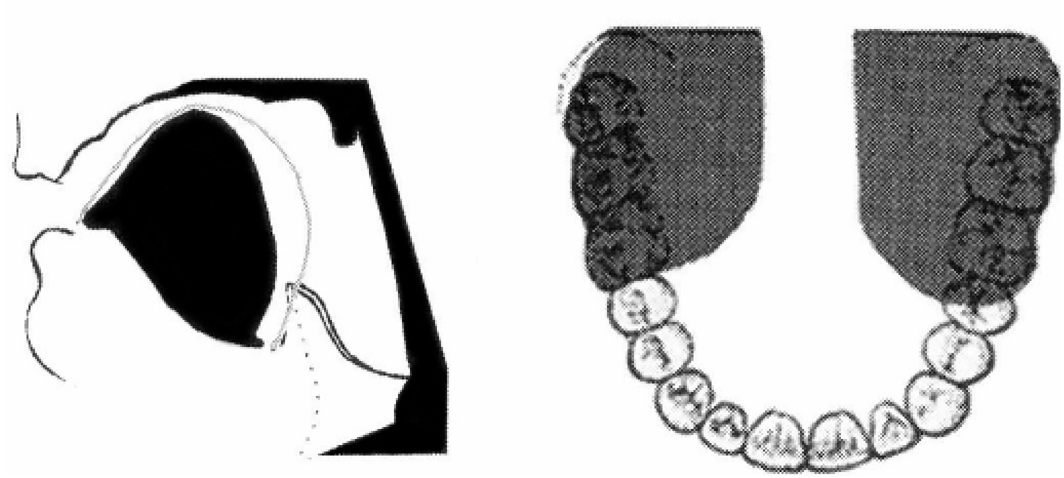


Figure 5.2: Sagittal view and palatogram of the vowel [i] showing linguadental occlusion (after Nair [71] p.97 and Appelman [72])

5.3.4 Lip shape

In combination with tongue position, vowels are also described in terms of the mouth aperture for which the controlling mechanisms are the lips and jaw height. The four main types of lip shape for the production of vowels are; spread, neutrally open, open rounding and close rounding. Examples[3] of these are shown in Figure 5.3.

[3] Images are reproduced with the kind permission of Gary Martin [73] Martin, G.C. *CG Imagery Development*. 1996 [cited 2005 January 2004]; Available from: http://www.garycmartin.com/.

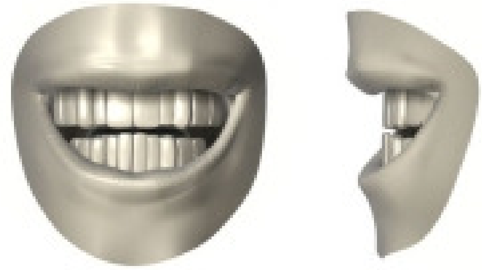


Figure 5.3(a): Lip spreading with raised jaw as in the vowel [i]

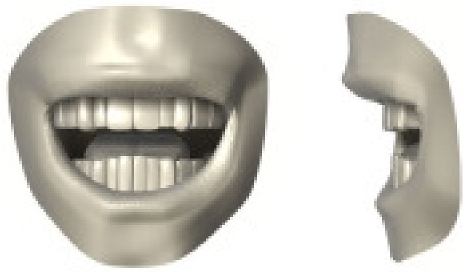


Figure 5.3(b): Lip neutrally open with lowered jaw as in the vowel [a]

Figure 5.3(c): Lip rounding with lowered jaw as in the vowel [o]

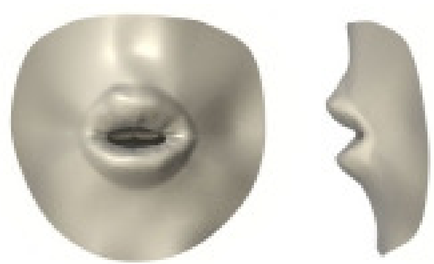


Figure 5.3(d): Lip close rounding with raised jaw as in the vowel [u]

5.4 The investigated vowels

Five non-nasalised American English vowels representing middle and corner placements within the F_1/ F_2 vowel chart were studied using APR (see Table 5.1).

Vowel	Example Words
[a]	*Ask, half, past*
[e]	*Ate, made, they*
[i]	*Each, free, keep*
[o]	*Obey, note, go*
[u]	*Ooze, too*

Table 5.1: Vowels with example words (southern British English pronunciation)

Continuing a method adopted in the literature and with reference to the discussion on resonances of the vocal tract in Chapter 2, vertical dashed lines are shown at the values 500Hz and 1,500Hz on the graphs of vowel frequency spectra in the following sections. These lines represent the formant frequencies of a tube open at one end (referred to as a closed tube – see Chapter 2, section 2.4), with a uniform cross section. The neutral schwa vowel [ə] matches closely this representation whereby the tongue is neither high nor low, nor is it placed to the front or the back, and in effect approximately makes the cross section of the vocal tract constant from glottis to lips. With reference to the formant frequencies of this neutral vowel, the following describes the vocal tract shape and corresponding spectra for each of the vowels studied. In addition, an explanation of the displacement of F_1 and F_2 from neutral for the corner vowels [a], [i] and [u] is presented.

5.4.1 Vowel [a]

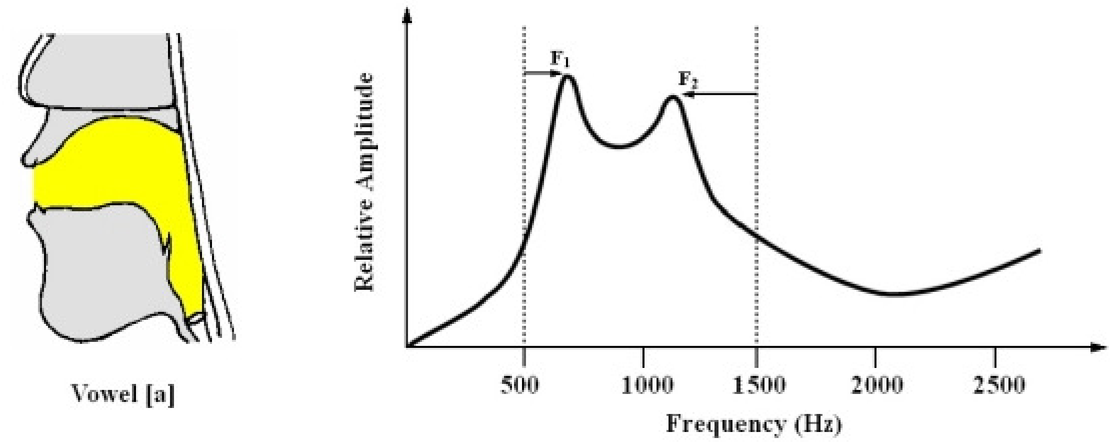


Figure 5.4: Midsagittal vocal tract profile and F_1/ F_2 frequency spectra for vowel [a]

As illustrated in Figure 5.4, the vowel [a] requires a comparatively open mouth cavity and a tongue that is low and pulled back into the pharynx. Fant [33] showed that this arrangement can be simplified by representing the vocal tract as two coupled tubes; a large tube representing the mouth and a narrower tube representing the pharynx (shown in Figure 5.5(a)). In Chapters 2 and 3, the transmission of waves at a junction between two discontinuous tubes was described. From this we saw that a wave experiences partial reflection and partial transmission at a change in impedance, occurring at a change in cross-sectional area. In this case, the area expansion (from pharynx tube to mouth tube) causes the transmitted acoustic pressure (p_t) to be less than the incident pressure (p_i) because the reflected pressure (p_r) has a negative polarity (rarefaction). It is this reduction in acoustic pressure that shifts the formant frequencies when compared with a uniform tube of equal length (see Figure 5.5(a)).

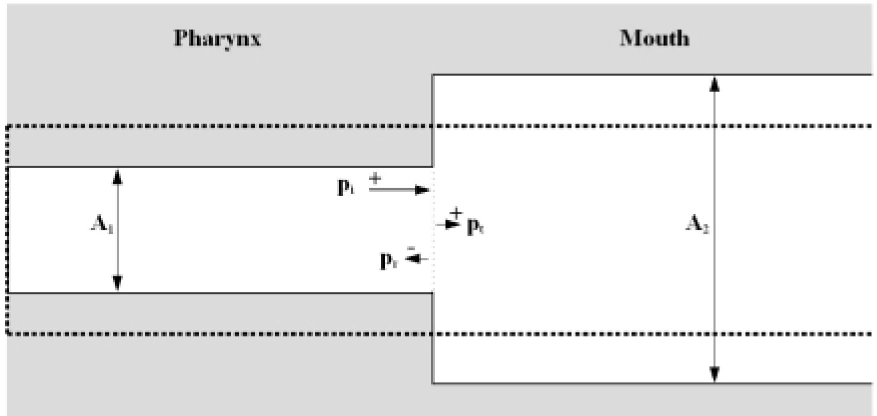


Figure 5.5(a): Two tube approximation for vowel [a]

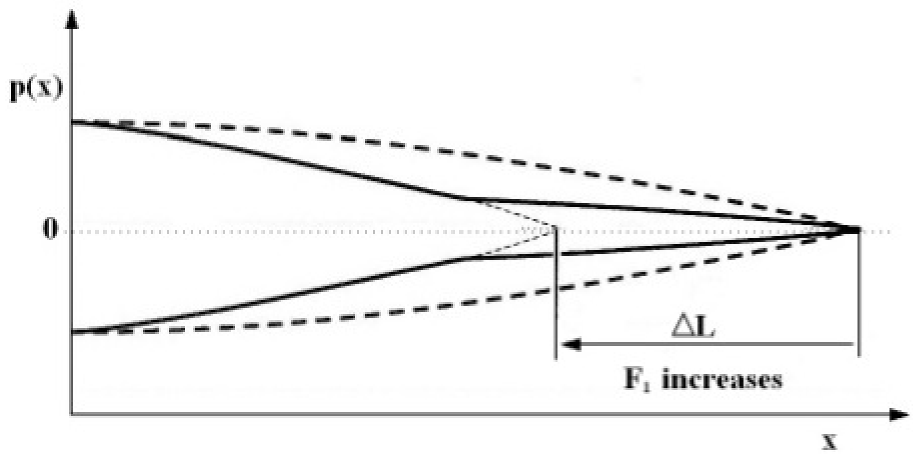


Figure 5.5(b): Standing wave pressure pattern for F_1

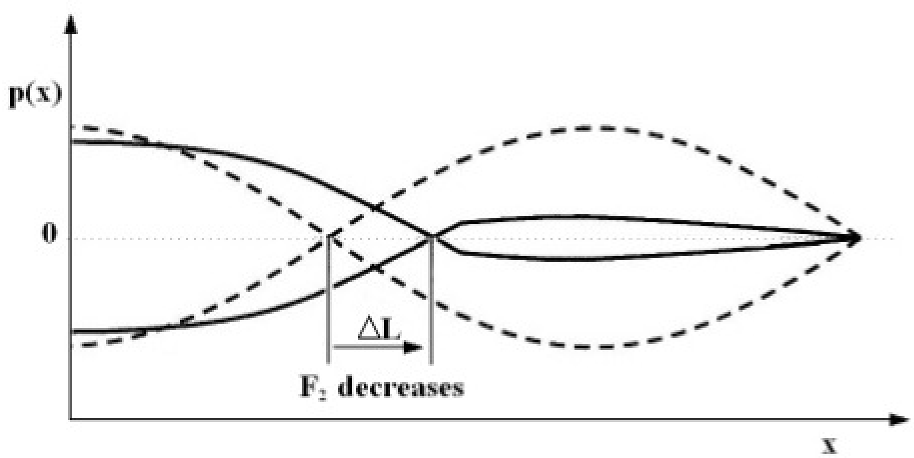


Figure 5.5(c): Standing wave pressure pattern for F_2

The standing wave pressure pattern for F_1 across the entire two-tube approximation of the vocal tract is shown in Figure 5.5(b) (solid line). The pressure wave pattern for a single uniform tube of equivalent length is shown as a dashed line (adapted from Titze [34] p.165).

From this graph we can see the system acts as a closed tube, with a pressure antinode (i.e. displacement node) at the glottis since no particle motions are possible at the closed boundary. The open end (at the lips) shows a pressure node in which air particles are free to move in and out of the tube end (i.e. displacement antinode) [32]. The reduced pressure in the mouth tube makes the pressure approach a 'virtual node' to the left of the real pressure node (shown by smaller dashed lines). i.e. the acoustic length of the combined tubes is diminished in comparison to a uniform tube by an amount ΔL (as shown).

If we recall from equation 2.1, the inverse relation between formant frequencies (F_n) and tube length (L) shows that shorter tubes have higher resonance frequencies, therefore the net effect on this two-tube system will be a rise in F_1.

This situation is reversed for F_2, as shown in Figure 5.5(c). Overall pressures are much higher in the pharynx, therefore the tube representing this region tends to govern this formant. F_2 is effectively lowered because the pressure node is shifted toward the right, close to the two-tube junction. That is to say, due to the higher pressures in the pharynx tube (and lower pressures in the mouth), the acoustic length of the pharynx has been lengthened for F_2 in comparison to a uniform tube by an amount ΔL (as shown).

In contrast to F_1, the longer acoustic length of the pharynx tube will result in lower resonance frequencies, therefore F_2 will be decreased.

It should be noted that, although one tube is much wider than the other, this is immaterial to the calculation of the resonant frequency as the wavelength (which determines the frequency) depends only on the length of the tube in the model of the vocal tract that we are considering [39]. This holds true as long as the width is less than a quarter of the wavelength. This is the case in a normal vocal tract although not in the diagrams we have been considering, in which the width has been exaggerated to emphasise important features.

Appropriate values for vocal tract resonances for the vowel [a] are:

- F_1 = 700Hz
- F_2 = 1500Hz

5.4.2 Vowel [e]

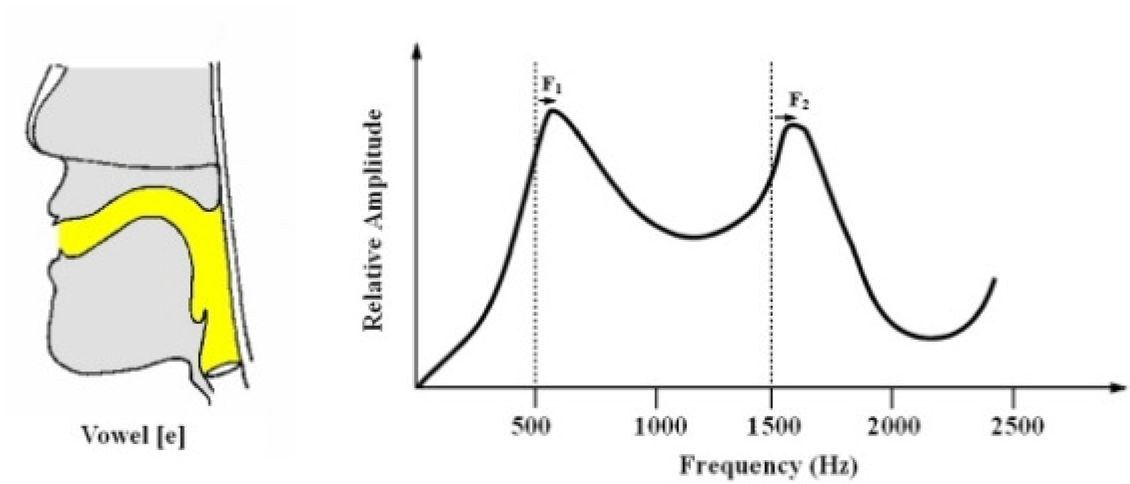


Figure 5.6: Midsagittal vocal tract profile and F_1/ F_2 frequency spectra for vowel [e]

As shown in Figure 5.6, the vowel [e] requires a comparatively open mouth cavity with a central medial tongue position and a raised jaw height. This vowel does not form one of the cardinal (corner) vowels but exhibits an articulatory shape (area function) in between that of the [a] and [i] vowels. It is characterised by a particularly large cavity at the back of the mouth leading into the pharyngeal cavity.

Appropriate values for vocal tract resonances for the vowel [e] are:

- F_1 = 550Hz
- F_2 = 1750Hz

5.4.3 Vowel [i]

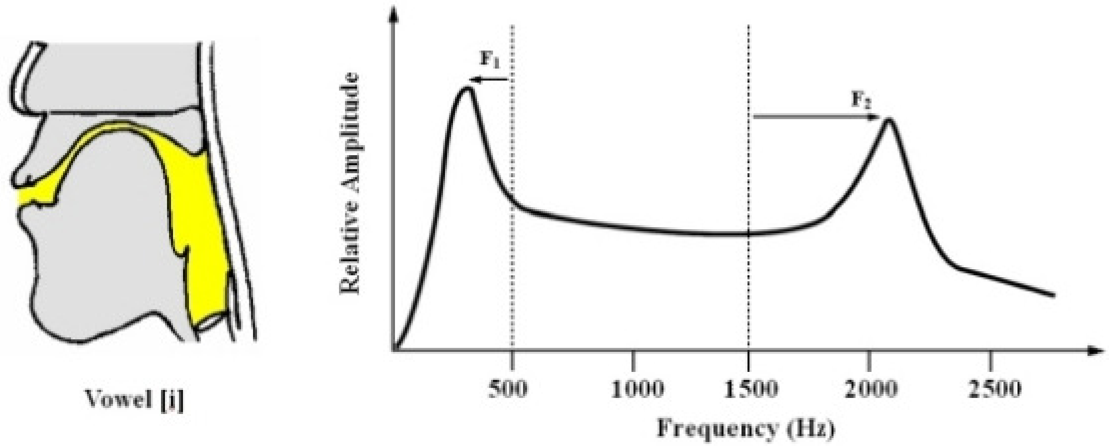


Figure 5.7: Midsagittal vocal tract profile and F_1/F_2 frequency spectra for vowel [i]

For the production of the vowel [i] (a high front vowel), the tongue assumes a position that is forward and up in the oral cavity, forming a constriction with the pharyngeal cavity (Figure 5.7).

Our comparison with a neutral tube has shown that the lowest resonance that can be produced is 500Hz. From Figure 5.7, we can see that the first formant of the vowel [i] can be below 300Hz, and therefore we must consider how such resonances arise. The solution to this problem is to consider this configuration of the vocal tract as a Helmholtz resonator [39, 74]. The front of the tongue is raised toward the hard palate (forming a narrow channel with a small body of air), and behind the tongue constriction there is a large body of air in the back of the mouth and the pharynx. It is an arrangement in which a small body of air is pushed into the larger body and then released. Acting somewhat like a mass on a spring displaced from equilibrium, the body of air will compress then expand (driven out by the increase in cavity pressure) and overshoot its point of equilibrium. This will produce a slight vacuum in the cavity causing the body of air to oscillate in and out of the system for a few cycles at a natural frequency. If oscillation is sustained, the system will exhibit a single resonant frequency (Helmholtz resonance).

With reference to the two-tube approximation of the vowel [i] (Figure 5.8(a)), the frequency of the Helmholtz resonance can be calculated as follows [39]:

$$f = \frac{c}{2\pi}\sqrt{\left(\frac{A}{VL}\right)} \tag{5.1}$$

where A is the area of the opening port, of length L, and V is the volume of the cavity. From values published in [33, 39, 69, 75], based on a constriction (L) in the area of the hard palate (12-13cm from the glottis), the area of the opening port can be taken as 15mm^2, with a constriction length of 1cm, and a cavity volume behind the constriction in the mouth and pharynx of 60cm^3. The frequency of the Helmholtz resonance would then be in the region of 270Hz, an appropriate figure for the first formant of the vowel [i] (see Figure 5.1).

The standing wave pressure patterns for F_1 and F_2 of the two-tube approximation of the [i] vowel are shown in Figures 5.8(b) and 5.8(c) (adapted from Titze [34] p.166). Figure 5.8(b) shows the first formant pressure pattern to be nearly constant in the pharynx (although it would approach a 'virtual' pressure node far beyond the diagram). The effect of this is to acoustically lengthen the tube, lowering F_1, and producing the

Helmholtz resonance as described previously. Figure 5.8(c) shows the mouth tube to be the dominant resonator for the second formant. The reduction in pressure in the pharynx shifts the pressure node toward the tube interface.

The mouth tube is approximately a half wavelength resonator as it is nearly open at both ends. The half wavelength of the mouth tube is shorter compared to that of the neutral tube, with the pressure node shifted a distance ΔL, thus raising F_2.

Comparing the displacement of F_1 and F_2 with the neutral vowel, F_1 is lowered and F_2 is raised.

In contrast to the [a] vowel, where the pressure patterns moved toward each other over the length of the vocal tract, the mouth narrowing for the [i] vowel spreads F_1 and F_2 apart.

Appropriate values for vocal tract resonances for the vowel [i] are:

- F_1 = 270Hz (Helmholtz resonance)
- F_2 = 2100Hz

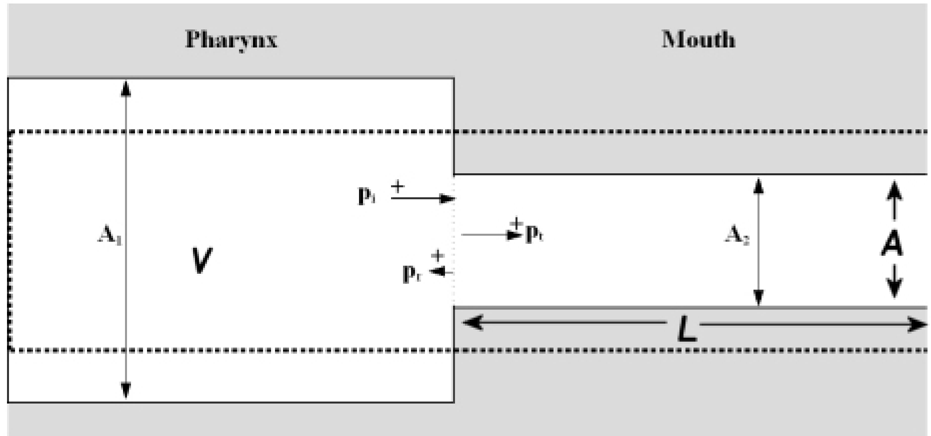


Figure 5.8(a): Two tube approximation for vowel [i]

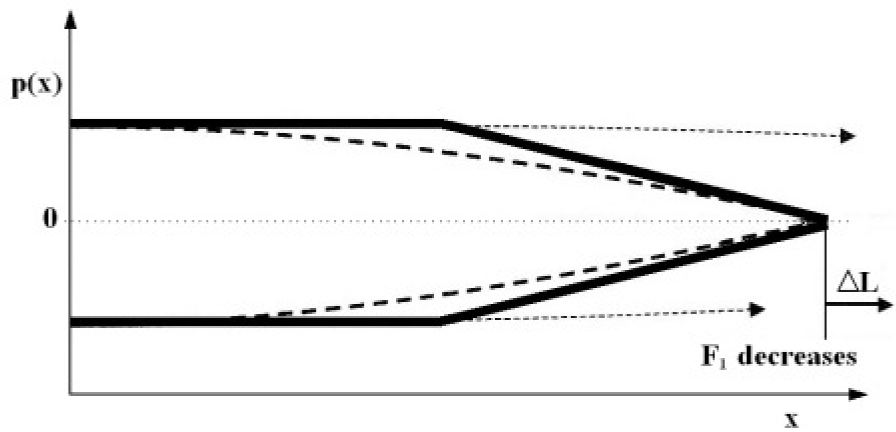


Figure 5.8(b): Standing wave pressure pattern for F_1

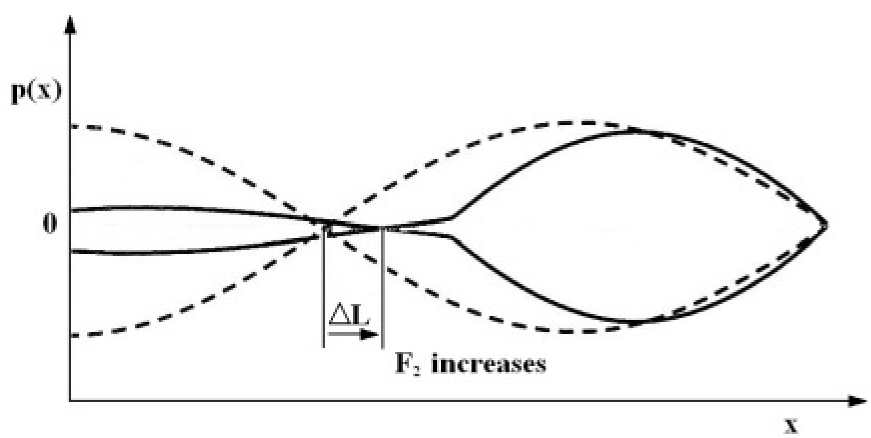


Figure 5.8(c): Standing wave pressure pattern for F_2

5.4.4 Vowel [o]

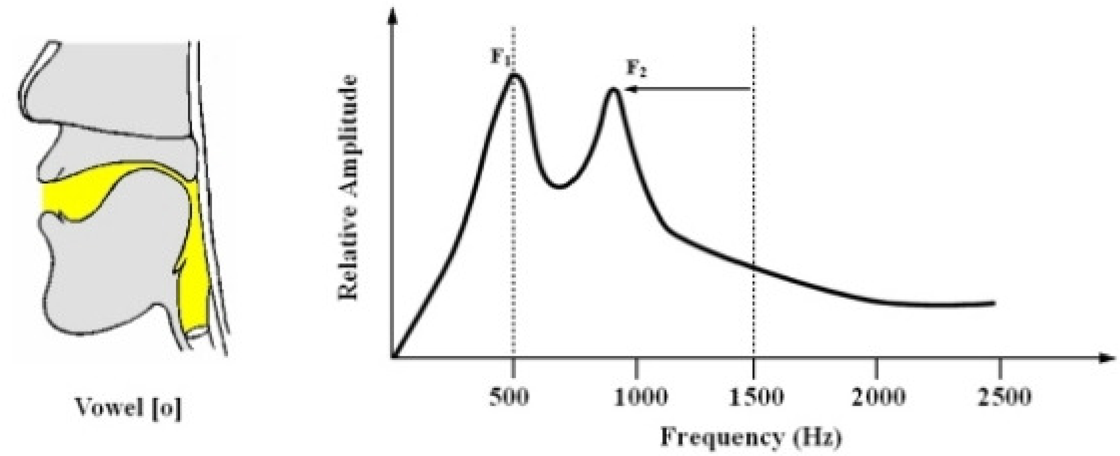


Figure 5.9: Midsagittal vocal tract profile and F_1/ F_2 frequency spectra for vowel [o]

The vowel [o] exhibits a vocal tract profile similar to that of the [a] vowel with a low back tongue position, but with a raised jaw height and a degree of lip rounding that acoustically elongates the vocal tract (Figure 5.9).

The combined effect of enlarging the mouth cavity and acoustically lengthening the vocal tract with lip rounding is to significantly lower the F_2 formant when compared to that of a neutral tube. This effect is described more fully for the corner vowel [u] in section 5.5.5.

Appropriate values for vocal tract resonances for the vowel [o] are:

- F_1 = 500Hz
- F_2 = 800Hz

5.4.5 Vowel [u]

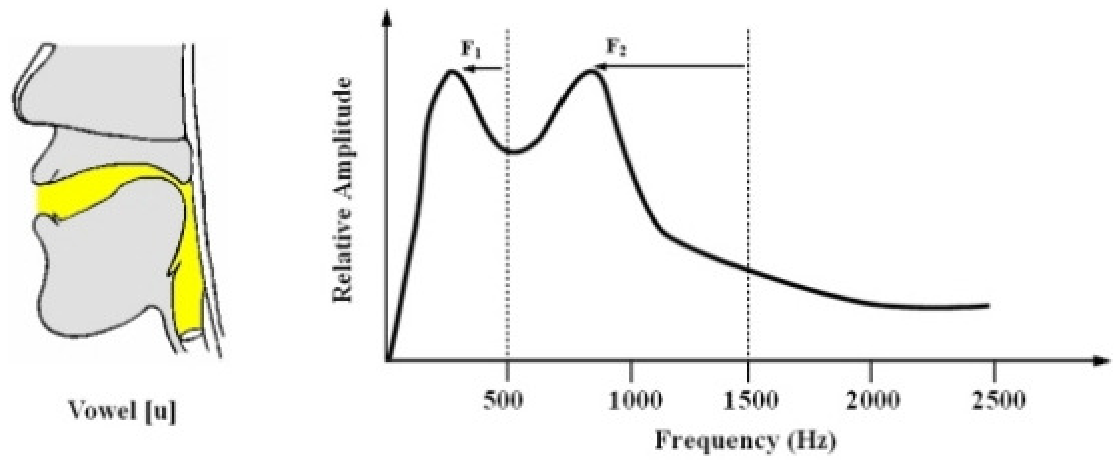


Figure 5.10: Midsagittal vocal tract profile and F_1/ F_2 frequency spectra for vowel [u]

For the production of the vowel [u] (a high back vowel), the tongue assumes a position that is back and up in the oral cavity, forming a constriction with the pharyngeal cavity (Figure 5.10). Although a similar size of constriction compared with that of the [i] vowel, the position of this constriction is further back in the velar region, and is approximately 10-12cm from the glottis. This vocal tract configuration can also be treated as a Helmholtz resonator with a cavity volume (V) of ~50cm^3. From equation 5.1, we can see that the frequency of this Helmholtz resonance would be around 300Hz, which is an appropriate F_1 value for an [u] vowel.

For the configuration of the [u] vowel, a two-tube approximation is insufficient to capture the effect of velar narrowing in combination with lip rounding. Other more detailed multi-tube models have been proposed by Fant et al. [33], to more accurately represent this configuration of the vocal tract. It has been found however [39, 69], that apart from forming a constriction with the velar region, the tongue is no longer the primary articulatory feature for the [u] vowel, and it is instead, the effect of lip rounding that is more prominent.

Lip rounding has little effect on F_1 (Helmholtz resonance) but has a much greater influence on F_2. This is because F_2 is affected by changes in the front (mouth) cavity, where the lips are the boundary to free space. To isolate the effect that lip rounding has on the standing wave pressure pattern, we can approximate the system using the two-tube model as before, but acknowledge that this is insufficient to model the overall configuration of the [u] vowel since the pharynx has been excluded from the model. Figure 5.11(a) shows a two-tube approximation, with the large tube representing the mouth cavity and the small tube representing the lips (approximating the aperture for lip rounding). The standing wave pressure patterns for F_1 and F_2 are shown in Figures 5.11(b) and 5.11(c) respectively (adapted from Titze [34] p.167). Comparing them with those for a uniform tube without lip rounding (shown as dashed lines), we can see that as pressure builds up behind the lips, the pressure node is shifted beyond the lips, therefore acoustically lengthening the tubes which lowers all formants (having a greater effect on F_2 than F_1).

In relation to the vowel [u], the combined effect of the constriction formed by the tongue moving further back towards the glottis, lengthening the front cavity, and the effect of lip rounding, both contribute to lowering F_2 as shown spectrally in Figure 5.10.

Appropriate values for vocal tract resonances for the vowel [u] are:

- F_1 = 300Hz (Helmholtz resonance)
- F_2 = 800Hz

The effect of lip rounding will be examined further in Chapter 7.

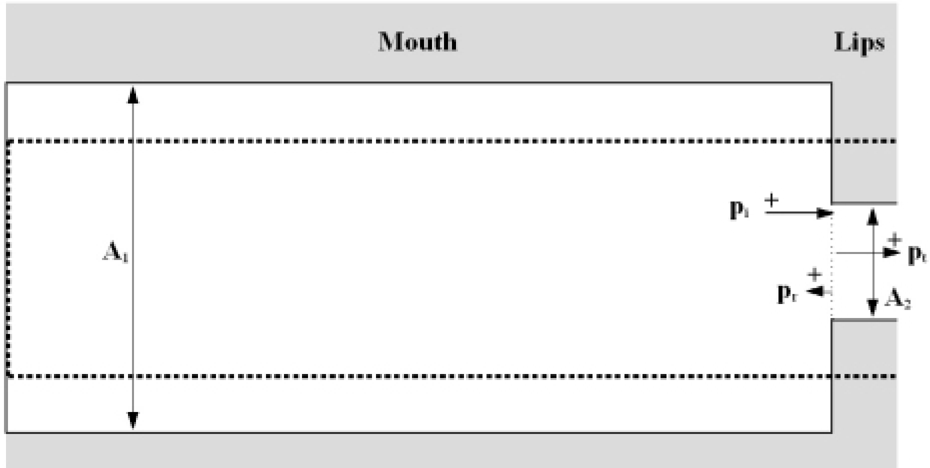


Figure 5.11(a): Two tube approximation of mouth and lips for vowel [u]

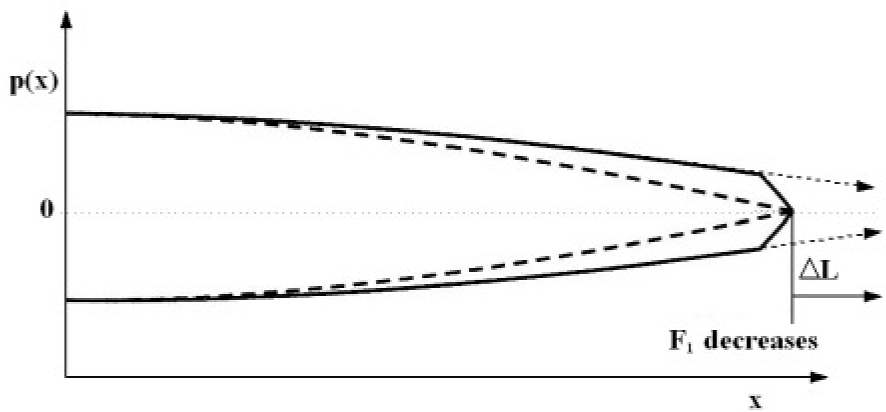


Figure 5.11(b): Standing wave pressure pattern for F_1

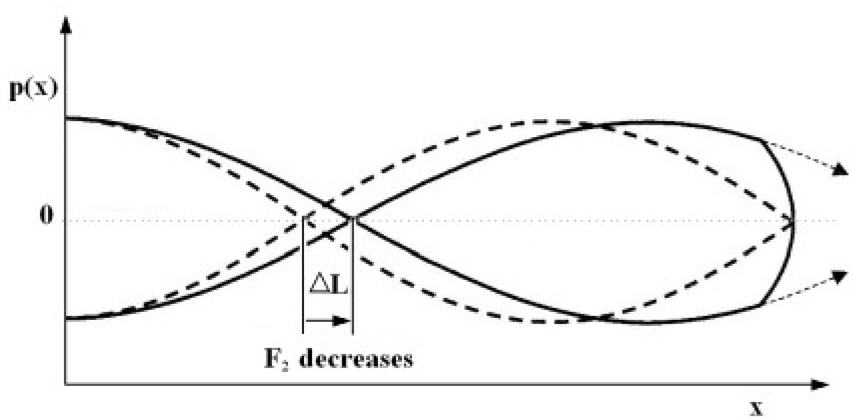


Figure 5.11(c): Standing wave pressure pattern for F_2

5.5 Optimisation of the APR procedure for measuring the vocal tract

In contrast to previous studies of airway measurement [7, 14-17, 22, 23, 76] this study is focussed specifically on the vocal tract (the section of the airway from lips to glottis).

Two issues must be addressed in using APR for vocal tract measurement.

Firstly, APR is unable to measure a branching pipe such as the nasal cavity at the same time as measuring the primary tubular object (vocal tract). For the purposes of this study, this restriction is of little consequence as the vowels being measured are non-nasalised and are most prevalent in singing. Therefore, the nasal cavity will not be measured.

Secondly, the subject must refrain from breathing during the measurement and is therefore unable to phonate the vowel when coupled to the source tube as this would pressurise the tube and interfere with the acoustic input signal.

A mouthpiece with valve system that enabled the subject to breathe whilst coupled to the reflectometer was implemented by Marshall [7] for airway measurement. However, this was required to accommodate longer measurement times using individual input pulses, and the subject was still unable to phonate. A solution employed throughout this study was to ask the subject(s) to imitate the phonation of the vowel whilst a measurement was taken. The subject, in an upright position, was first asked to sing the vowel at a comfortable pitch within his/her range. After several seconds, the subject was instructed to halt phonation by holding their breath, freezing the position of their articulators, and then placing their lips against the mouthpiece connected to the source tube (Figure 5.12).

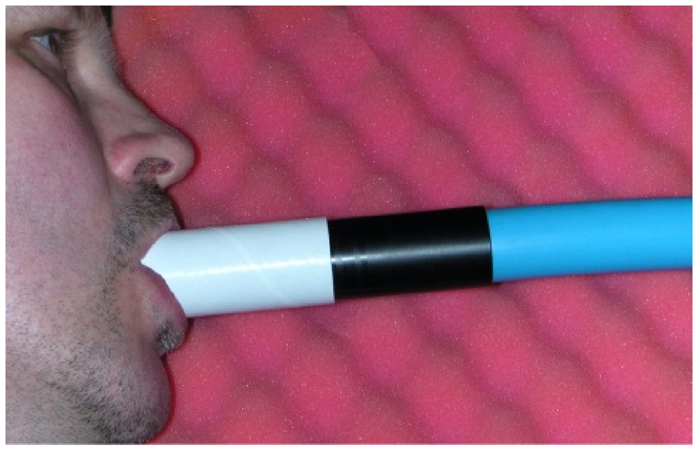


Figure 5.12: Subject coupling with mouthpiece and source tube

To prevent any unnecessary repositioning, the mouthpiece was positioned close to the subject's mouth and could be readily moved by the subject into position once phonation was halted. Once comfortably coupled with the mouthpiece, the measurement phase was initiated by a single button press on the computer keyboard. Twenty measurements of each vowel for each order of MLS signal (m = 16, 17 and 18) were performed using an adult male, resulting in a data set of 300 measurements. As the procedure for performing each of the measurements was quite tiring, the study spanned several consecutive days to prevent fatigue.

During the course of the experiment, it was clear that not every measurement would provide a satisfactory result. However, in order not to affect the outcome of repeatability, the area reconstructions and data analysis were not performed until the end of the study. The only time when measurements were redone was when the subject interrupted the measurement by decoupling himself from the mouthpiece during the recording phase. Reasons for decoupling included the feeling of being about to sneeze or choke, acoustic leakage from lip seal with mouthpiece or nasal cavity with velar lowering, forgetting articulatory positioning, and fatigue.

At the end of the study, the bore reconstruction algorithm was used with each of the impulse response measurements of the vocal tract for the studied vowels. Before averaging the twenty profiles of each MLS data set, the reconstructions were inspected manually. Those profiles showing changes in cross sectional area that were outwith the feasible constraints of the vocal tract during vowel production (i.e. infinite or zero cross sectional area) were marked as 'failed readings', and these were not included in the averaged result. An analysis of 'failed readings' is given in section 5.7. Also, to investigate the repeatability of the technique, deviations from the mean profile were calculated using coefficients of variations (CV)[4] for each MLS data set.

Area reconstructions of the vocal tract from these measurements are presented in the following sections. The effect of a subject in an upright position compared to that of being in a supine position (as in later measurements for comparison with MRI), is explored in Chapter 7.

[4] Standard deviation divided by the mean, expressed as a percentage

5.6 APR measurements of the vocal tract

5.6.1 Vowel [a]

The area reconstructions of the vocal tract for the vowel [a] are shown in Figure 5.13(a). Several centimetres of the mouthpiece have also been reconstructed to provide a point of reference ($[a]_{MP}$) for the beginning of the reconstruction at the coupling with the lips. From this graph we can see that even although the absolute values for cross sectional area differ, the overall profiles of the vocal tract for each MLS signal are similar. Each reconstruction shows a large mouth cavity ($[a]_{ORAL}$) and a smaller pharyngeal cavity caused by a tongue that is low and pulled back into the pharynx ($[a]_{PHRGL}$), all of which are features of the vowel [a] described previously.

It is evident that the reconstruction for the lowest order input signal MLS = 16 differs significantly from the others in the pharyngeal area. At a distance of 0.15m from the mouthpiece, this signal measures an area profile that is up to 40% less than that measured by either of the other signals.

From the graph displaying the coefficient of variation as a function of axial distance, we can see that the largest measure of uncertainty across all profiles to be in the pharyngeal region, with deviations up to 50%. Overall however, the average coefficient of variation across all profiles is less than 18%.

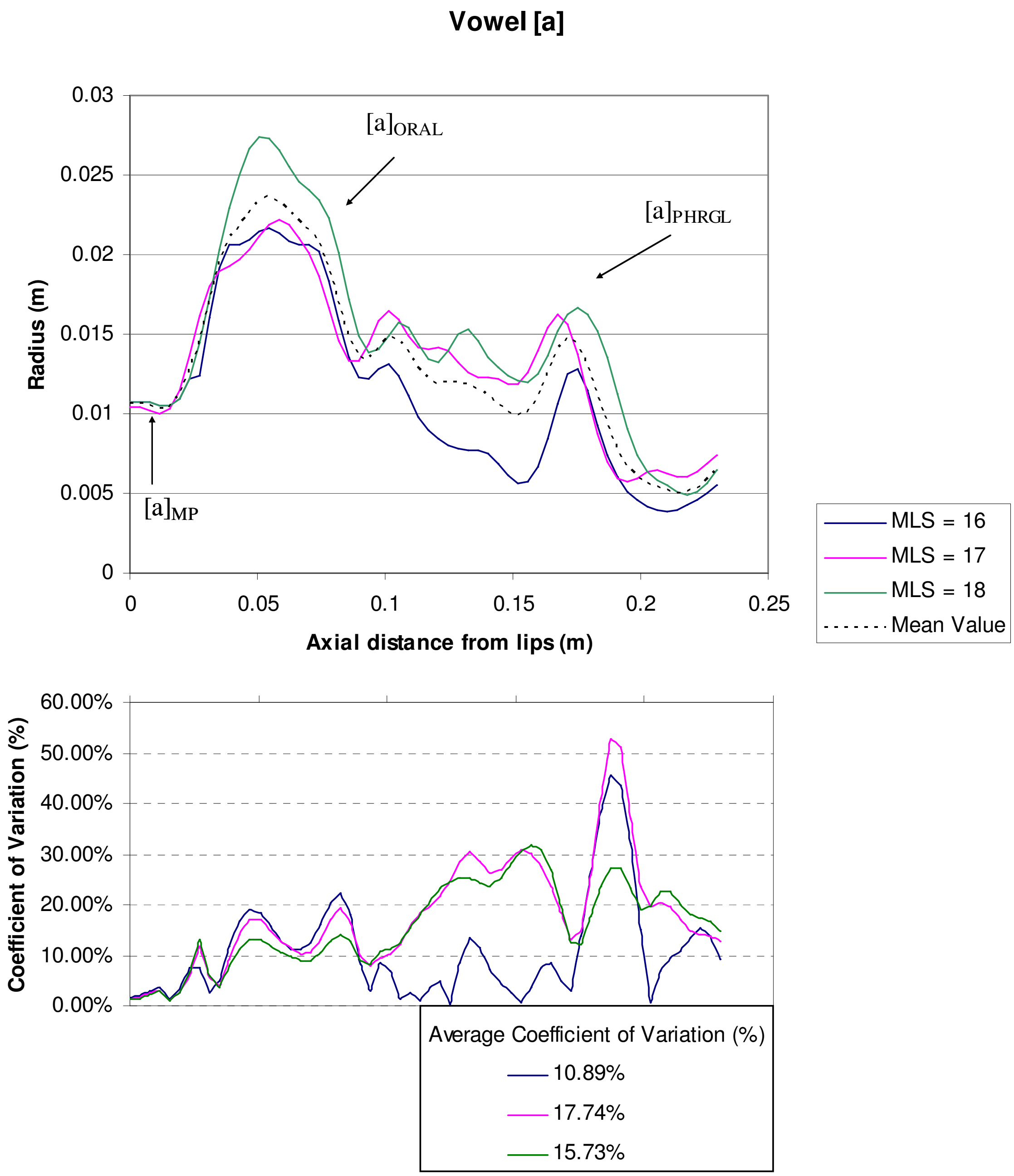


Figure 5.13(a): Vocal tract area reconstructions from APR measurements of a single subject for vowel [a], together with measure of variation in data.

Using the mean value for area reconstruction of the vocal tract for the vowel [a], a 3D area equivalent tube model was constructed (Figure 5.13(b)). Surface rendering and lighting effects have been applied to help visualise the model.

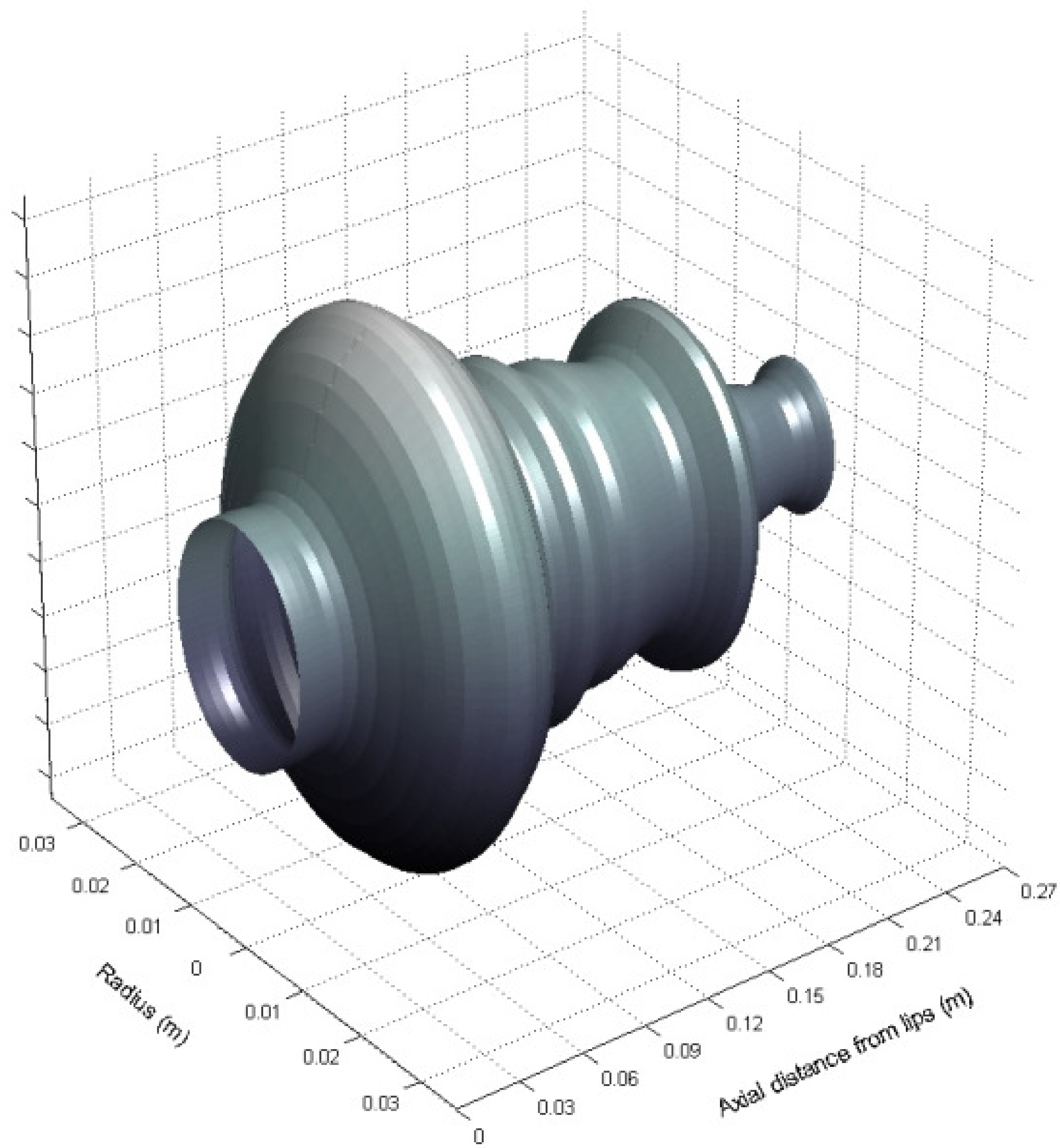


Figure 5.13(b): 3D area equivalent tube model for vowel [a]

5.6.2 Vowel [e]

The area reconstructions of the vocal tract for the vowel [e] are shown in Figure 5.14(a). As before, the overall profiles of the vocal tract for each MLS signal are very similar.

The coupling between the lips and mouthpiece at point ($[e]_{MP}$) and region extending into the mouth cavity ($[e]_{ORAL}$) are very closely matched between the area functions, with deviations of less than 5% from the mean profiles of each MLS signal.

The characteristic central tongue position and raised jaw height of this vowel are particularly evident in these measurements as it creates a large cavity at the back of the mouth leading into the pharyngeal cavity, which can be seen at 10-15cm along the reconstructed length.

The average coefficient of variation over the whole length of the measured area functions is less than 20%. There are however, large differences in the measured values for cross sectional area in the pharyngeal cavity ($[e]_{PHRGL}$), with variations of over 40%.

A 3D area equivalent tube model using the mean value for the vowel [e] is shown in Figure 5.14(b).

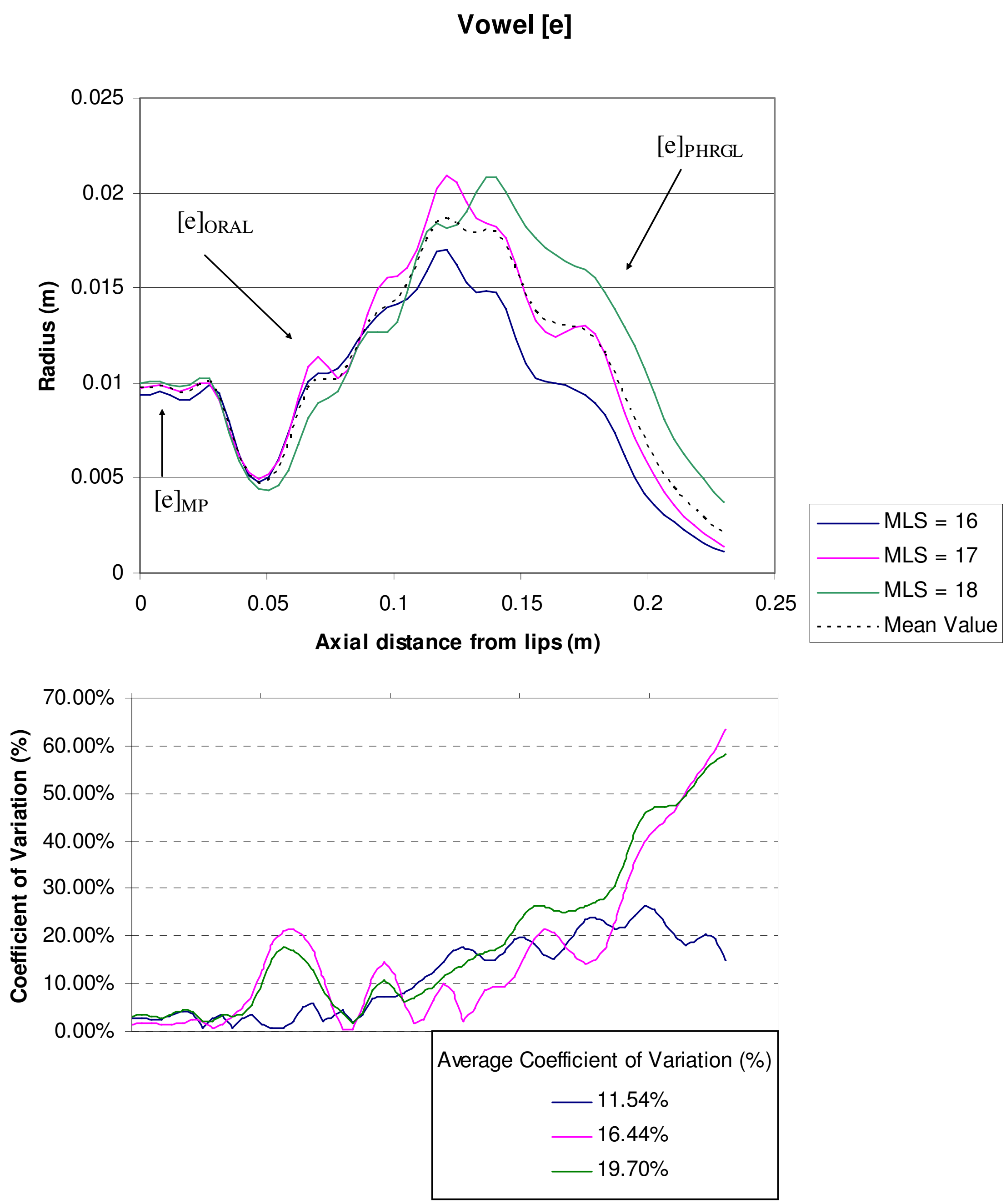


Figure 5.14(a): Vocal tract area reconstructions from APR measurements of a single subject for vowel [e], together with measure of variation in data.

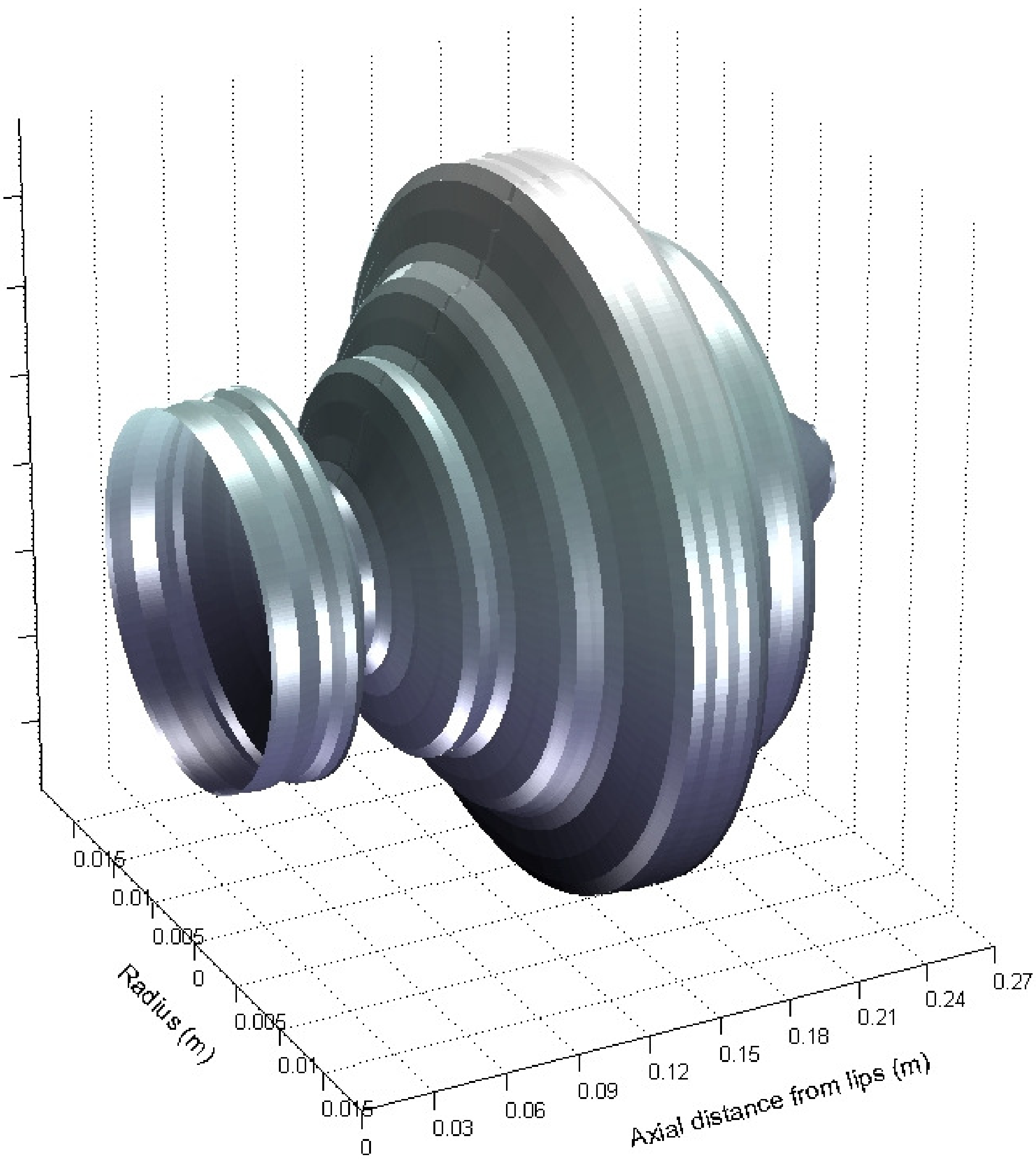


Figure 5.14(b): 3D area equivalent tube model for vowel [e]

5.6.3 Vowel [i]

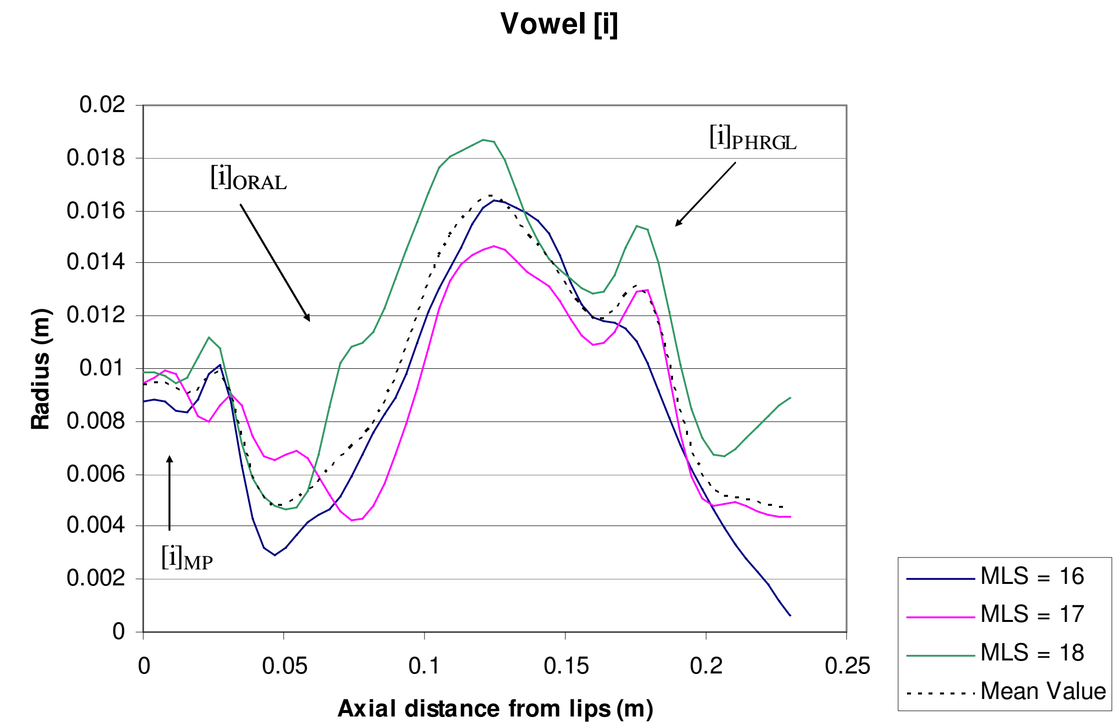
The area reconstructions of the vocal tract for the vowel [i] are shown in Figure 5.15(a). In comparison to the previous vowels, the vowel [i] features a high degree of lip spreading making it more difficult to couple with a cylindrical mouthpiece. Evidence of this can be found at point ($[i]_{MP}$) where there are greater deviations around the mouthpiece-lip coupling than seen in the previous measurements, with variations up to 20%.

A distinguishing feature of the vowel [i] is a constriction between the hard palate and the tongue that is forward and up in the oral cavity ($[i]_{ORAL}$). The size of the measured constriction is congruent between the signals, with a radius around 3-5mm, although the place of constriction ranges from 5cm to 8cm along the axial distance from the first point of reconstruction.

The large cavity created behind the constriction has a similar profile between the measurements with variations at the maximum radius of around 10%. The measured profiles for MLS=17 and MLS=18 show a rise in the pharyngeal area ($[i]_{PHRGL}$) before a sharp decrease in cross sectional area towards the glottis. The profile for MLS=16 on the other hand shows a much more gradual decrease in cross sectional area and does not feature this rise in the pharyngeal region.

From the graph showing the coefficient of variation for each mean profile, we can see large differences in the measured values for cross sectional area at the front of the mouth cavity and near the glottis, with variation in measured values in excess of 50%. However, the average coefficient of variation across profiles is less than ~27%.

A 3D area equivalent tube model using the mean value for the vowel [i] is shown in Figure 5.15(b).



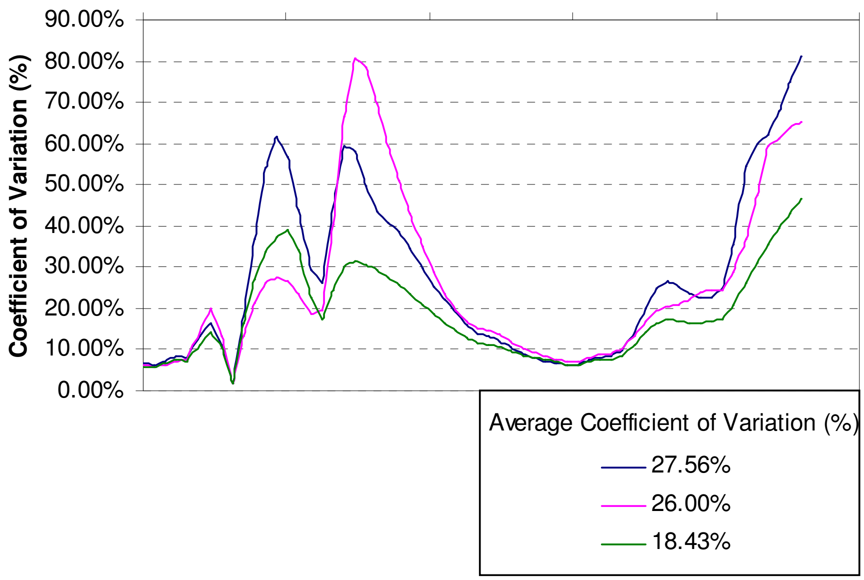


Figure 5.15(a): Vocal tract area reconstructions from APR measurements of a single subject for vowel [i], together with measure of variation in data.

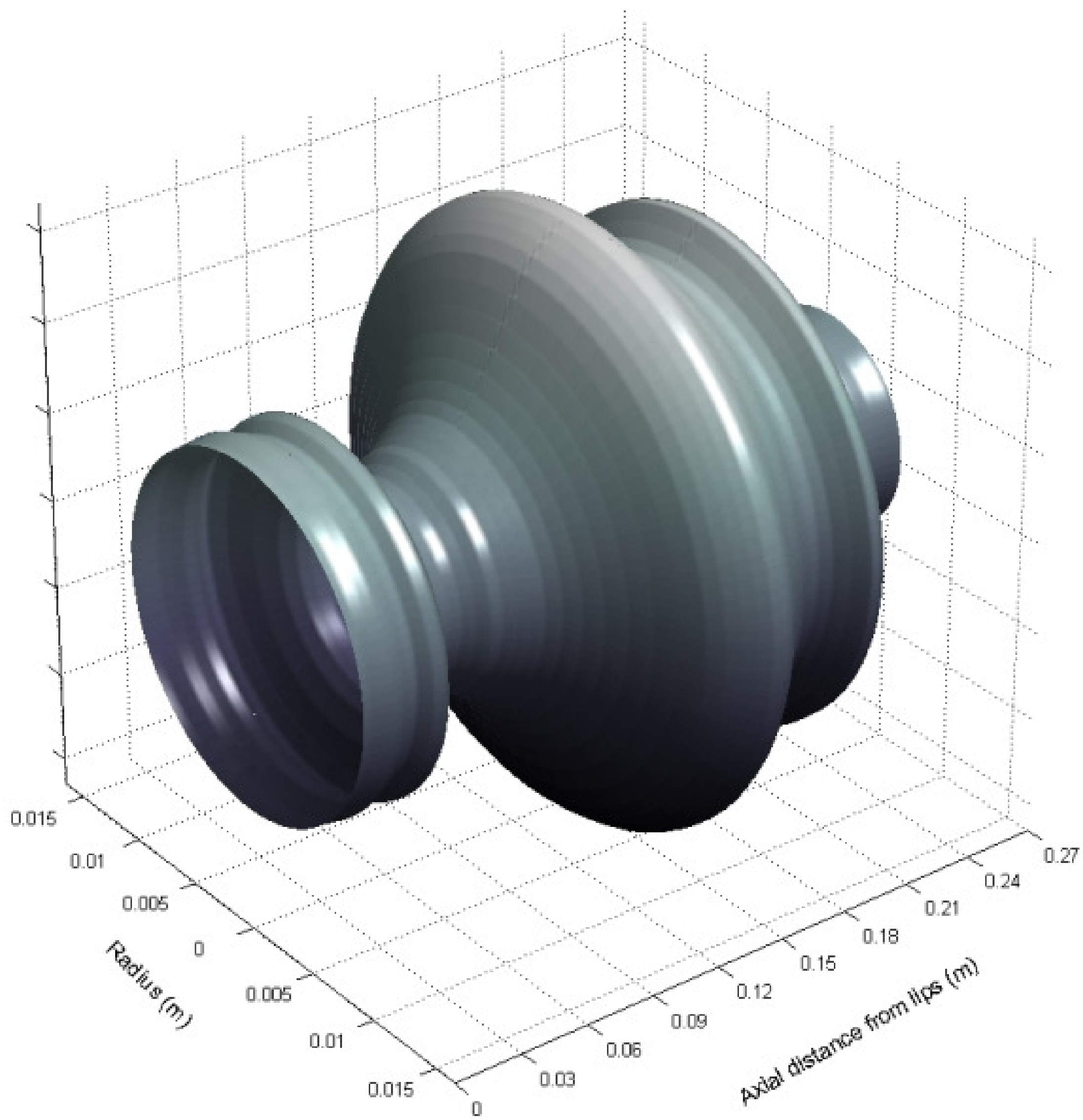


Figure 5.15(b): 3D area equivalent tube model for vowel [i]

5.6.4 Vowel [o]

The area reconstructions of the vocal tract for the vowel [o] are shown in Figure 5.16(a). In comparison to the vowel [i], this vowel exhibits a large degree of lip rounding. Although the mouth shape is more conducive to coupling with the cylindrical mouthpiece, the actual size of the mouth opening is smaller, therefore the subject has to accommodate this with a more emphasised lip opening ($[o]_{MP}$).

A large mouth cavity ($[o]_{ORAL}$) formed by a low back tongue position and raised jaw, together with a rise in cross sectional area in the pharynx ($[o]_{PHRGL}$), show similarities with the vocal tract profile of the vowel [a] as mentioned in the previous section. The main difference is in the lip shape (rounding and protrusion), which is not captured in these experiments due to the fixed dimensions of the cylindrical mouthpiece.

Although slightly offset from each other, the vocal tract profiles of the MLS=17 and MLS=18 signals are very similar, with only small measured differences between cross sectional areas. The profile for MLS=16 does however show increases in cross sectional area for both mouth and pharyngeal cavities, but underestimates these regions by up to 40% compared with those measured using the higher order sequences. Large differences in the measured radii are also present at the glottis where MLS=16 does not show the sharp constriction within the airway that has been measured using the higher order signals.

Examining the graph showing the coefficient of variation as a function of axial distance, we can see larger variations across all measurements over the length of the vocal tract compared to the previous vowels. Whilst there are points where the measured coefficients of variation are in excess of 50%, the average variation across all profiles is less than ~32%.

A 3D area equivalent tube model using the mean value for the vowel [o] is shown in Figure 5.16(b).

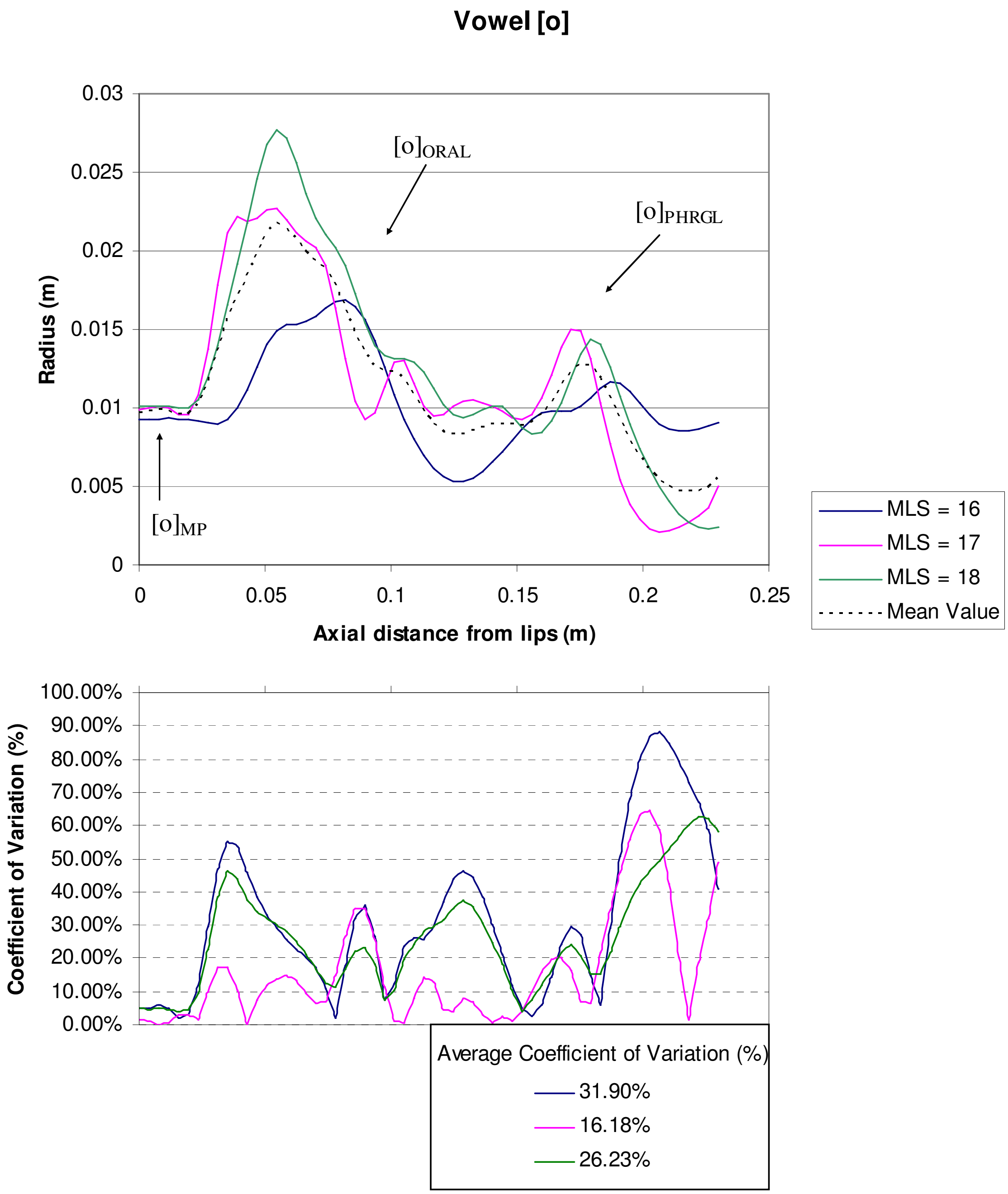


Figure 5.16(a): Vocal tract area reconstructions from APR measurements of a single subject for vowel [o], together with measure of variation in data.

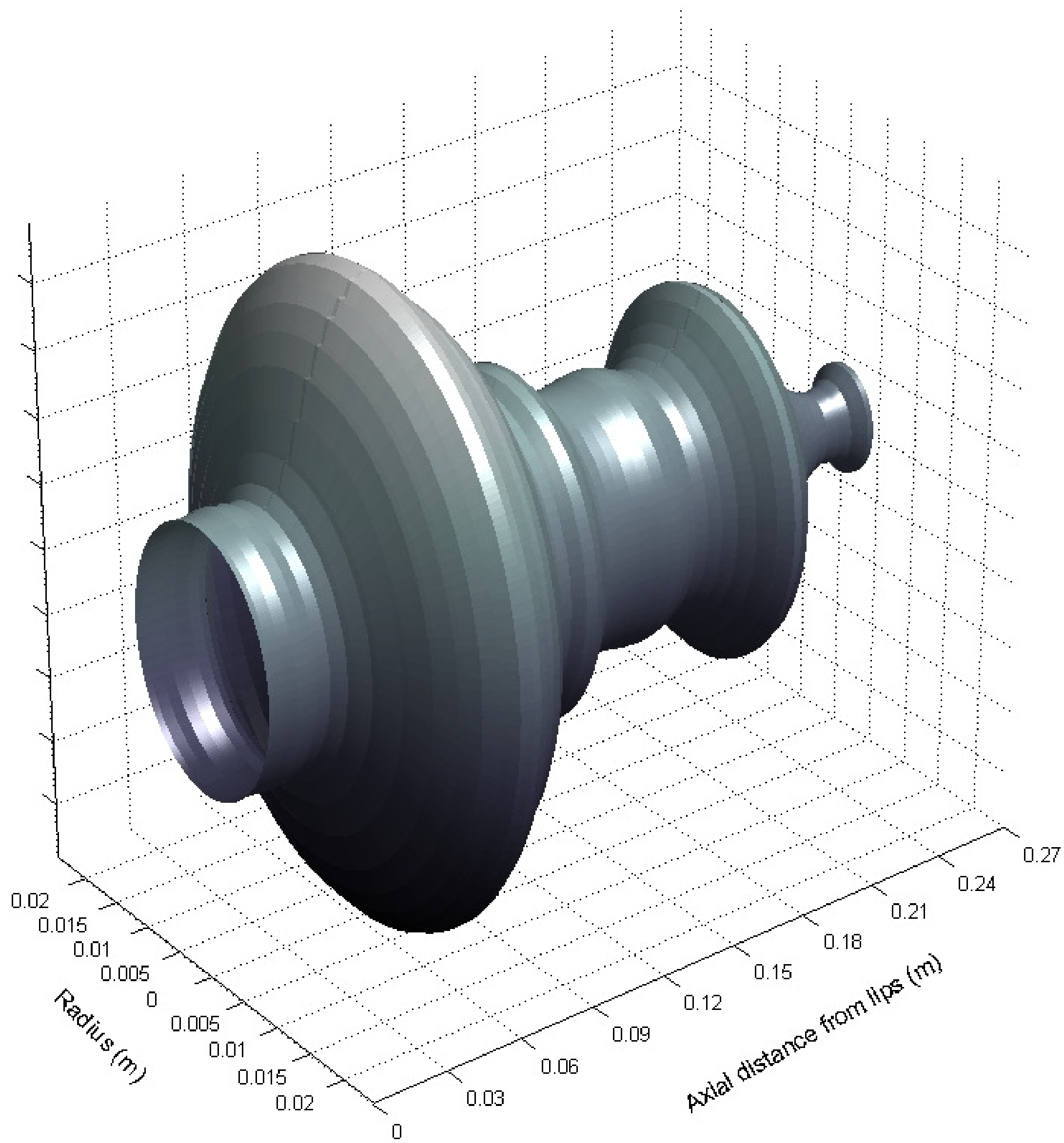


Figure 5.16(b): 3D area equivalent tube model for vowel [o]

5.6.5 Vowel [u]

The area reconstructions of the vocal tract for the vowel [u] are shown in Figure 5.17(a). Of all the vowels studied, the vowel [u] features the most significant amount of lip rounding and protrusion, which again caused difficulties in coupling with the cylindrical mouthpiece as there was a mismatch in area, causing the subject to widen his lips around the end of the mouthpiece to form a seal. We can see the effect of this from the point ($[u]_{MP}$) up to an axial distance of 7cm from the first point of reconstruction, where the lips have had to flare to accommodate the size of the mouthpiece. There is a large coefficient of variation (up to 80%) across all profiles in this region.

The small rise in area at ($[u]_{ORAL}$) is the area created by the tongue that is high and back in the mouth cavity .

The reconstructed profiles for both MLS=17 and MLS=18 are very similar, and exhibit a large increase in pharyngeal area ($[u]_{PHRGL}$) after a small constriction in the mouth cavity. However, the MLS=16 signal has significantly underestimated the complex profile of this vowel, particularly in the pharyngeal region at a distance of 15cm from the first point of area reconstruction. This profile does not exhibit the same large rise in pharyngeal area or sharp decrease towards the glottis that are found in the reconstructions of the higher order signals.

The average coefficient of variation is less than ~25% for all profiles, but there are at points large deviations from this, most notably with the MLS=16 reconstruction.

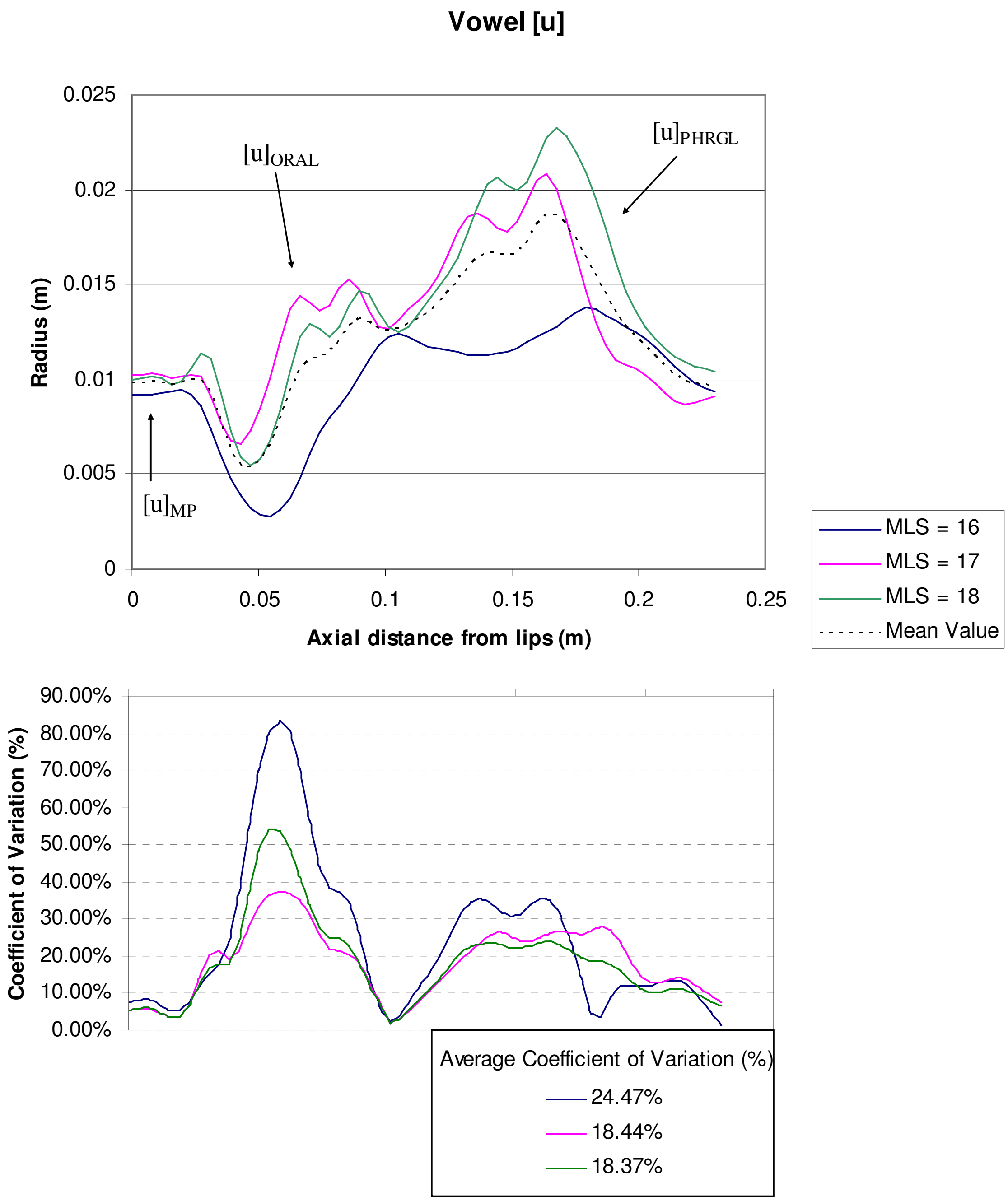


Figure 5.17(a): Vocal tract area reconstructions from APR measurements of a single subject for vowel [u], together with measure of variation in data.

A 3D area equivalent tube model using the mean value for the vowel [u] is shown in Figure 5.17(b).

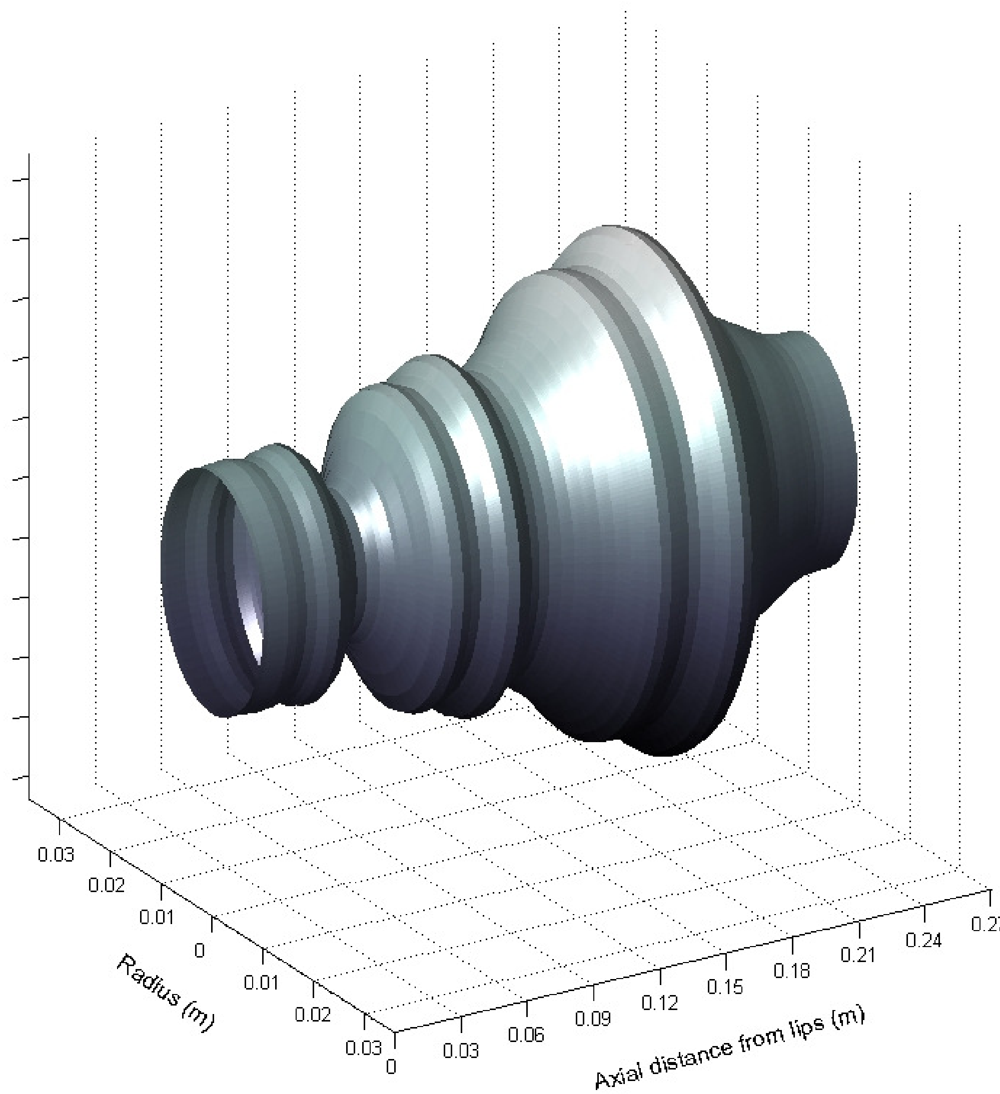


Figure 5.17(b): 3D area equivalent tube model for vowel [u]

5.7 Analysis of results

The study had two clear objectives: to measure the area profile for each vowel and to provide information as to the accuracy and repeatability of the technique.

The accuracy of the technique is a difficult quantity to measure as there were no pre-existing data sets to compare with. The overall profiles can be compared with those of similar studies [77, 78], but no direct comparisons can be made because of differing methods and subjects used in these experiments.

To assess the repeatability of the technique for a given order of MLS signal, twenty measurements per vowel were taken. By calculating the mean of the twenty profiles, deviations from the mean were calculated using coefficients of variation. This gave a measure of variability across a reconstruction and was adopted as a method to measure the repeatability of the technique.

The average coefficient of variation was generally less than 20%, with a few exceptions for the vowels [i] and [o]. We can therefore deduce that the technique is repeatable for each order of MLS signal and within acceptable limits when compared with a clinical physiological technique [7, 17].

A graph of the percentage of failed readings for each of the vowels and MLS input signal is shown in Figure 5.18.

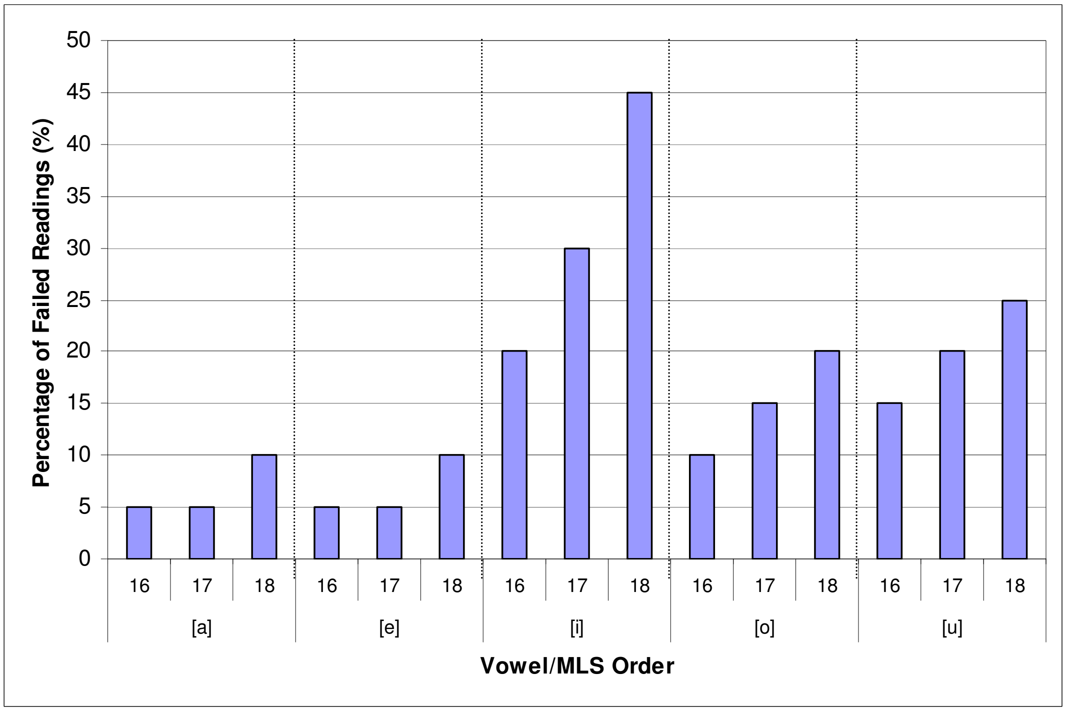


Figure 5.18: Percentage of failed readings for each vowel and MLS input signal

As would be expected, the data shows a higher incidence of failed readings for the higher order MLS input signals. That is to say, the longer the recording time, the greater the chance of movement within the vocal tract interfering with the measurement.

The proposed reasons for failed readings include:

- involuntary articulatory movement – the human vocal tract is a dynamic biomechanical system used in breathing, swallowing and phonation, and is never truly static for extended periods of time (a requirement for this technique).
- constriction – related somewhat to involuntary movement, the vowels that experience points of constriction along the vocal tract (such as [i] and [u]) are affected by tongue drift. As there is no airstream keeping the constriction open,

the muscular shaping of the tongue has a tendency to drift upwards and occlude the airway. This effect is shown in the reconstruction by a large decrease in cross sectional area to the point of closure (zero cross section).

- breathing – during the longer measurement times there was a natural tendency to breathe via the nasal airway. With the lowered velum, the nasal cavity is measured as a branching pipe, leading to an area reconstruction of infinite cross section.
- acoustic leakage – the two areas where acoustic leakage can occur are with the lip seal and during nasal breathing. The reconstructed area profile would exhibit infinite cross section.

It is also evident from Figure 5.18 that the vowels which exhibit large degrees of lip spreading or rounding (such as the vowels [i], [o] and [u]) have a higher percentage of failed readings. The previous section discussed the difficulties in trying to couple these lip shapes to the cylindrical mouthpiece.

Deviations between profiles were also presented in the previous section, and although this is a useful means of comparing the cross sectional differences between reconstructions, there is no absolute point of reference to align and compare profiles. For example, the area reconstructions for the vowel [i] (Figure 5.15(a)) show a similar overall profile, but when comparing cross sectional values at a fixed point along the axial distance from the mouthpiece-lip coupling, there can be large differences. It is clear in this example that there is a large difference in the place of constriction measured by the MLS=17 signal compared with the other two reconstructions. The size of the constriction (~4mm radius) is in line with the other measurements, but it is offset posteriorly in the mouth cavity by approximately 2cm. An explanation for this large difference can be found by examining the nature of the vowel and its mouthpiece-lip coupling. The vowel [i] has a raised jaw with high front tongue position together with a large degree of lip spreading. The constriction is formed between the tongue and the hard palate approximately 2cm from the alveolar ridge. By changing the lip shape from spread to round (to accommodate the cylindrical mouthpiece), the tongue experiences greater upward flexion forming a large tongue groove that displaces the point of constriction along the posterior axis. If care is not taken by the subject to adjust for this change then the measured point of constriction will be offset from the real place of articulation. After a short trial period, the subject became comfortable with the procedure of firstly singing the vowel then halting phonation, freezing the articulatory position and then coupling with the mouthpiece. However, imitating the articulatory position without phonation gave the subject no audible feedback as to the accuracy of placement. The unnatural lip shape of this vowel to accommodate the mouthpiece and involuntary tongue movement occluding the airway gave rise to a high percentage of failed readings, particularly with the higher order (longer duration) input signals (see Figure 5.18).

The previous section presented the area reconstructions for successful measurements of each vowel. The effect of the input signals on area reconstruction can be summarised as follows:

- MLS=16 – This was the lowest order signal which gave the quickest measurement time and least amount of failed readings. Although showing some of the main features for the reconstructed profiles of each vowel, the measured cross section of the mouth cavity is generally less than that of other studies [77, 78], and significantly underestimates the pharyngeal area for vowels with constrictions.
- MLS=17 – This middle order signal produced similar area reconstructions to the highest order signal but in half the time. There was a slight increase in the overall number of failed readings when compared with the lowest order signal but the measured cross section for both oral and pharyngeal regions is more in line with that of previous studies.
- MLS=18 – This highest order signal produced area reconstructions that were the closest to the measured profiles of previous studies. The measurement time is twice that of the MLS=17 signal and four times that of the MLS=16 signal. It also exhibited the largest number of failed readings for all the measured vowels.

From these results, two features are apparent:

Firstly, the area reconstructions using lower order MLS input signals show significant underestimation of the pharyngeal area. This can be explained by the shorter measurement time resulting in less energy being input into the system. The effect is compounded by the losses from the yielding wall boundaries of the vocal tract which become more significant as axial distance increases from the source.

Secondly, the technique requires the object being measured to be static. The vocal tract and articulators are by nature non-stationary. The tongue in particular experiences involuntary movements so although the subject is able to imitate the vocal tract shape for each vowel without phonating, deviations which may not be apparent in everyday speech are noticeable on a time averaged measurement technique such as APR using MLS input excitation. Therefore, due to the essential biomechanical functions used in breathing, swallowing and phonating, it is important that experiments are carried out to obtain results speedily and non-invasively.

From these experiments, the input signal MLS=17 is a good compromise between the need for a quick measurement and the desire for a long time interval to put energy into the system.

5.8 Discussion

A summary of the work and conclusions in this chapter is found here.

An analysis of five non-nasalised vowels has been presented and the effect of altering the air cavity within the vocal tract has been described in both acoustic and articulatory terms. The first two resonances of the vocal tract (F_1 and F_2) for each of the five vowels have been shown to depend on three factors:

- the position of the point of maximum constriction along the vocal tract (which is controlled by the forward and backward movement of the tongue)
- the cross sectional area of the maximum constriction (which is controlled by the movements of the tongue toward and away from the roof of the mouth and the back of the pharynx)
- the position of the lips (where lip narrowing has been shown to diffuse the vowel spectrum, spreading F_1 and F_2 apart, as in the case of the vowel [i]; lip rounding shifts all formant frequencies of the vocal tract downward and is the primary articulatory gesture for the vowel [u]).

The procedure for measuring a human subject using APR was introduced, and the results from a study measuring five non-nasalised vowels using different excitation signals were presented. Bore reconstructions of these measurements were evaluated together with a discussion on errors and uncertainties in the measurements. A discussion on the effect of different order MLS excitation signals followed, and a conclusion on the most appropriate length of signal for future measurements was reached.

In the following chapter we will apply the technique of acoustic pulse reflectometry for vocal tract measurement to a larger group of human subjects using MLS excitation signal m=17.

Chapter 6

A group study of vocal tract measurement using APR

6.1 Introduction

Before embarking on a study to compare results of vocal tract measurement using APR against an accepted gold standard technique such as magnetic resonance imaging (MRI), it was important to establish the repeatability of the APR technique and standardise a measurement procedure (protocol) so that a benchmark could be set for comparison of results with future studies. Optimisation of the APR technique for measurement of the human vocal tract was described in Chapter 5 and the results from using a single subject were presented. The primary purpose of this series of experiments was not to construct a commercial or clinically robust reflectometer for daily use, but instead to investigate the repeatability of the technique under controlled conditions. The trial was performed using a subject who was familiar with the technique and apparatus. The subject also had the opportunity to conduct several trial runs to experience and adapt to such facets as lip-tube coupling and the positioning of articulators before any measurements were recorded. However, the opportunity of multiple trial runs and adjustment of parameters could not be readily afforded when utilising the assistance of volunteers who were unfamiliar with the APR technique, design and construction. An analysis of variance identified the most important factors to be the coupling of mouthpiece with subject, control of breathing and ability of subject to remain still throughout the measurement phase. These factors needed to be conveyed to each volunteer in a clear and concise manner in order to obtain a set of useable measurements from a relatively small sample.

This chapter describes the procedure and application of the APR technique for vocal tract measurement in a larger group of volunteers; from children to adults.

6.2 Procedure for APR in a group study

Measurements were carried out on five healthy adult volunteers having no history of airway disease other than the common cold (see table 6.1). All measurements were completed in a single sitting for each volunteer, lasting approximately 45 minutes which included rest breaks. To ensure consistency, a pre-prepared verbal statement citing the purpose of the experiment together with basic instructions and a demonstration of measurement procedure were given to each volunteer. An experiment sheet listing the five vowels for which the vocal tract would be measured was also issued. This sheet provided the volunteers with example words to ensure the correct sounding of each vowel and in turn, the static place of articulation during which they would be required to hold their breath for a stipulated time to allow a measurement to be taken. The measurement procedure was explained in detail in section 5.5 and was adopted as protocol for this group study. Only when the subject was comfortably coupled with the mouthpiece of the reflectometer and holding their breath was the measurement phase initiated. For each volunteer, 5 measurements per vowel were taken using the MLS=17 signal, and for each vowel an average profile was calculated for each subject together with coefficients of variation (CV). As with previous experiments, the area reconstructions and data analysis were not performed until the end of the study so as not to affect the outcome. Also, the only exception to retake a measurement was when the subject interrupted the measurement by decoupling themselves from the mouthpiece during the recording phase.

	Sex	Age	Height (cm)	Voice Professional	Avg. Vocal Tract Length	Failed APR Readings (%)
Subject 1	M	50	182cm	No	18.5cm ±0.8cm	28%
Subject 2	M	23	165cm	Yes	17.2cm ±1.0cm	16%
Subject 3	M	25	178cm	No	17.6cm ±1.2cm	28%
Subject 4	M	28	176cm	No	17.9cm ±0.8cm	40%
Subject 5	F	37	158cm	Yes	16.2cm ±0.8cm	24%
Child 1	*F*	*7*	*122cm*	-	*11.2cm ±0.4cm*	*80%*
Child 2	*F*	*11*	*150cm*	-	*13.4cm ±0.6cm*	*60%*

Table 6.1: Volunteer information for APR group trial of vocal tract measurement

At the conclusion of the study, the bore reconstruction algorithm was run on each of the 125 impulse response measurements of the vocal tract gathered from the five volunteers. Before calculating the average bore profile for each subject in each vowel group, the reconstructions were manually inspected and those profiles exhibiting cross sectional areas outwith the feasible constraints of the vocal tract during vowel production were marked as ‘failed readings’, and were discarded from the averaged result. An analysis of ‘failed readings’ comparing the single subject study (using MLS=17) and group study is given in section 6.5.

Between each volunteer sitting, new cap measurements were taken and the reflectometer was recalibrated and tested by performing measurements and bore reconstructions on objects of known internal area. In readiness for the next volunteer and in the interests of hygiene, the reflectometer was also cleaned and new sterile sleeves were fitted to the mouthpiece.

Area reconstructions of the vocal tract for each volunteer are presented in Figures 6.1 to 6.5, together with histograms showing the average coefficient of variation between subjects.

6.3 Results from APR measurement of the vocal tract: Adult Group Study

6.3.1 Vowel [a]

The reconstructed profiles for the vowel [a] (Figure 6.1) exhibit the characteristic features observed in 5.6.1 across the majority of subjects. The most prominent of which include the large mouth cavity ($[a]_{ORAL}$) and a smaller pharyngeal cavity caused by a tongue that is low and pulled back into the pharynx ($[a]_{PHRGL}$). The profile for subject 4 deviates slightly from this observation as there is a marked decrease in cross sectional area just beyond the point of lip coupling with the mouthpiece ($[a]_{MP}$), displacing the measured oral cavity. Additionally, there is no discernable increase in cross sectional area in the pharynx. The constriction situated past the end of the mouthpiece suggests that the subject was incorrectly coupling with the reflectometer by raising his jaw and flaring his lips to form a seal rather than butting his teeth against the outer rim of the mouthpiece.

Subjects 1 and 3 display unusually large cross sectional areas just above the glottis which could be an artefact introduced by averaging over multiple area profiles, some of which may contain elements of acoustic leakage from breathing. This is supported somewhat with coefficients of variation for subject 3 exceeding 50% in this region. However, average variations of <17% for all subjects is within acceptable limits when compared with a clinical physiological technique [7, 17] and in line with previous APR measurements [67].

A significant feature to note throughout all measurements in this study is the shorter vocal tract length of subject 5 in comparison to the others. This should be unsurprising as the subject was the only adult female participant in the group study (see table 6.1) and as described in Chapter 2, the physiology of the adult female vocal tract is on average 2cm (or 15%-20%) shorter than her male counterpart [33, 69] which results in higher formant (resonance) frequencies, accounting for the higher pitched voice in females.

6.3.2 Vowel [e]

The reconstructed profiles for the vowel [e] (Figure 6.2) show the characteristic large cavity at the back of the mouth leading into the pharyngeal cavity at approximately 12-17cm along the reconstructed length, which is created by a central tongue position and raised jaw height. There are however large differences between profiles at a distance of 5cm from the start of reconstruction, which is the junction where the lips and mouthpiece couple. The constriction created in this region from subjects 3 and 4 suggests an improper coupling with the mouthpiece, as previously observed with the [a] vowel. Indeed the range of values shows that all subjects experienced difficulty in coupling the circular mouthpiece whilst trying to retain the widened mouth shape of this vowel. Coefficients of variation in excess of 30% for some subjects in this region are also evidence of this. However, average variations of <15% for all subjects is still well within acceptable limits.

6.3.3 Vowel [i]

The reconstructed profiles for the vowel [i] (Figure 6.3) are distinguished by the constriction between the hard palate and the tongue that is forward and up in the oral cavity ($[i]_{ORAL}$), and the corresponding cavity created behind the constriction entering the pharyngeal area ($[i]_{PHRGL}$). The point of constriction from 5cm to 8cm along the axial distance is similar to that found in 5.6.3, however there is a greater variation in the measured size of constriction between subjects, ranging from 2-7mm in group study versus 3-5mm in single person study. As observed in the single subject experiments, the widened mouth shape of this vowel is the least conducive out of all the vowels measured to couple with a cylindrical mouthpiece.

The coefficients of variation in Figure 6.3 show the distribution of largest variation to be behind the point of constriction and downstream towards the glottis. This is evident for subjects 2 and 5, who both show variations in excess of 50% in these regions.

There is a large range in the average coefficient of variation between subjects (from ~9% to ~41%). However, as will be shown in section 6.5, the vowel [i] experienced the highest percentage of failed readings across the subject group resulting in a smaller number of successful measurements with which to calculate an average area profile. Large deviations between measurements would result in high coefficients of variation.

The high number of failed readings can be accounted for due to occlusion of the airway in the absence of air stream in the area which would be normally constricted. The involuntary tongue movement and effort of the subject to allow linguadental contact but prevent airway occlusion would also account for the large variation in profile between measurements.

6.3.4 Vowel [o]

The reconstructed profiles for the vowel [o] (Figure 6.4) feature a large oral cavity ($[o]_{ORAL}$) across all subjects which is formed by a low back tongue position and raised jaw. As discussed previously in sections 5.6.4 and 5.7, the lip rounding and small mouth shape caused by lip protrusion of this vowel have needed to be exaggerated by the subjects in order to couple with the mouthpiece, which in turn affects the jaw height and relative size of the oral cavity.

The coefficient of variation in oral region measurements for each subject (<25%) are in line with previous experiments, however pharyngeal reconstruction is poor with only two out of the five subjects showing the expected increase in area.

As will be shown in section 6.5, the percentage of failed readings for this vowel is high, which in turn leads to a small number of readings to calculate mean profiles. Therefore the average coefficient of variation for subjects 1 and 3 of <11% should be treated with caution.

The absence of pharyngeal area in the reconstructed profile of subject 5 is further evidence to support the previous assertion of airway occlusion during breath hold in a region that would normally be constricted during phonation. This is presented as being the primary reason for the high number of failed readings for this vowel across all subjects.

6.3.5 Vowel [u]

The reconstructed profiles for the vowel [u] (Figure 6.5) feature the most significant amount of lip rounding and protrusion out of all the vowels studied and as noted in 5.6.5, this caused difficulties in the subjects' coupling with the cylindrical mouthpiece.

The vocal tract shape for this vowel should be characterised by a small rise in area at ($[u]_{ORAL}$) which is created by the tongue that is high and back in the mouth cavity, giving rise to a large pharyngeal area ($[u]_{PHRGL}$), but this profile is only featured in subject 5. The other four subjects exhibit a pronounced constriction at the point of coupling with the mouthpiece ($[u]_{MP}$) 5cm from the first point of reconstruction, where the lips have had to flare to accommodate the size of the mouthpiece. A variation across all subjects' measurements of up to 25% can be seen at this point.

Average coefficients of variation of <16% for all subjects is surprisingly low but this must be taken in context with a relatively high proportion of failed readings (36%).

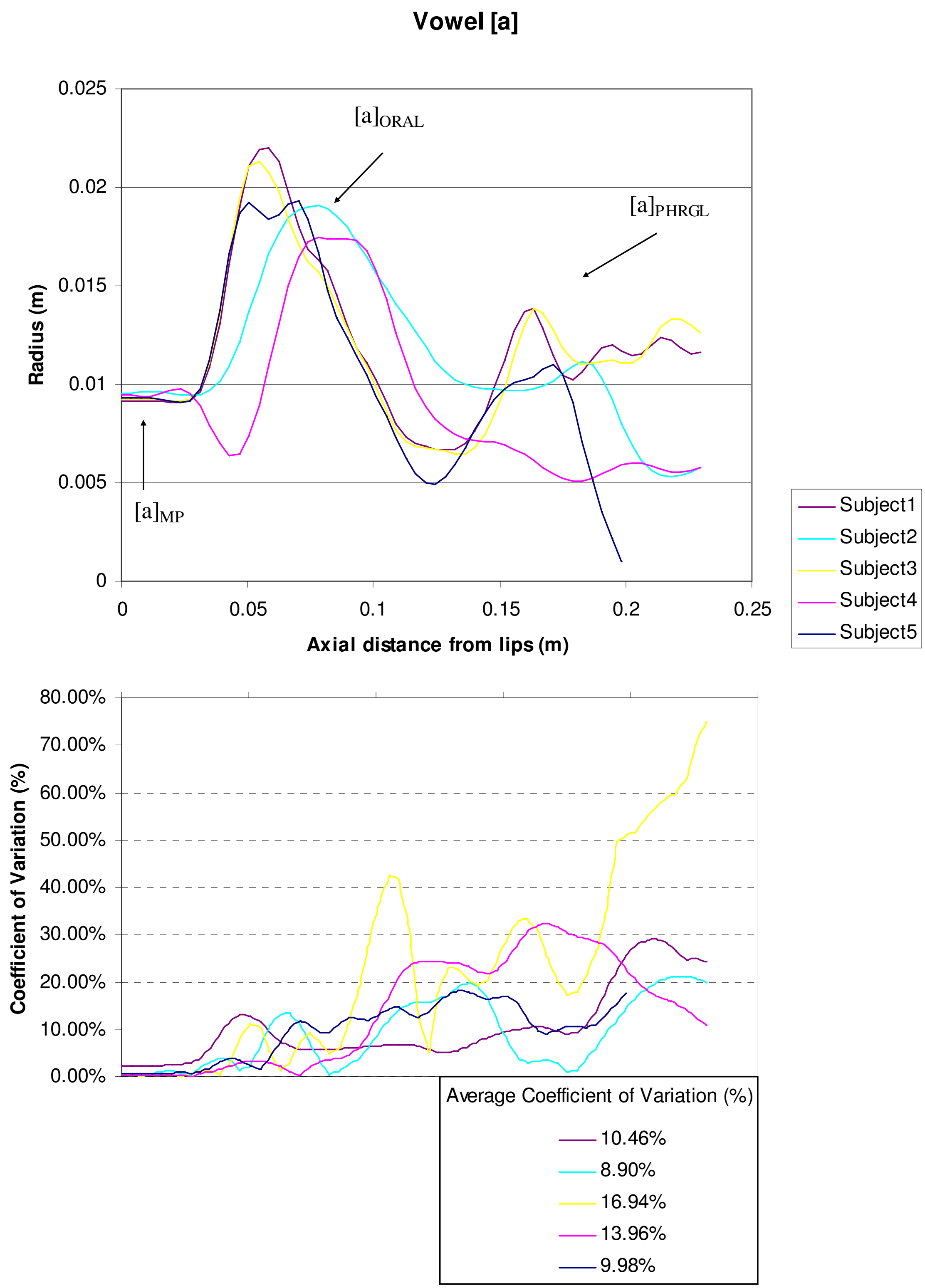


Figure 6.1: Vocal tract area reconstructions from a group study for vowel [a]

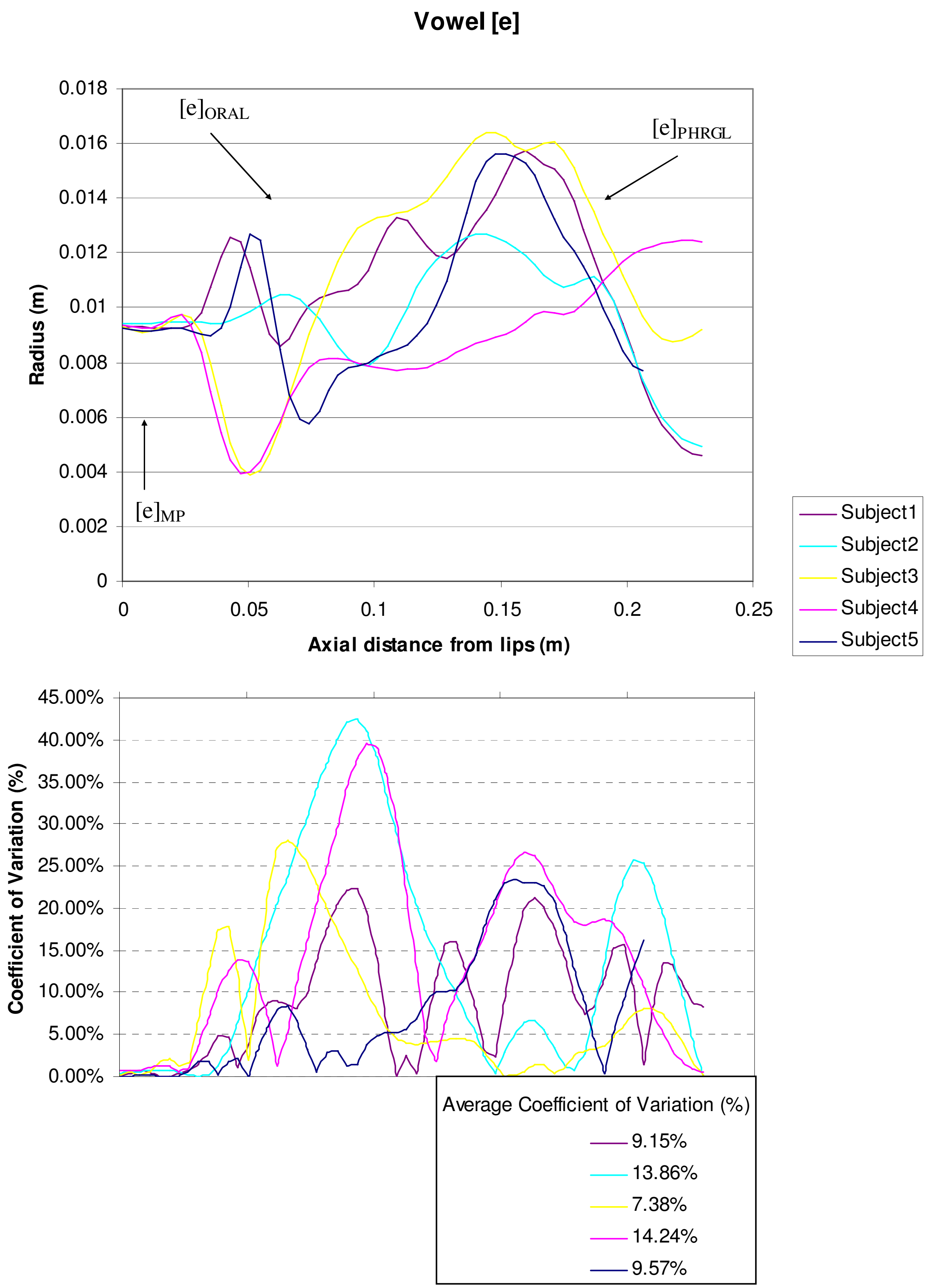


Figure 6.2: Vocal tract area reconstructions from a group study for vowel [e]

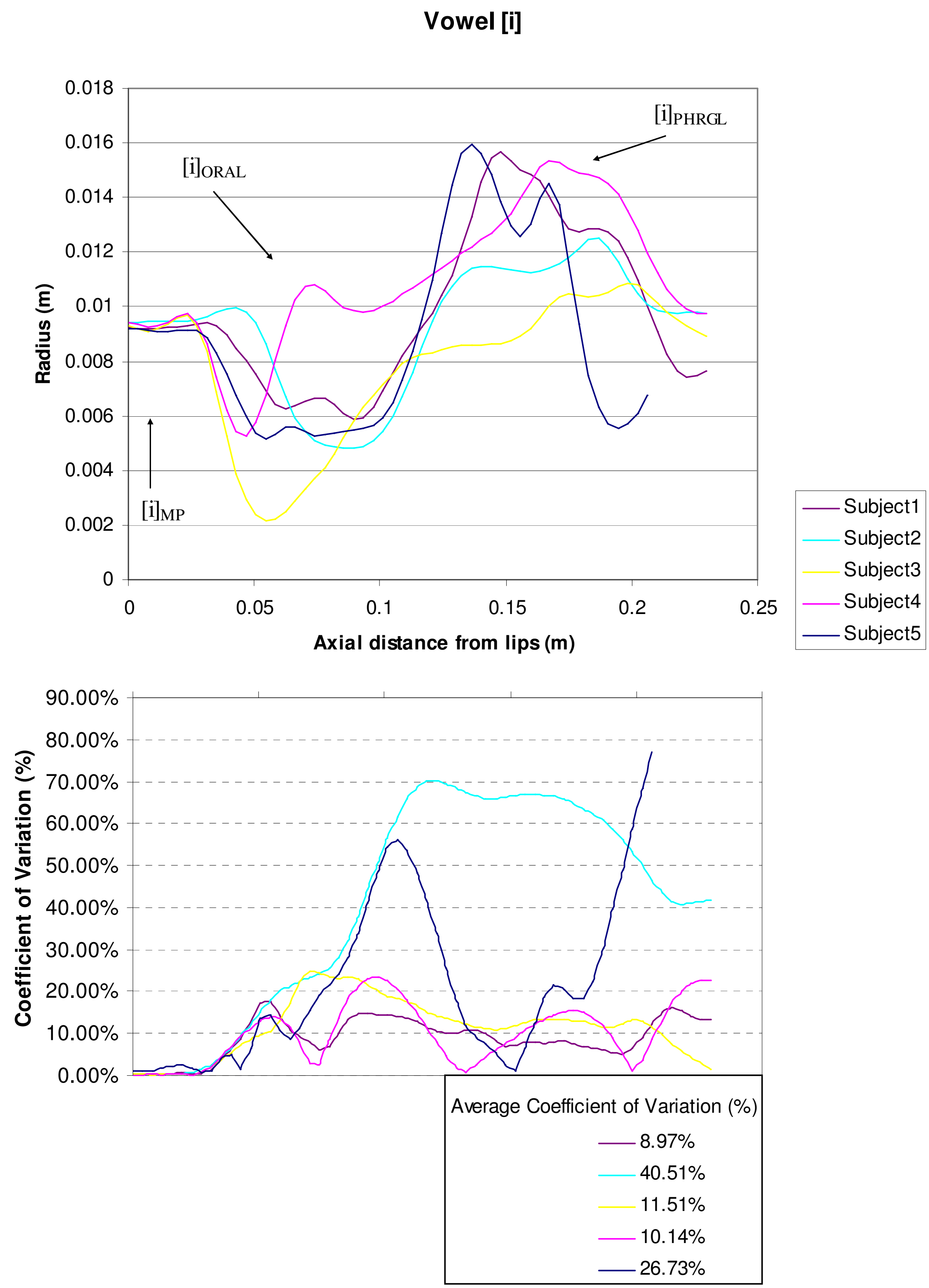


Figure 6.3: Vocal tract area reconstructions from a group study for vowel [i]

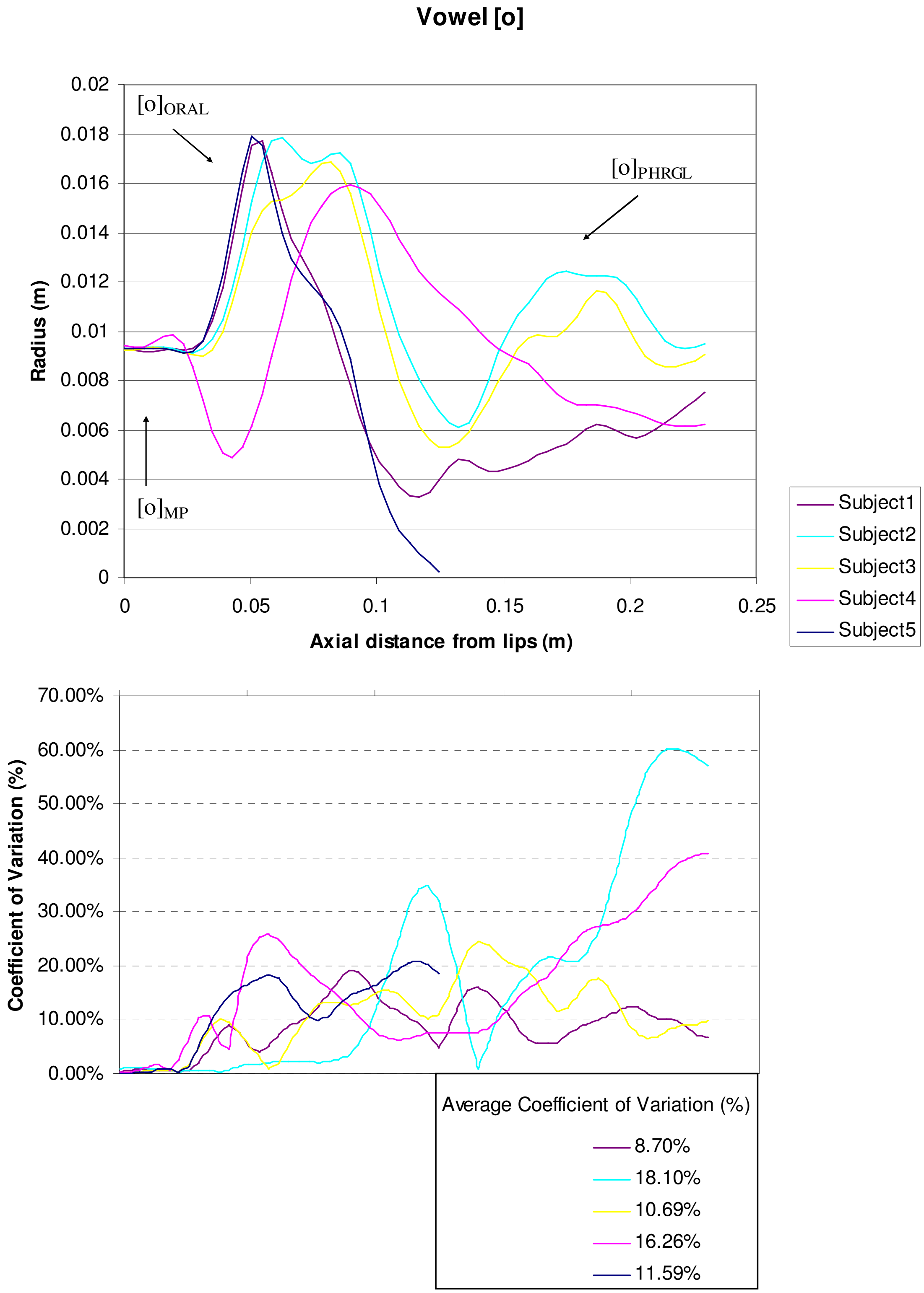


Figure 6.4: Vocal tract area reconstructions from a group study for vowel [o]

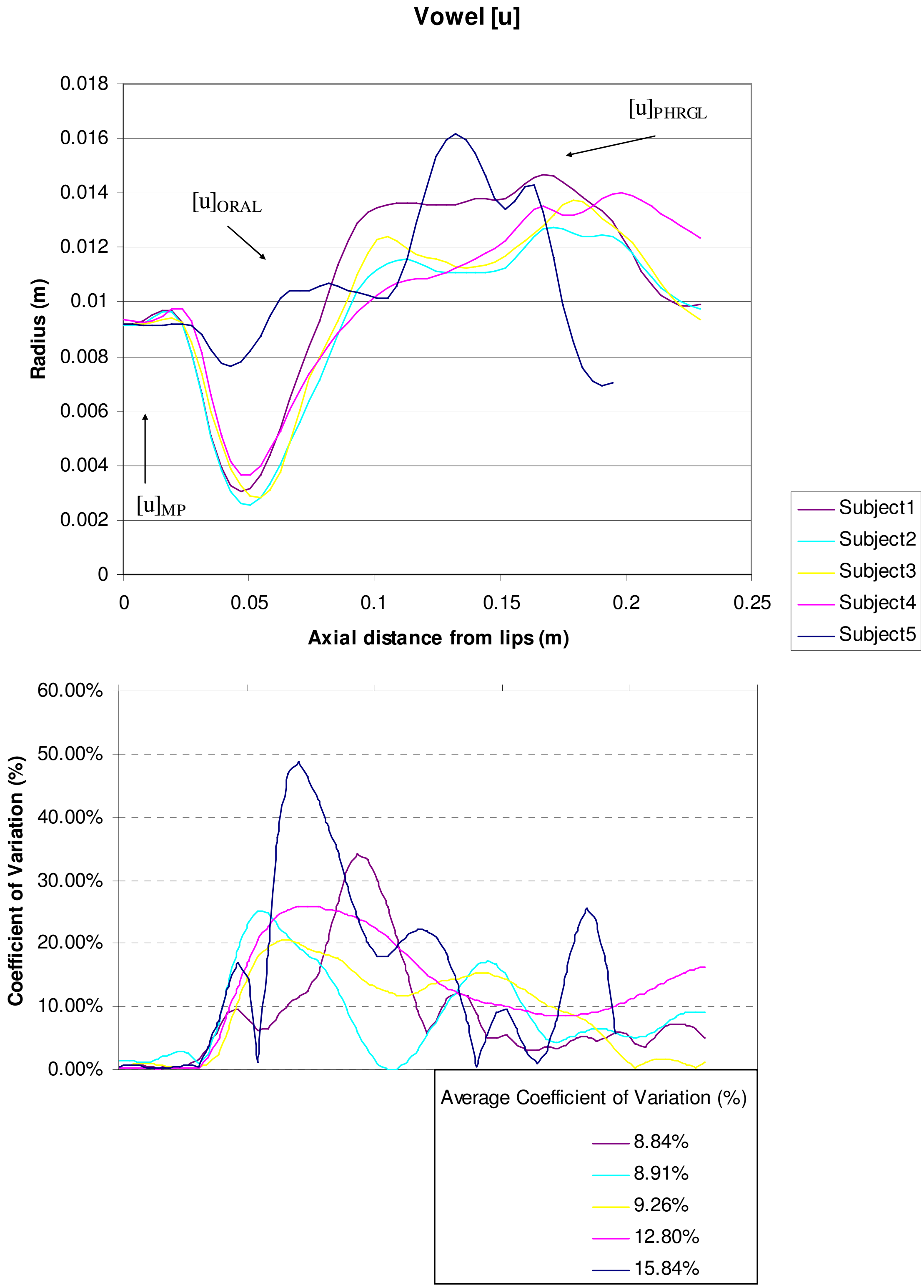


Figure 6.5: Vocal tract area reconstructions from a group study for vowel [u]

6.4 Results from APR measurement of the vocal tract: Child Study

APR measurements from the adult group study showed observable differences in vocal tract length between male and female volunteers. During the course of this trial there was the opportunity to use APR for the vocal tract measurement of two female children aged 7 and 11 years old, to ascertain the potential of APR in measuring much smaller vocal tracts. The reflectometer was constructed with a view to performing vocal tract measurements with adult subjects and was therefore not optimised for use with children. Figure 6.6 shows images of an adult male, adult female and child volunteer coupling with the standard mouthpiece during measurement phase, however it is clear that this would have to be scaled down for future measurements involving children in order to accommodate their smaller mouth size.

Figure 6.6: Mouthpiece coupling showing adult and child subjects

Only two measurements per vowel for each child subject were able to be taken, therefore a qualitative comparison with the adult group study cannot be made. Figure 6.7 shows example measurements from each child subject next to that of the adult female volunteer from the group study. Despite being given instructions to concentrate on forming a seal whilst trying to pretend to imitate the vowel sounds, the child subjects had great difficulty coupling with the mouthpiece. As a result, too much emphasis should not be placed on the detail of the vocal tract profiles. However, information can still be gained regarding the lengths of each subject's vocal tract.

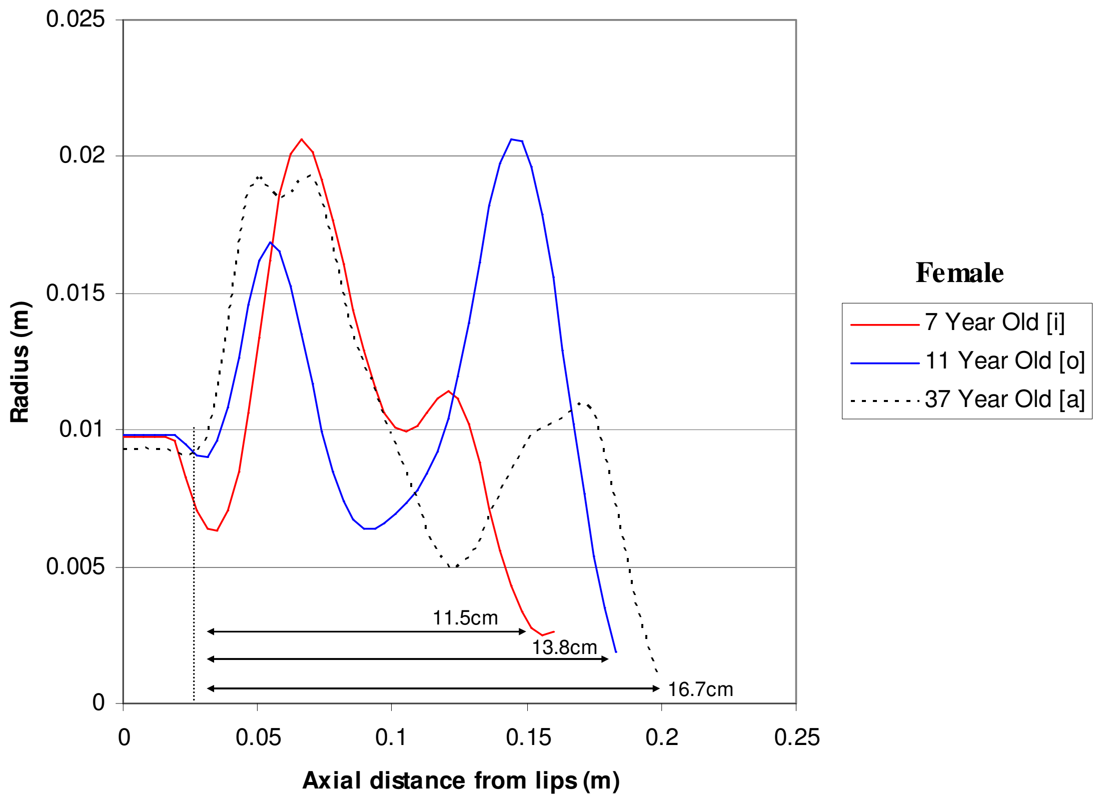


Figure 6.7: Example measurements from child and adult subjects

The length of the vocal tract was measured from the bore reconstruction from the end of the mouthpiece to the point of greatest constriction in the likely area of the glottis. The 7 year old subject displayed an average vocal tract length of 11.2cm ±0.4cm, with the 11 year old measuring 13.4cm ±0.6cm. These results are in line with the generally accepted standards for their age groups [34, 69, 70, 79]. Nevertheless, the number of failed readings was extremely high in both cases (80% and 60% respectively).

6.5 Analysis of results

As in Chapter 5, the repeatability of the technique was assessed through repeated measurements of each vowel, from which a mean area profile and coefficients of variation were calculated. With only a few exceptions, the average coefficient of variation was generally less than 20% which is within acceptable limits when compared with a clinical physiological technique [7, 17]. However, the incidence of failed readings across the group study was significantly higher than that of the single subject study. A graph comparing the percentage of failed readings between the single subject study (using MLS=17) and group study is shown in Figure 6.8. There is however a trend between studies showing higher fail rates with the vowels [i], [o] and [u], all of which feature varying degrees of constriction within the vocal tract and were highlighted as being difficult to couple with the cylindrical mouthpiece. This data supports the belief that persons familiar with the technique and experience of the measurement procedure are more likely to provide successful measurements. In addition, preliminary evidence from the group study suggests that voice professionals (i.e. those with voice training: singers, linguists etc.) show a greater incidence of successful measurements using the APR technique than those with no formal voice training, although the difference is marginal and a larger scale study would be required for more conclusive results. It can be argued however that voice professionals are more aware of, and therefore more capable of controlling their articulators and breathing, which are skills fundamental to this technique.

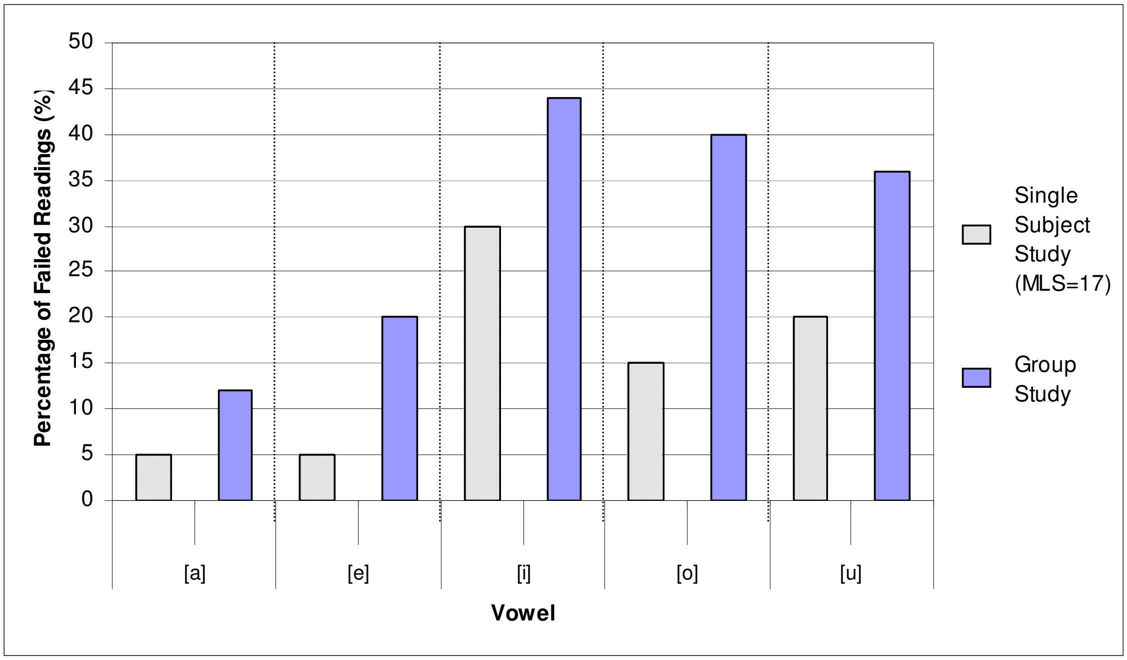


Figure 6.8: Percentage of failed readings for each vowel between single subject and group study

As the study involved several individuals each of whom had a different size and length of vocal tract, there could be no absolute point of reference to align and compare profiles and differences in measurement between subjects. Therefore the coefficient of variation as a function of axial distance was computed between successful measurements for each subject. This allowed an inter-subject comparison of regions of largest deviation in bore profile.

As stated in the previous chapter, the variation between measurements should not always be taken as absolute. Each sample point corresponds to a ~4mm slice at a fixed position from the end of the mouthpiece. The current configuration of APR using MLS excitation signals is a time-averaged technique and is therefore sensitive to changes that occur during the measurement phase. The unnatural length of time a subject is required to imitate the phonation of vowels raises the possibility that the articulators will drift from the target due to fatigue and involuntary muscle movement, which in turn offsets their axial positioning leading to sizeable deviations in the bore between measurements. With the limited resolution of the acoustic technique tending to broaden features, it is the profile (shape of the bore) that is at present the more useful factor in determining the success of each APR measurement.

In addition, the required process to imitate rather than phonate a vowel sound necessitates the absence of an air stream from the pressurised lungs, the force of which would have otherwise supported the vocal tract cavity particularly in cases of localised constriction, a feature of some of these vowels. In some cases occlusion of the airway occurs and the reflectometer is unable to measure past the point of occlusion, most often the pharyngeal area. A prime example of this can be found in Figure 6.4 which shows subject 5 to have had an occluded airway in all measurements for the vowel [o]. It is believed that airway occlusion was the predominant reason for the large number of 'failed readings' within this group study.

6.6 Discussion

APR measurements of the vocal tract are influenced by factors related to subject posture, breathing, inclination and positioning of the source tube, leaks and distortion at the mouthpiece and ambient noise [80]. The objective of the study was to apply the APR technique for vocal tract measurement in a group study and to establish its repeatability under controlled conditions. By using volunteers who were unfamiliar with the APR technique, a simple set of procedural instructions were issued to each subject. This formed part of the APR measurement protocol which was established to ensure consistency across measurements and allow for inter-subject comparison.

The results from this study have shown the APR technique to be successful in distinguishing measurements of cross sectional area and vocal tract length between adult subjects, but to have a more limited success when applied to child subjects.

Through the use of a group study, enhancements to the scan protocol for the standardisation of APR measurements of the vocal tract have been made. Guidelines for quality control and optimal application of APR include:

- Following a standard operating procedure
- Control of environmental conditions (temperature and noise)
- Attention should be given to the mouthpiece and the coupling between the APR equipment and the mouth to obtain correct position, and sufficient seal without disturbing the anatomy

- Regular calibration procedures and maintenance to address hygiene, environmental and safety standards

In the following chapter we will compare results of vocal tract measurement using APR against the gold standard technique of magnetic resonance imaging (MRI).

Chapter 7

Comparative study between APR and MRI

7.1 Introduction

The previous chapters have presented several studies using the APR technique for vocal tract measurement, which have shown to be repeatable under controlled conditions and in line with those of other similar studies. However, a true validation of measurements can only be realised by comparing results with an accepted gold standard technique. This chapter describes the design, process and results from a peer reviewed study entitled "*Comparative study between acoustic pulse reflectometry (APR) and magnetic resonance imaging (MRI) for volumetric measurement of the vocal tract during vowel production*".

7.2 Study design

The purpose of this study is to compare the results of vocal tract measurements from one volunteer using data from both MRI and APR scans. The study was assisted by a female volunteer, aged 28, who was both a professional singer and linguist. A voice professional was chosen to be the subject as the study required the individual to be able to hold their breath comfortably for short periods of time and have good control over voice, articulators and breathing during both APR and MRI scanning procedures. The study was spread over four days and divided into two parts:

(a) APR measurements - performed in an acoustics lab in James Clerk Maxwell Building, The King's Buildings, Mayfield Road, Edinburgh.

(b) MRI measurements – performed at SHEFC Brain Imaging Research Centre for Scotland, Western General Hospital, Edinburgh.

Due to the load being asked of the volunteer and the financial cost of access to such expensive equipment, the comparative study between APR and MRI employed the use of 3 of the 5 vowels previously studied, shown in table 7.1. These three vowels are categorised as corner vowels since they represent extreme placements of the tongue, forming the corners of a triangle in articulatory space, as shown in the F_1/F_2 Vowel Chart (Figure 5.1). All other vowels assume shapes somewhere between these extremes.

Vowel	Tongue Placement
[a]	low and back
[i]	high and front
[u]	high and back

Table 7.1: List of vowels together with tongue placement studied using APR and MRI

The APR data set consists of 60 measurements in total and is divided between measurements with the subject in the normal upright position and the subject lying in a supine position to replicate the conditions of the MRI study. For the MRI study, the subject could only be scanned in a supine position. The opportunity was taken however to measure the subject with the mouthpiece, to replicate the conditions of the APR study, and without the mouthpiece, to analyse the effect of the mouthpiece on the positioning of articulators. Each MRI data set comprises up to 36 image slices covering the entire vocal tract, with each midsagittal scan consisting of a single image. Over 600 MR images were collected throughout the course of the study. A summary of the collected data for measurement of the vocal tract using APR and MRI is presented in Table 7.2[5].

Vowel	MRI		APR	
	With Mouthpiece	Without Mouthpiece	Upright	Supine
[a]	4 image sets + 2 midsagittal	2 image sets + 1 midsagittal	10	10
[i]	4 image sets + 2 midsagittal	2 image sets + 1 midsagittal	10	10
[u]	4 image sets + 2 midsagittal	2 image sets + 1 midsagittal	10	10

Table 7.2: Summary of collected data for APR and MRI study

[5] The summary excludes calibration, test measurement and localiser scans.

7.3 Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging (MRI) is a non-invasive method of obtaining images of internal bodily tissue such as the brain and the spinal cord through the use of powerful magnets and radio waves. The origin of MRI as a medical imaging technique stems from research in the field of nuclear magnetic resonance. Originally called Nuclear Magnetic Resonance Imaging (NMRI), the word 'nuclear' was dropped in clinical applications to avoid negative connotations with nuclear energy, and to prevent patients from associating the examination with radiation exposure.

Medical MRI relies on the relaxation properties of excited hydrogen nuclei in water. Other elements can be used for MR imaging (i.e. those which have an odd number of protons in the nucleus), but hydrogen is used as it is the most abundant element in the body and it has the largest 'Gyromagnetic Ratio' at 42.57 MHz/Tesla.

When a person or object to be imaged is placed in a powerful, uniform magnetic field, the spins of the atomic nuclei with non-zero odd number spin numbers within the tissue all align with the magnetic field. This is done in one of two opposite directions: parallel to the magnetic field or anti-parallel.

The hydrogen protons precess or "wobble" due to the magnetic momentum of the atom. They precess at the Larmor frequency which is given by the following equation

$$\omega_0 = \gamma B_0 \tag{7.1}$$

where ω_0 is the angular (Larmor) frequency of a precessing proton (MHz), B_0 is the magnetic field strength (T), and γ is the gyromagnetic ratio (MHz/T) which is a constant unique to the nucleus of each element.

The tissue is then briefly exposed to pulses of electromagnetic energy (RF pulse) in a plane perpendicular to the magnetic field, causing some of the magnetically aligned hydrogen nuclei to assume a temporary non-aligned high-energy state.

In order to selectively image the different voxels (3-D pixels) of the tissue or object, three orthogonal magnetic gradients are applied. The first is the slice selection, which is applied during the RF pulse. Following this is the phase encoding gradient, and finally the frequency encoding gradient, during which the tissue is imaged. The three gradients can be applied in the X, Y, and Z directions of the machine so that the patient is 'sliced' from head to toe, however, the MRI technique is particularly flexible in that various combinations of the gradients can be combined to enable slices to be taken in any orientation.

As the high-energy nuclei relax and realign, they emit energy which is recorded to provide information about their environment. The realignment with the magnetic field is termed 'longitudinal relaxation'. The time required for a certain percentage of the tissue nuclei to realign (given in milliseconds) is termed "Time 1" or T1. This is the basis of T1-weighted imaging.

T2-weighted imaging relies upon local dephasing of spins following the application of the transverse energy pulse. The transverse relaxation time is termed "Time 2" or T2.

In order to create the image, spatial information must be recorded along with the received tissue relaxation information. For this reason, magnetic fields with an intensity gradient are applied in addition to the strong alignment field to allow encoding of the position of the nuclei. A field with the gradient increasing in each of the three dimensional planes is applied in sequence. When received, the signals are recorded in a temporary memory termed K-space. This is the spatial frequency weighting in two or three dimensions of a real space object as sampled by MRI. The information is subsequently inverse Fourier transformed by a computer into real space to obtain the image of the tissue or the object, with a typical resolution of around $1mm^3$ in clinical studies.

A comprehensive review of magnetic resonance and magnetic resonance imaging is outwith the scope of this investigation. Sources of information on this subject can be found in [81-83].

As discussed in Chapter 1, X-ray CT and MRI are considered gold standard imaging techniques and are the only systems capable of producing vocal tract measurements for comparison with APR measurements. CT scanning was ruled out at an early stage as the technique would expose volunteer subjects to ionising radiation. APR estimates of airway areas have previously been compared with MRI measurements [7], and there have been several successful studies using MRI for airway measurement (whilst breathing) [8] and vocal tract measurement (whilst phonating) [77, 78, 84-87]. There has however been no comparisons between APR and MRI measurements of the vocal tract during phonation (or imitating phonation) of vowel sounds.

7.3.1 Procedure for research trials using human subjects

The study required the services of a healthy volunteer to undergo a medical imaging technique for research purposes with no clinical benefit to the volunteer, therefore the project had to undergo a thorough and rigorous review from medical, scientific and ethical panels, before access was granted to MRI facilities. These included:

- **Lothian University Hospitals NHS Trust – Research & Development**
- **Scientific Committee of the SHEFC Brain Imaging Research Centre for Scotland (SBIRCS)**
- **Lothian Research Ethics Committee (LREC)**

In accordance with the guidelines set by the Central Office for Research Ethics Committees (COREC), a range of documents and protocols were required to be produced in addition to the written informed consent of the subject. Some of these are referenced as follows[6]:

- **Patient Information Sheet (Appendix A.1)**

[6] Volunteer personal data has been removed. Data produced during this study has been archived in compliance with the principles set out in the Data Protection Act 1998.

- **Volunteer Consent Form (Appendix A.2)**
- **Study Protocol (Appendix A.3)**
- **General Practitioner Approval Letter (Appendix A.4)**

7.3.2 Methodology

7.3.2.1 Image acquisition

MR imaging was performed on a 1.5T Signa LX scanner (General Electric Healthcare, Slough, UK), using a neurovascular phased array coil designed to give good signal coverage of the head and neck region. The subject lay supine, using a mouthpiece to reproduce the conditions of the APR measurements as far as possible (see Figure 7.1), and also without the mouthpiece to replicate normal phonation.

Figure 7.1: Subject lying supine with plastic mouthpiece for use in MR scanner to reproduce conditions of APR measurements

7.3.2.2 MRI scanning parameters

The MR examination consisted of a 3-dimensional localiser scan, from which T1-weighted gradient-echo images were prescribed. Imaging parameters for the T1 images were: field-of-view (FOV) 260 mm; slice thickness 4 mm; acquisition matrix 256×256; echo time (TE) 4 ms; repetition time (TR) 80 ms; and flip angle 80°. For each vowel sound, a midline sagittal image was acquired, followed by groups of slices approximately perpendicular to the airway. Slice spacing was 8 mm, i.e. there were gaps of 4 mm between grouped slices. 36 images covered the vocal tract from lips to glottis and were acquired in groups of 6 slices, taking approximately 13 seconds per group. During this time, the subject held the articulation in full apnoea (a pause in breathing). Short rest breaks in between each acquisition allowed the subject to breathe, but no repositioning of or by the subject occurred once the localiser scan was performed. All T1 imaging was repeated without the mouthpiece to compare the effect it had on the vocal tract (displacement of lip shape, jaw height etc.). Selected images shown in Figure 7.2 are: (a) localiser midsagittal scan; (b) a coronal section through the oral cavity showing the maxillary sinuses either side of the nasal cavities, and ethmoid sinuses between the eyeballs; (c) an oblique section through the oropharynx, with the location of the roots of the molars identified as dark grey areas; and (d) a transverse section at the tip of the larynx, above the vocal folds.

The images show the air passages and calcified structures (teeth and bones) in black, muscular tissue in grey and fatty tissues and bone marrow in white, indicating large amounts of free hydrogen atoms [88].

A complete set of MR images for each vowel in normal phonation (excluding mouthpiece) is given in Appendix B.

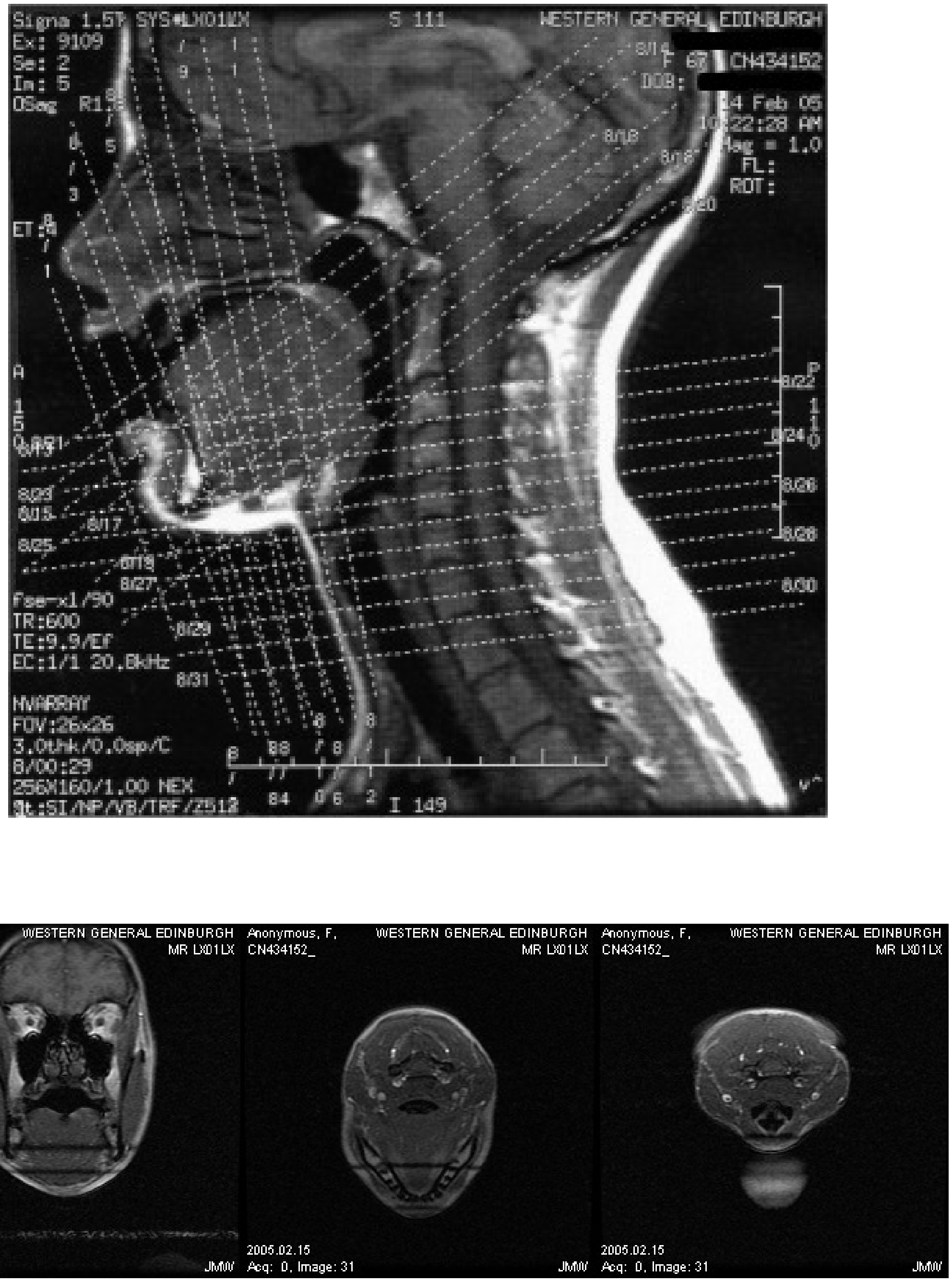


Figure 7.2: (a) Localiser scan (top), together with selected MR images taken perpendicular to the airway axis (located bottom going from left to right) (b) oral cavity, (c) oropharynx, (d) tip of larynx. Air passages and calcified structures appear in black, hydrogen rich parts, such as fat and marrow, in white.

7.3.3 Image analysis

The post processing of MRI data was performed at the Wellcome Trust Clinical Research Facility, Image Analysis Core, at the Western General Hospital, Edinburgh, using Analyze 6.0[7] software on a UNIX platform for the extraction of airway area information. Areas were taken from slices positioned perpendicular to a line drawn midpoint through the vocal tract in the midsagittal image. In practice this meant areas in the mouth were taken directly from the coronal images, whilst for pharyngeal measurements the areas were taken from oblique (tilted) and horizontal images. The overlap of slices between regions ensured full coverage of the vocal tract, from which best positioned slices could be chosen or interpolated. In all images, the pixel size was $1.032 mm^2$.

To align with the reconstructed profiles from APR measurements, the mouth termination of the vocal tract was taken as the first MR image slice preceding the incisors for scans using the mouthpiece. For scans without the mouthpiece, the area profiles were aligned with incisors as before, but the mouth termination was taken as the first MR image slice preceding the lips, allowing for the inclusion of cross sectional areas formed by the lip shape. Other models exist for the calculation of the mouth termination [89, 90].

Areas were determined from "regions of interest" (ROI) which were identified using object maps (see Figure 7.3). The auto detection feature was only partially successful in

[7] © 2002 Mayo Biomedical Imaging Resource: A comprehensive software package developed at the Mayo Clinic for multi-dimensional display, processing, and measurement of multi-modality biomedical images.

segmenting the airway from surrounding tissues, therefore the majority of images relied on the operator segmenting the ROI using the ‘closed trace’ function to position ‘spline points’. Using the ‘flood fill’ tool, the ROI was highlighted before advancing to the next slide to repeat the process. Once a ROI had been selected and highlighted in all slices of the image set, the object map was saved and then using the image analysis and measurement tools, each ROI was propagated to extract area information.

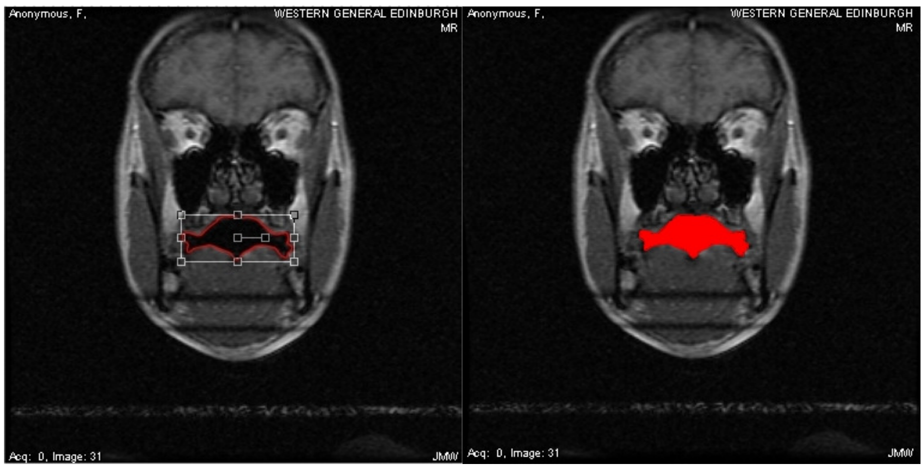


Figure 7.3: Selection and measurement of a ROI using spline points to segment air cavity from surrounding tissue (left), and highlighted using flood fill (right).

Air passages and calcified structures, such as teeth and bones, contain no free hydrogen atoms and therefore do not appear in MR images. Previous studies have devised various procedures to take account of their absence, some of which have included coating the teeth with paraffin wax or pastes containing gadolinium and barium sulphate in order to make the structures visible in MR images, but the results have often proved unsatisfactory [91]. Rather than trying to image the teeth directly during the MR scan, Story et al. [77] used Electron Beam Computed Tomography (EBCT) to generate a cast of the teeth, which was subsequently subtracted from the vocal tract shapes at the image analysis phase. As part of a 3D study of the vocal tract, Engwall and Badin [88] used a similar solution to account for the 'missing structures' by scanning a cast from their subject's dental impression using the same MRI machine as the subject. During image processing, the contours of the casts were positioned to fit the upper and lower incisors in a midsagittal image of the reference position, which depicted a neutral vocal tract shape with upper and lower incisors touching and aligned.

As the use of ionising radiation was ruled out for this study, EBCT could not be used to scan the subject's teeth. Also, to properly align a 3D scan of a dental cast at the post processing stage, 3D scans of the vocal tract would be required, which was not an objective of this study. Therefore, no allowances have been made for the exclusion of teeth in this MRI study.

7.4 Results from APR and MRI measurements of the vocal tract

The following sections present the results from measuring the area profiles of the subject's vocal tract for three corner vowels using APR and MRI. For MRI, the procedure for the measurement and extraction of areas from MR images was described in the previous sections, and the following represents the subject lying supine, with and without a mouthpiece. For APR, the measurement procedure has been described in Chapter 4 and 5, and in addition to the normal upright position, the subject was also measured whilst lying supine to reproduce the conditions of the MRI measurements as far as possible.

7.4.1 Vowel [a]

There is reasonable overall agreement between the APR and MRI area profiles for the vowel [a] (Figure 7.4), exhibiting features previously observed in 5.6.1 and 6.3.1, the most prominent of which is a large mouth cavity ($[a]_{ORAL}$) and a smaller pharyngeal cavity caused by a tongue that is low and pulled back into the pharynx ($[a]_{PHRGL}$).

The hypopharynx in the APR_{supine} profile deviates from this observation and displays an unusually large cross sectional area when compared with the $APR_{upright}$ and MRI profiles. A possible explanation could be the method with which APR_{supine}

measurements were taken. As the subject lay supine, the counterbalanced APR source tube and mouthpiece were self supporting and positioned directly over the subject's mouth. The subject was then able to make fine control positioning movements of the mouthpiece with her hand to couple with her lips. As oral breathing is impossible when connected to the closed source tube, the mouthpiece required to be lifted up and away from the mouth before a breath could be taken. In contrast, $APR_{upright}$ measurements only required the subject to move their head a small distance away or to the side of the mouthpiece in order to take a breath. Also, the $MRI_{mouthpiece}$ measurements used a mouthpiece with an appropriately sized aperture to allow free breathing. In sessions where multiple APR measurements are taken, fatigue can quickly set in, therefore APR_{supine} measurements could have made the subject more susceptible to nasal breathing and also a swallowing reflex rather than constantly repositioning the mouthpiece for an oral breath. Even a small lowering of the velum can cause the acoustic signal to find alternative paths leaking into the nasal cavity, which then appear "added to" the true area. It is considered that these acoustic leakage artefacts averaged over multiple area profiles could account for this measured rise in pharyngeal area.

A notable feature between APR and MRI measurements is the difference in cross sectional area at the lips and posterior to the incisors. The $MRI_{excl./mouthpiece}$ profile exhibits the true cross sectional area created by the lip aperture (radius ~15mm) whereas the other profiles are bounded by the internal mouthpiece opening (radius ~10mm). At an axial distance of ~30mm from the first point of reconstruction, and proceeding posteriorly for ~20mm, the MRI profiles display much larger cross sectional areas than the acoustic measurements. As described in the previous section, calcified structures, such as teeth, appear black in MR images since they contain no free hydrogen atoms, therefore their area is merged into 'airway space' during image analysis. Unless otherwise accounted for, the true vocal tract area (airway) will be overestimated, as in this case.

Although the $APR_{upright}$ area profile provides a better fit to the MRI data than the APR_{supine} profile, the positions of the hypopharynx and vocal folds measured using APR are shifted 10-15mm caudally relative to the MRI data, and are not as well defined. However, the average coefficient of variation across all profiles is <15%, and is in line with previous results.

7.4.2 Vowel [i]

There is good overall agreement between the APR and MRI area profiles for the vowel [i] (Figure 7.5), exhibiting features previously observed in 5.6.3 and 6.3.3. The high degree of lip spreading for this vowel and difficulty in coupling with a cylindrical mouthpiece has been documented in these sections. Evidence of the effect the mouthpiece is having on lip aperture can be found at point ($[i]_{MP}$) showing profiles that utilised the mouthpiece to have a much smaller cross sectional area (radius ~10mm) compared to the true area measured by the MRI scan excluding mouthpiece (radius ~14mm). Also, the MRI profile excluding the mouthpiece shows the constriction between the hard palate and the tongue, formed in the oral cavity, to be closer to the lip aperture than that measured with the mouthpiece, which are displaced posterior by a further ~1cm.

An unusual feature in the series of profiles utilising the mouthpiece is a cavity created behind the point of constriction between 6cm to 9cm along the axial distance from the first point of reconstruction. Throughout this investigation, this feature has only been observed with one other subject (see Figure 6.3 – Subject 1), but it is thought that this is an artefact due to tongue shaping, introduced by the subject's method of coupling with the mouthpiece.

The large cavity created in the pharyngeal region ($[i]_{PHRGL}$) is shown with a similar profile across all measurement groups, only varying in absolute size. On the other hand, there is a much larger difference in cross sectional areas between APR and MRI measurements of hypopharynx and glottal regions, although the overall profiles are very similar.

The acoustically measured position of the vocal folds are again shifted 10-15mm caudally relative to the MRI data and are not as well defined. The average coefficient of variation across all profiles is <20%, and is in line with previous results.

7.4.3 Vowel [u]

There is reasonable overall agreement between the APR and MRI area profiles for the vowel [u] in a supine position (Figure 7.6), exhibiting features previously observed in 5.6.5 and 6.3.5. The rounding and protrusion of the lips associated with this vowel can be seen at the first point of reconstruction for the MRI profile excluding the mouthpiece. The subsequent rise from 8mm to 13mm is exaggerated somewhat by not accounting for the area occupied by the incisors, which are not shown in the MR images.

At the point of lip-mouthpiece coupling ($[u]_{MP}$) for the other profiles, we can see that the mouthpiece has a larger area than the true lip aperture of the vowel [u], forcing the subject to widen her lips to accommodate the mouthpiece. The effect of this is a larger oral cavity ($[u]_{ORAL}$) than compared with the MRI profile excluding the mouthpiece, and is most likely caused by a lowering of the jaw. This difference in area can be seen from 3cm up to 7cm along the axial distance from the first point of reconstruction.

A significant constriction is visible in all profiles between 9cm and 11cm along the axial distance, with radial size ranging from 2-7mm. There are however large measured variations of up to 40% across profiles in this region.

The large pharyngeal cavity behind the point of constriction, created by the tongue that is high and back in the mouth cavity, is of particular note here. When lying supine, both techniques measure similar areas for the pharyngeal region ($[u]_{PHRGL}$). However, the APR measurement of the subject in an upright position is significantly larger. The size of this cavity is still within reasonable limits and suggests that, due to the effects of gravity by lying supine, the tongue and other yielding tissues mass at the back of the airway, reducing the overall volume. This feature is not as prevalent in the vowels [a] and [i] as the tongue assumes a more advanced position in the oral cavity.

As seen in the previous single subject and group studies, there are large variations across the measured profiles for the vowel [u], however, the average coefficient of variation remains within acceptable limits at ~20% across all groups.

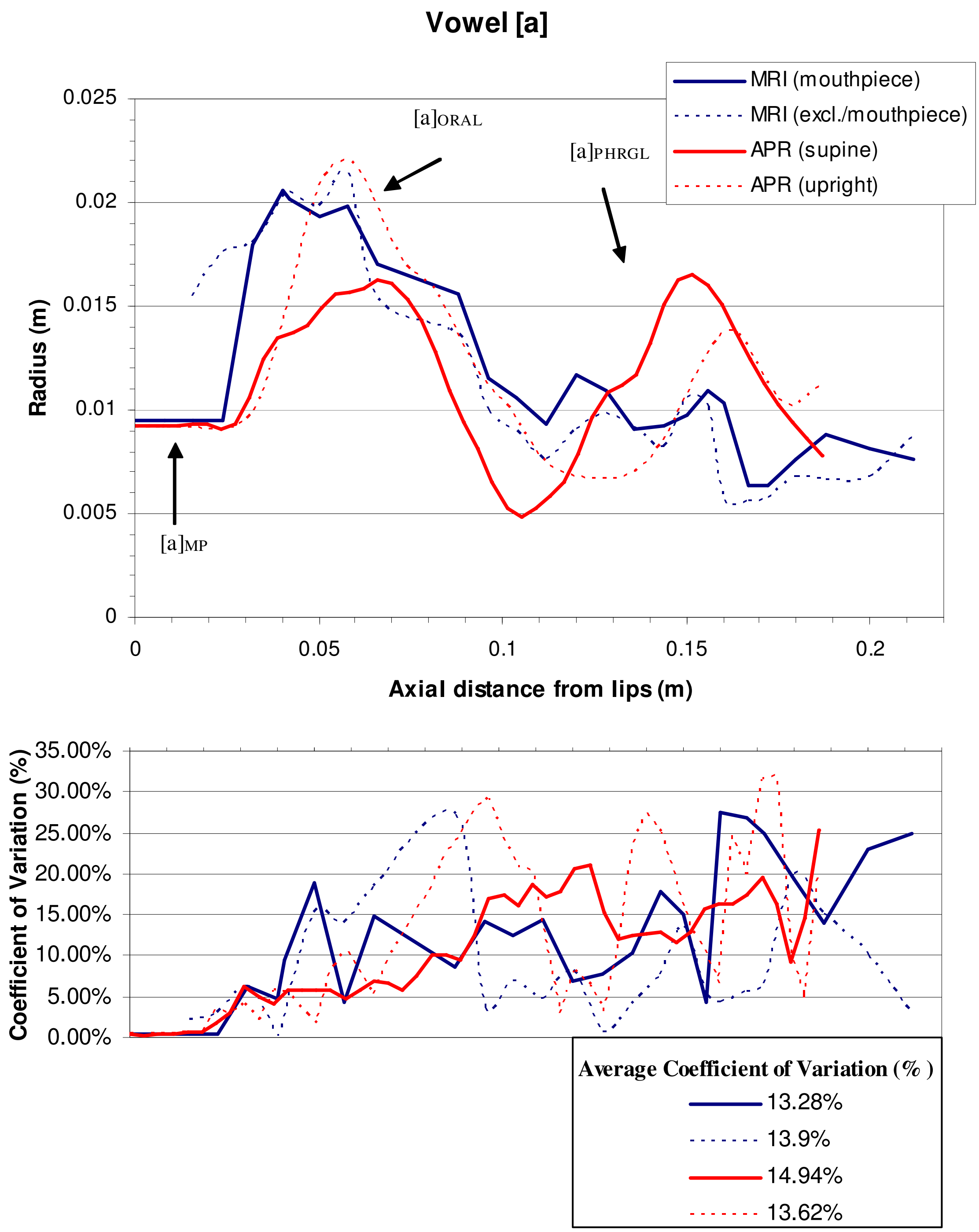


Figure 7.4: Vocal tract area reconstructions from APR and MRI measurements of a single subject for vowel [a], together with measure of variation in data.

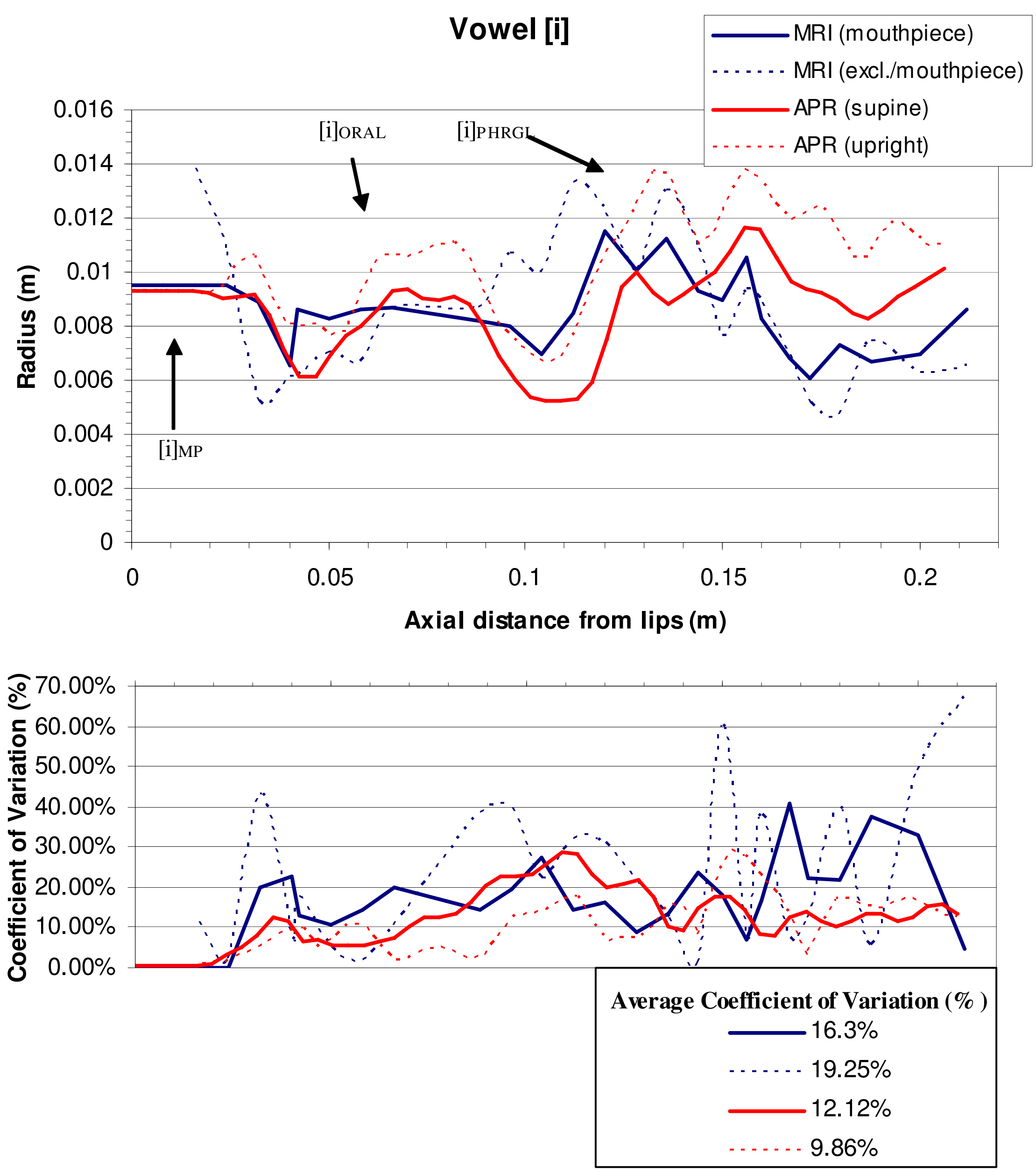


Figure 7.5: Vocal tract area reconstructions from APR and MRI measurements of a single subject for vowel [i], together with measure of variation in data.

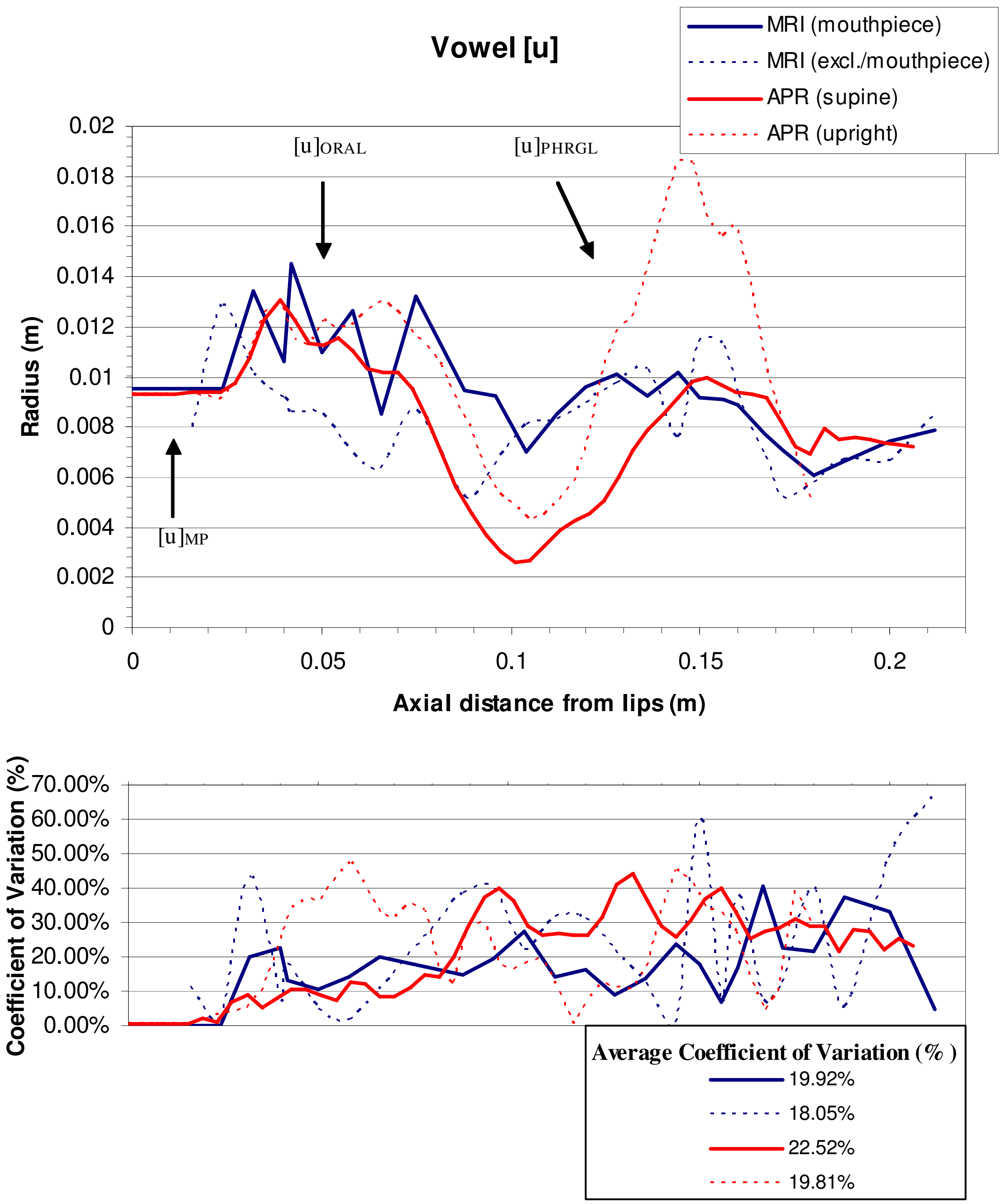


Figure 7.6: Vocal tract area reconstructions from APR and MRI measurements of a single subject for vowel [u], together with measure of variation in data.

7.5 Analysis of results

In a like-for-like comparison between measurements of the vocal tract using APR and MRI (i.e. under the same conditions of lying supine and using a mouthpiece for each vowel), a measure of vocal tract volume was obtained by calculating the cross sectional area of each profile and then integrating over the axial distance. In all cases, the APR technique underestimated the total vocal tract volume compared with those measured by MRI due to the fact that the loss treatment assumes rigid walls and energy is lost to higher order modes. At the same time, the MRI technique overestimates due to the fact that teeth are included in the vocal tract volume.

Volumes for the vowels [a] and [i] agree very well between the two techniques, with APR values within 9% and 3% respectively, of those measured using MRI. The total vocal tract volume using APR for the vowel [u] was within 17% of that measured by MRI. The mean distance from the incisors to the vocal folds using the mouthpiece was 152±12mm by MRI, and 164±11mm with APR. The MRI study without the mouthpiece had a mean length of 167±15mm, measured from the outer edge of lip protrusion to the vocal folds.

7.5.1 Analysis of MR images

Although the software automatically performed a 'best fit' of operator positioned spline points, image analysis was time consuming (taking between 30 minutes to 1 hour per image set) and the resulting area measurements were highly dependent on the thresholds used. Area uncertainties were generally less than ±20%, however there were a number of incidences which exceeded this value, but are in line with those reported by [7] up to ±50%, and [92] who reported coefficients of variation in the range 5-33% for repeated estimates of glottal areas from CT images.

7.5.1.1 Irreducible uncertainty in MR image analysis

Although many would regard MRI as the "Gold Standard" for airway measurement, the problems of determining "Regions Of Interest" (ROI) areas are substantial, even when the process is partially automated (best fit of adjoining spline points). Marshall [7] showed that a lower bound can be set on the area uncertainties by considering the image pixels defining the airway wall (Figure 7.7).

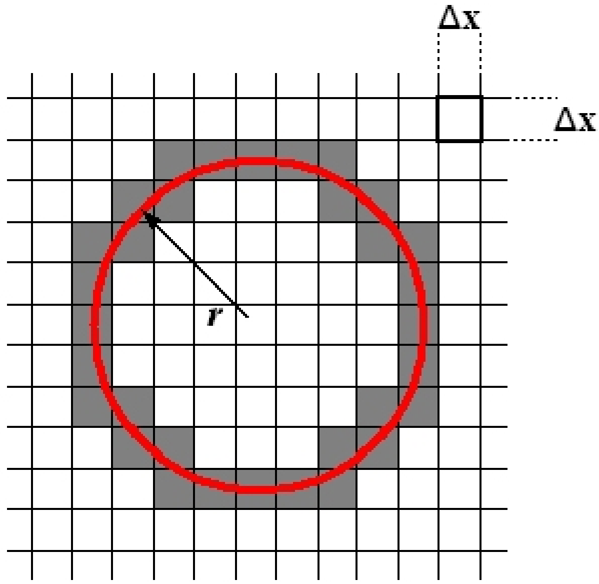


Figure 7.7: The intrinsic uncertainty in area estimation caused by the pixel size

The wall passes through approximately $2\pi r/\Delta x$ perimeter pixels, each of area Δx^2, where r is the radius of the airway and Δx is the length of the pixel side. Since we cannot tell what proportions of these pixels lies inside the airway, the irreducible area uncertainty is:

$$\pm\frac{1}{2}\times\frac{2\pi r}{\Delta x}\times\Delta x^2 = \pm\pi r\Delta x \tag{7.2}$$

compared with an actual area of $A = r^2$, i.e. a proportional uncertainty of $\pm\Delta x / r$ or $\pm\Delta x / \sqrt{A/\pi}$. With a pixel side length of 1.01563mm, this expression evaluates to 18% for A = 1cm^2, 13% for A = 2 cm^2 and 10% for A = 3 cm^2. This uncertainty is inherent in the discrete nature of the image, and would apply even if the contrast between the airway and the surrounding tissue were infinite. Although these errors may be small for large oral cavities, they become more significant when evaluating smaller areas such as those commonly found in the oro- and hypopharynx, and especially the much smaller areas at points of constriction.

7.5.2 Acoustic and articulatory effect of the mouthpiece

Chapter 5 discussed in detail the relationship between the vocal articulators and how they shape the acoustic signal. Whilst the tongue is recognised as the prime mediator of oral and pharyngeal resonance in the production of vowel sounds, lip shape and jaw height also play a role [71, 89]. Lip rounding increases the acoustic length of the vocal tract and therefore shifts the formant frequencies of the vocal tract downward ([39] p.132). On the other hand, lip spreading reduces the acoustic length and therefore increases all formant frequencies of the vocal tract ([34] p.168). Jaw lowering has a similar effect, particularly with regard to F_1 [93].

A current limitation of the APR technique highlighted throughout this work is the need for a mouthpiece to couple the source tube with the lips. The fixed size of the mouthpiece has meant that subjects have had to adjust their lip shape and jaw height for some vowels in order to form an acoustic seal. For the vowels [a] and [i] this has meant lip rounding, and for the vowel [u], lip spreading. Figure 7.8 shows pictures of lip shape and midsagittal MR images for normal phonation compared with midsagittal MR images using a mouthpiece for each vowel. From these images, it is clear that the mouthpiece is impacting on the normal phonatory position of the lips, particularly in the case of vowels [i] and [u], and also altering the cavity posterior to the incisors.

The acoustic effect of the mouthpiece was investigated by recording the subject whilst singing each vowel at a comfortable pitch (A4 – 440Hz), with and without the mouthpiece[8], and then performing a spectral analysis on the recorded audio signal. The acoustic effect of changing the lip shape to accommodate the mouthpiece is shown in Figures 7.9(a-c), which display a section of the frequency spectrum (250Hz – 2.8kHz) for each vowel in normal phonation, compared with that of phonation using the mouthpiece.

[8] To prevent acoustically lengthening the vocal tract or introduce filtering properties of the plastic mouthpiece, the mouthpiece was removed prior to recording.

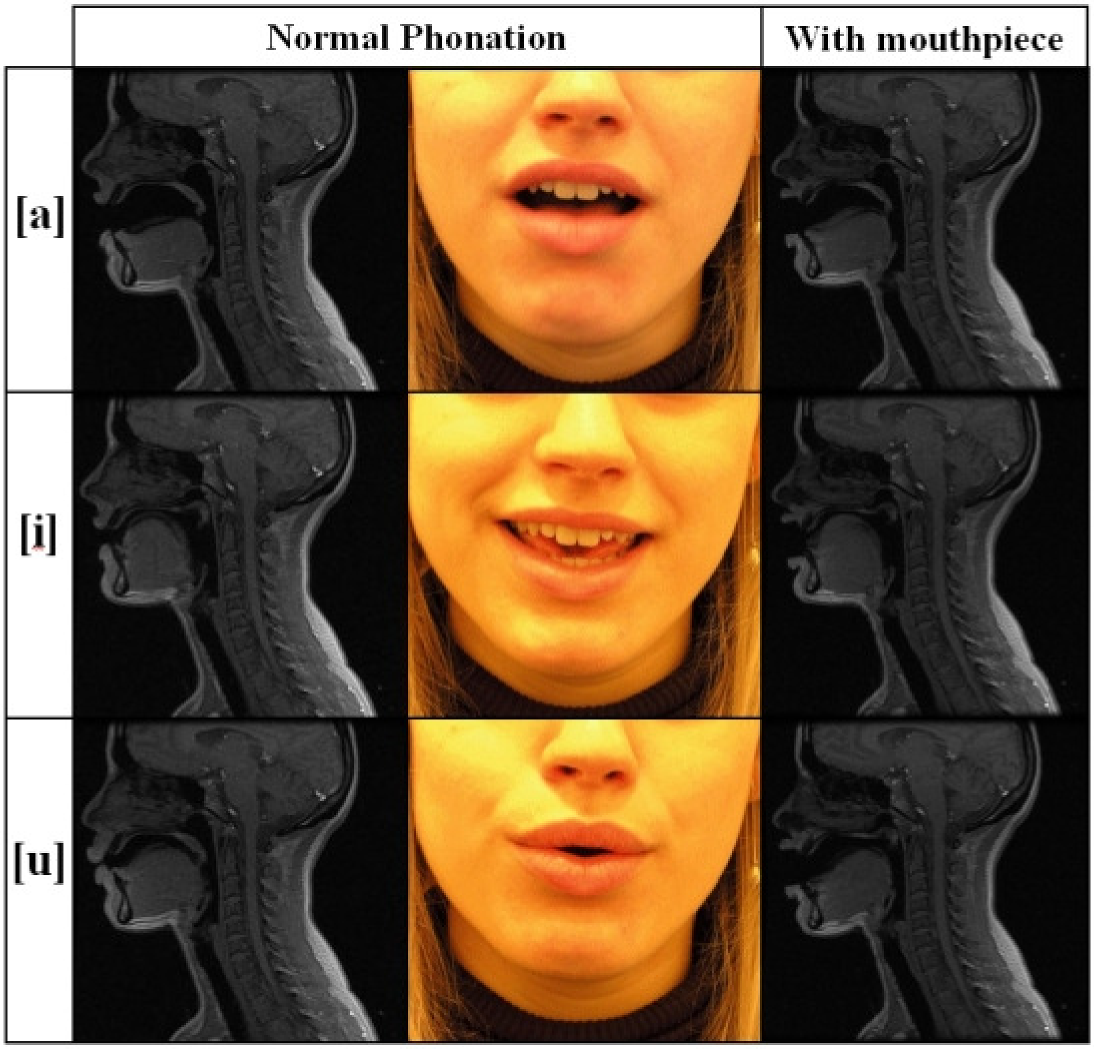


Figure 7.8: Midsagittal MR images together with photo of lip shape for normal phonation (left), and MR images using a mouthpiece (right).

For the vowel [a] (shown in Figure 7.9(a)), vocal tract resonances around 700Hz and 1500Hz (see section 5.4.1) are shown to boost the 2nd and 3rd harmonics when singing normally at A4 pitch. When singing at the same pitch with the mouthpiece in place, the lip shape becomes more rounded thereby increasing the acoustic length of the vocal tract, shifting the resonant frequencies of the vocal tract downward. The result in this case is attenuation of the 3rd and 4th harmonics.

For the vowel [i] (shown in Figure 7.9(b)), vocal tract resonances around 270Hz and 2100Hz (see section 5.4.3) are shown to boost the 5th and 6th harmonics when singing normally at A4 pitch. By using the mouthpiece, the spread lip shape changes to a round lip shape, and as in the previous vowel, has the effect of lowering the resonant frequencies of the vocal tract. The result in this case can be seen in the boost given to the 3rd, 4th and 5th harmonics.

For the vowel [u] (shown in Figure 7.9(c)), vocal tract resonances around 300Hz and 800Hz (see section 5.4.5) boost the 2nd harmonic when singing normally at A4 pitch. By using the mouthpiece, the lip shape is spread slightly, decreasing the acoustic length of the vocal tract and raises its resonant frequencies. This rise in F_2 boosts the 2nd harmonic.

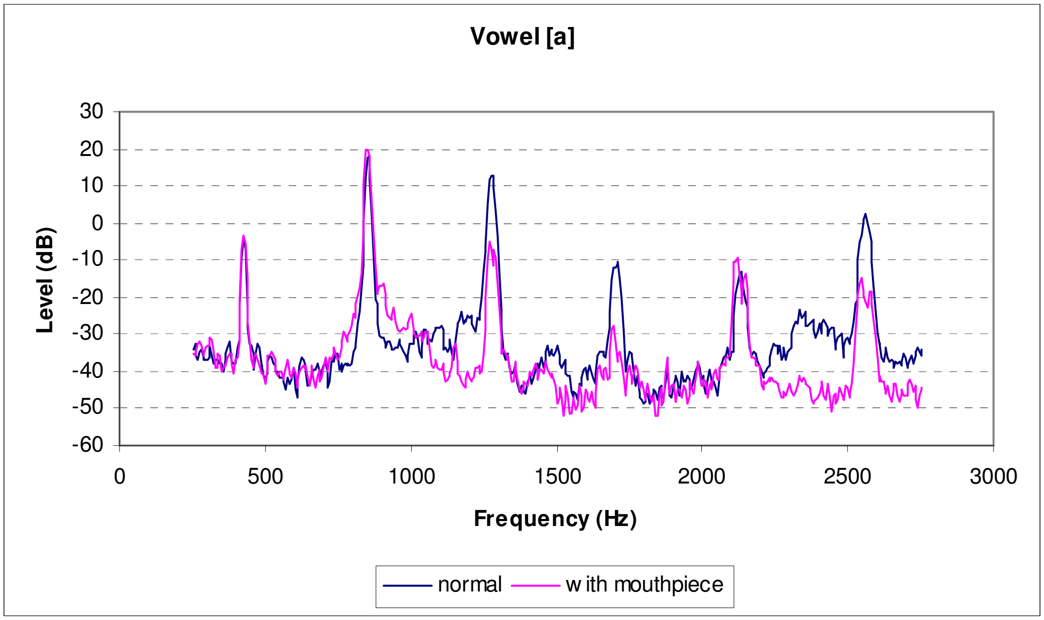


Figure 7.9(a): Frequency spectrum for vowel [a] sung at A4 (440Hz) pitch, with and without a mouthpiece.

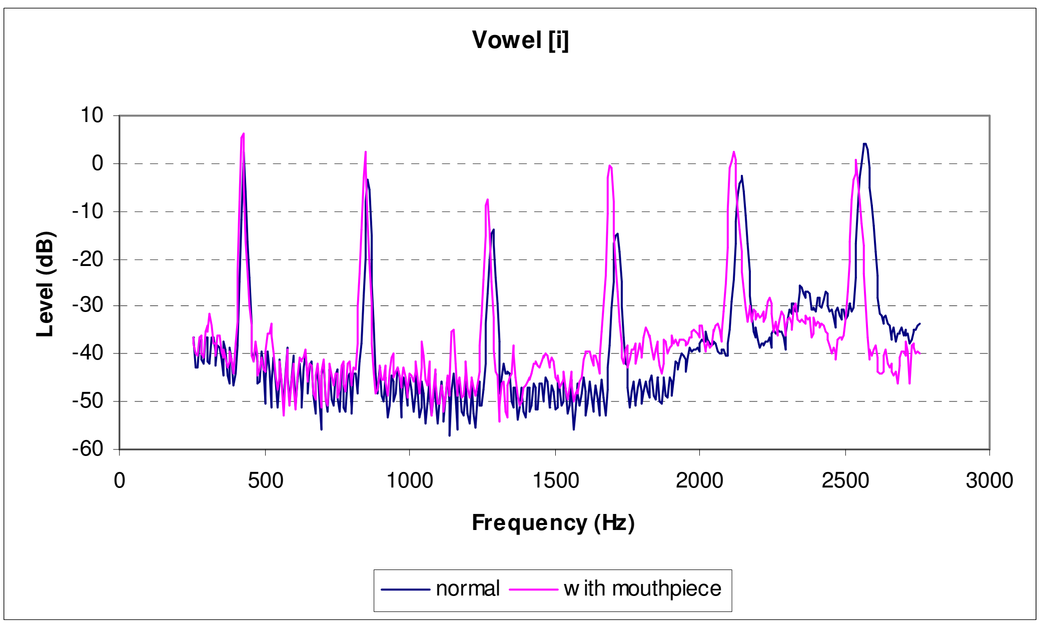


Figure 7.9(b): Frequency spectrum for vowel [i] sung at A4 (440Hz) pitch, with and without a mouthpiece.

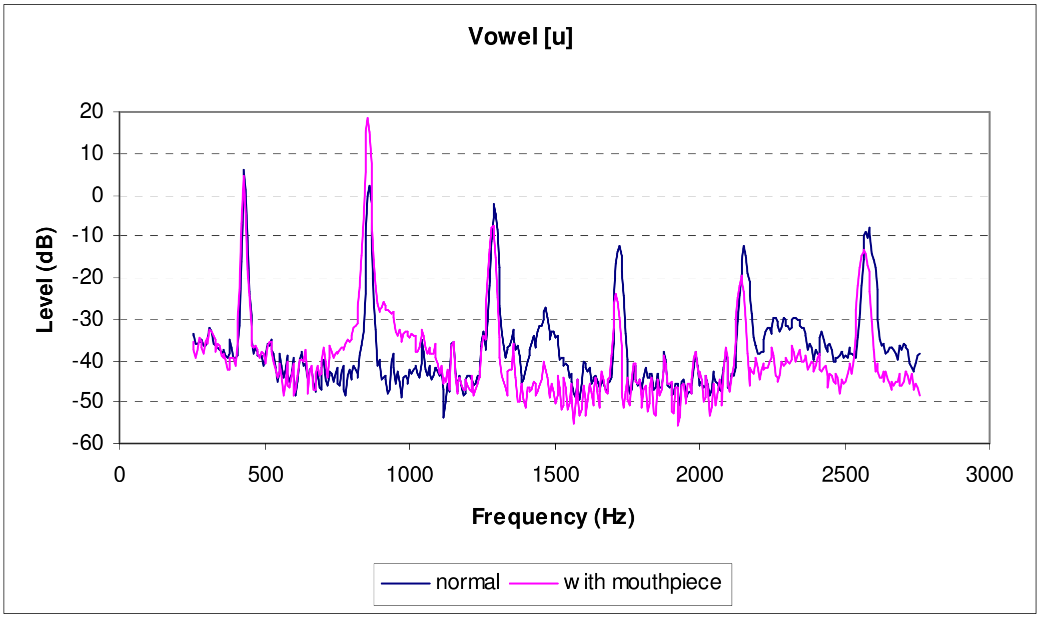


Figure 7.9(c): Frequency spectrum for vowel [u] sung at A4 (440Hz) pitch, with and without a mouthpiece.

7.5.3 Effect of posture

Studies have shown that the control of head position during airway measurements using acoustic reflection is not critical, provided there is no obvious neck flexion [7, 18]. Most APR measurements throughout this study have been on subjects in an upright position in order to recreate the normal the conditions of phonation. The MRI study however, required the subject to lie supine in order to be scanned. The effect of posture on upper airway dimensions has shown the supine position to significantly alter the pharyngeal area [94-98]. The APR results of this study have also shown the subject in a supine position to have a significant reduction in pharyngeal vocal cavity volume when compared with measurements of a subject taken in an upright position. In the absence of airstream (expiration pressure in the case of phonation), gravity makes it more likely for the tongue to fall back over the airway and/or for the airway muscles and other tissues to occlude the airway [99]. The link between supine position and 'collapsibility' of the upper airway [100, 101] are well known in respiratory medicine and have been shown to be a major contributory factor in sleep disorders [95].

7.6 Discussion

The purpose of the study was to compare results from the APR technique for vocal tract measurement against those of MRI, an accepted gold standard technique. This involved using the same subject, a voice professional, in both studies for the measurement of three corner vowels [a], [i] and [u]. The same conditions were reproduced for each technique as far as practically possible, which meant a primary data set measuring the subject lying supine and using a mouthpiece. For comparison, a secondary data set was collected which measured the subject in an upright position (APR) and without the mouthpiece (MRI). The articulatory and acoustic effects of the mouthpiece, and the effect of subject posture on measurements have also been discussed.

The results from this study have shown reasonable overall agreement between the APR and MRI data, with the limited resolution of the acoustic technique tending to broaden features, particularly in the region of the pharynx and glottis. The spatial smoothing of acoustic measurements also has its counterpart in the temporal smoothing of MRI measurements [7]. The imaging time of several minutes, encompassing 6 breath holds and rest periods for one complete scan of the vocal tract (~31 MR images), inevitably means that the data is averaged over all phases of respiration. It is important that the subject remains still for this time, ideally without swallowing. With the acoustic technique, on the other hand, a full vocal tract measurement can be made within a single breath, with little inconvenience to the subject.

Chapter 8

Conclusions and future work

The research aims set out in Chapter 1 have all been achieved. In this chapter, a summary of the conclusions is presented with recommendations for future work.

8.1 Achievement of aims

8.1.1 Aim 1: Development of APR

The first aim of the study was to develop the APR technique as a non-invasive method for measuring the bore of the vocal tract for static targets of articulation.

Chapter 2 introduces vocal tract anatomy and physiology. Chapter 3 discusses the basic theory for plane wave propagation within tubular cavities together with a method for solving the inverse problem for determining the dimensions of an object from its input impulse response. This formed the theoretical foundation from which an Acoustic Pulse Reflectometer could be built. Chapter 4 describes the production of a working reflectometer and its application to the measurement of the vocal tract.

Several developments to the standard APR technique are described in these chapters, one of which is to apply MLS signals rather than using a single pulse as an input signal. By applying a signal which continues over a longer time interval, more energy is able to be injected into the system which improves the signal-to-noise ratio and removes the need to average over a large number of measurements. Experiments on test objects confirmed that higher order MLS excitation signals enabled greater accuracy in subsequent bore reconstructions. It is also observed however that MLS excitation signals are normally applied to linear time invariant (LTI) systems. Its application must be used with care as the vocal tract is most often considered a dynamic system. The assumption made in this study shows that by imitating the phonation of vowel sounds, with a specific place of articulation, and refraining from breathing, the vocal tract can be considered a time invariant system. This is shown in Chapter 4 which describes how recordings were made of individuals to determine the maximum length of time they could remain still before fatigue and involuntary movement set in. By doing this, a maximum length of time for an APR measurement was deduced, which in practice gave the maximum length of an MLS excitation signal that could be applied. Chapter 4 also describes the application of the 'virtual DC tube' method which simulates the effect of the DC tube without the need of an intermediate physical coupling, resulting in a quicker and more reliable measurement.

8.1.2 Aim 2: Group study

The second aim was to apply, and investigate the repeatability, of the APR technique for the measurement of the vocal tract in a group study of volunteers ranging from children to adults.

The optimisation of the APR technique for measurement of the human vocal tract using a single subject is described in Chapter 5. The results from repeated measurements of five non-nasalised vowels using MLS excitation signals in the range $16 \leq m \leq 18$ show good overall agreement when compared with vocal tract profiles from previous studies, but no direct comparisons can be made because of differing methods and subjects used in these experiments.

Measurements using the lowest order MLS input signal (m=16) consistently underestimated the pharyngeal area when compared with those using higher order signals. The application of the highest order MLS signal (m=18) exhibited the greatest incidence of 'failed' readings, therefore a compromise between the need for a quick measurement and the desire for a long time interval to put energy into the system was required. An MLS signal of order m=17 was adopted for all subsequent measurements.

The repeatability of the technique was found through repeated measurements of each vowel, from which a mean area profile and coefficients of variation were calculated. With a few exceptions, the average coefficient of variation was generally less than 20%. Factors pertaining to the success of measurements and the repeatability of the technique have been discussed in detail in Chapter 5 and a measurement protocol was devised to create a benchmark for comparison of results with future studies.

Chapter 6 describes how this protocol was successfully applied to the vocal tract measurement of children and adults. The results show the APR technique to be successful in distinguishing measurements of cross sectional area and vocal tract length between adult subjects, but have a lesser success when applied to child subjects. Enhancements to the scan protocol for the standardisation of APR measurements of the vocal tract were made and methods to facilitate inter subject comparisons were developed.

8.1.3 Aim 3: Comparison of APR and MRI

The third aim was to implement a comparative study between APR and MRI for the measurement of the vocal tract during the (imitation) of vowel sounds.

Full details of the study are presented in Chapter 7.

The services of a voice professional were employed for the vocal tract measurement of three corner vowels [a], [i], and [u] using both APR and MRI techniques. The results from the study show reasonable overall agreement between the APR and MRI data, with the limited resolution of the acoustic technique tending to broaden features, particularly in the region of the pharynx and glottis. The average coefficient of variation between area profiles across all vowel groups is less than 20%, which is in line with that of the previous APR single subject and group studies.

In a like-for-like comparison between the APR and MRI techniques for measurement of the vocal tract, measurements were conducted with the volunteer lying supine and using a mouthpiece for each vowel. A measure of vocal tract volume was obtained by calculating the cross sectional area of each profile and then integrating over the axial distance. In all cases, the APR technique underestimated the total vocal tract volume compared with those measured by MRI. Volumes for the vowels [a] and [i] agree very well between the two techniques, with APR values within 9% and 3% respectively, of those measured using MRI. The total vocal tract volume using APR for the vowel [u] is within 17% of that measured by MRI. The mean distance from the incisors to the vocal folds using the mouthpiece is 152±12mm by MRI, and 164±11mm with APR. The MRI study without the mouthpiece has a mean length of 167±15mm, measured from the outer edge of lip protrusion to the vocal folds.

The second part of this aim was to produce a framework for future APR and MRI investigations, which includes all supporting documentation to satisfy scientific, clinical & ethical bodies for the use of human volunteers in research trials.

The documents and protocols produced in this study are in accordance with guidelines set by the Central Office for Research Ethics Committees (COREC) and have been approved by the appropriate scientific, clinical and ethical bodies.

Their inclusion in this work should assist future researchers to facilitate what can often be a lengthy and protracted process.

8.1.4 Aim 4: Vocal tract data

The fourth aim was to produce a corpus of data of vocal tract measurements from APR and MRI investigations, which can be used as ‘real’ geometric data for application in physical models of the voice.

Results from over 400 APR vocal tract measurements have been presented as area profiles in Chapters 5 and 6. The profiles represent the shape of the vocal tract (cross-sectional area as a function of distance) for the vowels [a], [e], [i], [o] and [u], and were measured from a variety of adult male and female volunteers (including a voice professional), and also female children.

For the comparative study between APR and MRI, 60 APR measurements were taken of a female voice professional in upright and supine positions. In addition, over 600 MR images of the same individual have been collected whilst imitating the phonation of the corner vowels [a], [i] and [u] (split between full vocal tract data sets and single slice midsagittal images). The extracted area profiles from these measurements are shown in Chapter 7. Selected MRI data sets, together with midsagittal ‘profile’ images for each vowel during normal phonation (excluding mouthpiece), are shown in Appendix B.

8.2 Limitations

No technique is without methodological problems, ranging from the invasiveness of the procedure to the limitation on temporal and spatial resolution. APR for vocal tract measurement is a technique that has been shown to provide reasonable accuracy in the measurement of cross sectional area of the upper airway. However, it is a technique that is not exempt from aforementioned limitations.

Assumptions being made in the present work are plane wave propagation and rigid airway walls. As discussed in chapter 3, if the bandwidth of the input signal is greater than the lowest cut-off frequency of the cavity (which may vary along its length), a mixture of plane waves and higher order modes will propagate. In this case, the assumption of plane wave propagation upon which the layer peeling algorithm is based becomes invalid, the result of which is a lowering of the accuracy of the bore reconstruction with underprediction of the cavity cross-section, as energy is lost to higher order modes. Also, the rigidity of the airway wall is frequency dependent and it is only above approximately 1 kHz that airway non-rigidity could be neglected [16]. Further investigation is required in both these areas.

A practical limitation of the APR technique highlighted throughout this work is the need for a mouthpiece to couple the source tube with the lips. The fixed size of the mouthpiece has meant that subjects have had to adjust their lip shape and jaw height for some vowels in order to form an acoustic seal. For the vowels [a] and [i] this has meant lip rounding, and for the vowel [u], lip spreading. The mouthpiece has been shown to impact on the normal phonatory position of the lips, particularly in the case of vowels [i] and [u], and also altering the cavity posterior to the incisors. The acoustic and articulatory effects of the mouthpiece are examined in detail in section 7.5.2.

The measurements from the child group showed an extremely high number of failed readings. Accepting that the technique was not optimised to accommodate a child's much smaller mouth shape, the current technique requires a certain level of cooperation from volunteers, which includes control of breathing and the ability to remain still throughout the measurement phase, both of which are difficult to obtain with young children. The present format of the technique therefore has limitations in its application to young children. However, the technique was well tolerated by adult volunteers, although shorter measurement times would be beneficial.

8.3 Future research

Real-time imaging or a sufficiently fast capture rate with which to observe dynamic changes (such as articulation and swallowing) is the ultimate goal of vocal tract and airway imaging studies. In practice however, no technique currently exists that has a rapid response and satisfies the requirements of being non-invasive, portable, inexpensive and simple to operate.

The primary focus of this research has been on the development of the APR for measuring the vocal tract bore for vowel sounds. It has been noted previously however that determining the internal dimensions of objects of varying cross-section is a problem common to many branches of science and industry. Therefore the development of the APR technique is of interest to a number of disciplines that require a non-invasive method of measurement. For example, acoustic rhinometry (acoustic reflectometry of the nose) is increasingly used for the objective clinical assessment of nasal patency. The benefits of this technique as a method which can be applied easily and quickly, in addition to being portable, are well documented [51]. APR measurements of the nasal and oral airways are however influenced by factors related to subject posture, breathing, inclination and positioning of the source tube, leaks and distortion at the nosepiece/mouthpiece, and ambient noise. Recommendations for future work to overcome these and other limitations are discussed in this section.

In the present study, volunteers were issued with the same set of verbal and written instructions to ensure consistency across measurement procedures and allow for inter-subject comparisons to be made. However, no standardised measurement protocols currently exist for measuring the vocal tract using the APR technique. Such a procedure would be useful when comparing results between different reflectometers and set a benchmark for future clinical studies.

The MLS excitation signals employed throughout this work have proved suitable in the limited context of studying a static vocal tract. However, the required length of signal is at the fringes of how long a human subject can remain immobile for the vocal tract to still be considered a linear time invariant system. The current method requires two sequences of the MLS signal to be played end to end to satisfy the periodicity condition of the technique, but only the second half of the signal is used. Whilst of little consequence when applied to the measurement of static objects, a faster method which would ideally use the whole signal is required for the vocal tract. Further work is required to see if such an adaptation of the MLS method is possible. Alternatively, new input signals should be investigated.

A practical development to the current system would be to adapt the mouthpiece or introduce a face mask method to alleviate the problem of fixed jaw height. This would allow a better measurement of the vocal tract and lip shape when imitating the phonation of some vowels.

The imitation rather than phonation of vowels is not ideal as the articulators can drift out of their target position during the long measurement time. The subject has no auditory cue with which to make fine adjustments to articulators in order to achieve an articulatory target, therefore methods to enable phonation and breathing whilst coupled with the source tube should be investigated.

A study of airway wall characteristics with regards to frequency dependent attenuation factors (losses) is required. Their inclusion in the bore reconstruction algorithm may have the potential to improve the accuracy of future vocal tract measurements. In addition, further work is necessary to investigate what role, if any, higher order modes play in the APR method for vocal tract measurement.

Finally, another potential area for future research involves the multiple microphone technique [102]. By being able to decrease the length of the source tube, high frequency losses within the source tube would be less significant and increase the bandwidth of the measurement. This may prove useful when applied to the measurement of the vocal tract only if higher order modes can be taken into account in the bore reconstruction algorithm.

Appendix A

A.1 Patient Information Sheet

A.2 Volunteer Consent Form

A.3 Study Protocol

A.4 General Practitioner Approval Letter

ACOUSTICS & FLUIDS RESEARCH GROUP
SCHOOL OF PHYSICS – COLLEGE OF SCIENCE & ENGINEERING
THE UNIVERSITY OF EDINBURGH

Mr Calum Gray
Rm.4201 - James Clerk Maxwell Building
The King's Buildings
Mayfield Road
Edinburgh EH9 3JZ

Tel: 0131 650 5257
Fax: 0131 650 5902
Email: cgray@ph.ed.ac.uk
Web Site: http://www.ph.ed.ac.uk/fluids/

PATIENT INFORMATION SHEET

COMPARATIVE STUDY BETWEEN ACOUSTIC PULSE REFLECTOMETRY (APR) & MAGNETIC RESONANCE IMAGING (MRI) FOR VOLUMETRIC MEASUREMENT OF THE VOCAL TRACT DURING VOWEL PRODUCTION

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

LREC Ref: 04/S1101/22

Version: 1.2 (06/09/04)

Contents

1. What is the Purpose of the Study?

The measurement of human airway dimensions has been a frequent objective in the field of medical and speech research, but vocal tract imaging is inherently difficult due to its important biomechanical function (breathing, swallowing, speaking etc.).

The Chief Investigator for this project is a researcher in the field of Acoustics (the study of sound) and is involved in developing a computer simulation of the voice.

A major part of this research is to obtain ‘real data’ of the voice source – with the size/shape of the vocal tract playing a primary role in the acoustic output (what we hear).

One of the main methods used in measuring the vocal tract is *Magnetic Resonance Imaging (MRI).* This technique uses a combination of powerful magnets and radio waves to create very high quality pictures of particular parts of the body.

There are however disadvantages with using MRI for vocal tract measurement. Some of these include:

- **Time** - It takes time to perform a complete scan using MRI - during which the subject must remain immobile. Scan time is typically longer than what a human can comfortably hold their breath for.
- **Expense** - MRI is an expensive procedure and access to this equipment is limited
- **Image Analysis** – Once the MRI scan is complete, ‘post processing’ of the images is still required so that the data can be presented in a useable form to the researcher/clinician.

As an alternative to MRI, the Chief Investigator has continued the development of a technique known as Acoustic Pulse Reflectometry (APR) and has adapted it for measuring the vocal tract.

APR uses audible sound waves from a loudspeaker (directed down a tube with a subject placing their lips against a mouthpiece) to calculate the vocal tract shape/volume.

The advantages associated with APR include:

- **Non-Invasive** – The subject places their mouth against the sterile mouthpiece and is asked to refrain from moving & breathing for up to 10 seconds.
- **Repeatable** – The quick measurement time means the subject can perform several repeats with minimal fatigue
- **Inexpensive** – The equipment to construct an APR is readily available and inexpensive

The purpose of this study is to compare the results of vocal tract measurements from one volunteer using data from both MRI and APR scans.

2. What is required of the volunteer?

The main requirements for a volunteer to participate in this study are:

- Adult (over 18 years of age)
- To be available for both parts of the study (*see The Study & Schedule*)
- Able to hold breath comfortably for short periods of time (10 to 15 seconds)
- Good control over voice, articulators (lips & tongue) & breathing

NOTE: *Ideally, we would like a voice professional – singer, linguist, phonetician, public speaker, voice technologist, musician (brass/woodwind player) etc.*

- Able to lie still on your back whilst pretending to 'sing' vowel sounds with a mouthpiece against your lips/teeth
- **Able to have MRI scan**

NOTE: ***There are a number of criteria for participation in MRI studies. As the MRI machine uses strong magnets, you may be unsuitable for this scan if you have any of these conditions:***

- Cardiac pacemaker
- Artificial heart valve
- Vascular clips
- Cochlear implant
- Shunt
- Shrapnel or metal injury in the past
- Head operation
- Dentures, dental plate/bridge or brace
- Hearing aid
- Joint replacements
- Harrington rods
- Ever had metal in or around the eye
- Epilepsy
- Diabetes
- Metal plates, screws or pins
- Bone or joint pins
- Prosthesis (non-removable)
- Artificial limbs (non-removable)
- Metal screws or pins
- Insulin pump
- Seizures
- Claustrophobia
- Heart disease or rhythm disorder
- Recent surgery of any type (within the last six months)

Ladies

- Pregnant
- Breast-feeding
- IUCD or sterilisation clips

MRI scans will be performed at SHEFC Brain Imaging Research Centre for Scotland which is part of Clinical Neurosciences, School of Molecular and Clinical Medicine at the Western General Hospital. The staff are very experienced in imaging healthy volunteers as well as sick people. Their clinical decision as to a volunteer's suitability for MRI scans will be final.

Due to the number of scans required and therefore the corresponding times the volunteer will be asked to hold their breath for short periods, this study could be quite tiring (mainly affecting jaw and facial muscles when pretending to 'sing' vowel sounds – such as '*ooh*' and '*ah*'). Therefore the study has been designed to include rest periods in between scans and will span several days to prevent volunteer fatigue.

3. The Study & Schedule

This study is divided into two parts:

(a) **APR measurements** - To be performed in an acoustics lab in James Clerk Maxwell Building, The King's Buildings, Mayfield Road, Edinburgh.

(b) **MRI measurements** – To be performed at SHEFC Brain Imaging Research Centre for Scotland, Western General Hospital, Edinburgh.

One volunteer is required to participate in **both** parts of the study.

It is expected that the volunteer will be required for up to 8 hours in total which will be spread over four days (not necessarily consecutive).

	Scan Type	Location	Duration	Description
Day 1	APR	Acoustics Lab, James Clerk Maxwell Building, The King's Buildings	2 hours	• Meet with Chief Investigator who will explain the project & answer any questions. • Practice with mouthpiece & holding vocal tract steady for short periods. • APR scanning will commence.
Day 2	MRI	Centre for Brain Imaging Research, Western General Hospital	2 hours	• Accompanied by Chief Investigator, meet with staff of Brain Imaging Research Centre. • MRI scanning will commence.
Day 3	MRI	Centre for Brain Imaging Research, Western General Hospital	2 hours	• MRI scanning.
Day 4	APR	Acoustics Lab, James Clerk Maxwell Building, The King's Buildings	2 hours	• This time has been allocated to allow any further APR scanning and to organise any final administration details.

4. Part A: APR Measurements

4.1. What is APR?

An Acoustic Pulse Reflectometer consists of a computer connected to a loudspeaker (similar to that used in a home Hi-Fi system) which is coupled to a long plastic tube with a microphone.

The computer creates a sound from the loudspeaker which travels down the plastic tube and out the end into the open air.

A small microphone is placed inside the plastic tube and records the sound back into the computer. By blocking off the end of the pipe or by attaching another object, we change properties of the sound we hear from the loudspeaker. At the same time, these changes are recorded by the microphone back into the computer. We can then use this recording to calculate the properties of the object placed at the end of the tube (for example, its shape, size and volume).

We have successfully used this technique for many years in measuring the internal shape of musical instruments (such as trumpets, trombones and clarinets), but we are also developing this technique to measure the human airway (mouth and vocal tract). This introduces new challenges as at present we can only measure things that remain still – unfortunately for our research, the human mouth and vocal tract are very seldom inactive as they perform roles in breathing, speaking and swallowing!

Having said this, there are times where our mouth and vocal tract remain relatively still. One of these times is when we 'sing' vowel sounds, and it is from this we will start our research for measuring the vocal tract.

4.2. Is APR safe?

APR has been used with humans over many years in both clinical and research fields. It is deemed a very safe technique as it only uses audio signals created by a loudspeaker (exactly the same way as music is played on a home Hi-Fi system) and the loudness of the signal is kept to within a comfortable listening level.

4.3. What will I be asked to do for APR measurements?

You will be asked to lie on a bed with the mouthpiece of the reflectometer positioned next to your mouth. You will then be asked to make a vowel sound (such as '*ooh*' – as in the word '*ooze*'). Once you have performed this vowel for a few seconds you will be asked to freeze your mouth in that position and place your mouth to the mouthpiece of the reflectometer where you will hold your breath for 10-15 seconds.

During this time a sound is sent from the computer to the loudspeaker. As you put your mouth on the mouthpiece (attached to the plastic tube), the sound of the noise will be altered by the shape you have made with your mouth for that particular vowel.

It is this sound that the microphone records back into the computer and with which we can use to work out the shape of the mouth and vocal tract for that particular vowel.

4.4. How many measurements will be needed?

As it takes a relatively short time to obtain a measurement with a reflectometer, it would be preferable to take several measurements for each vowel.

It is intended that between 3 and 5 vowels will be measured with a minimum of 5 measurements for each vowel (one measurement takes 10-15 seconds).

As previously mentioned, it can be quite tiring holding your mouth & vocal tract in these positions for periods of time, so we will aim to minimise the amount of discomfort by taking breaks in between measurements and allow you to drink water.

4.5. What does the mouthpiece look like?

The mouthpiece is little more than a circular plastic tube to put your lips onto.

Here is a picture of one of our mouthpieces:

As we may use several volunteers throughout the course of our research with the reflectometer, the mouthpiece allows us to hygienically ‘connect’ volunteers to the plastic pipe of the reflectometer. Volunteers will be given an individual mouthpiece sleeve (sterilised cardboard tubes from peak flow meters) to slip on and cover the plastic mouthpiece (which is cleaned after each use with anti-septic wipes).

5. Part B: MRI Measurements

This part of the study will be carried out at:

SHEFC Brain Imaging Research Centre for Scotland
Clinical Neurosciences
SCHOOL OF MOLECULAR AND CLINICAL MEDICINE
The University of Edinburgh
Western General Hospital
Crewe Road
Edinburgh EH4 2XU

Centre Coordinator: Mr Lindsay Murray
Tel: 0131 537 2664
Email: lgm@skull.dcn.ed.ac.uk
Fax: 0131 537 2661
Web Site: http://www.dcn.ed.ac.uk/bic/

5.1. Is MRI Safe?

MRI does not use X-rays, and no drugs or injections will be involved.

If you agree to join the study, **we will check that it is perfectly safe for you to be scanned**. Although MRI is normally a very safe method of taking pictures, scans are not taken from people who have a heart pacemaker or who have had surgery involving the insertion of metal clips into the brain, or people who have metal fragments in their eyes, perhaps as a result of their job. Neither will we scan you if there is a chance that you might be pregnant. On the other hand, the metals used in operations such as hip replacements are very rarely a reason not to undergo scanning. Please also see the list of exclusions (Page 3 of this Patient Information Sheet)

The Radiographers will check if you are in any doubt.

5.2. What will I be asked to do for MRI measurements?

When you come to the Centre for your scan, a changing cubicle will be provided. You will be asked to place any metal objects, such as keys, watches, coins and credit cards, in a locker. Please do not wear any make-up or talc, and be prepared to remove contact lenses if you use them. You may be asked to wear a wrap-around gown while you are in the scanner.

You will be asked to lie on the scanner bed for up to one and a half hours, but not continuous, for each of the two days.

The format will be very similar to the APR measurements. While you are in the scanner you will have the mouthpiece in your mouth and will be asked to make a vowel sound for a few seconds. You will then be asked to freeze your mouth in that position, holding your breath for 10-15 seconds whilst a series of pictures are taken of your mouth and vocal tract.

It is envisaged that this part of the study will be more tiring for you than in the APR measurements as you are being asked to hold the mouthpiece in your mouth for longer periods of time. It is necessary to remain still as there is less opportunity to alter position whilst in the scanner.

A shorter version of the mouthpiece will be used to fit inside the scanner. It will also have extra side supports to rest comfortably against your lips.

You will be supplied with this version of mouthpiece at the first meeting with the Chief Investigator and will be encouraged to practice with it (both on site and at home).

MRI scanners are only able to take one picture at a time of a small region of your mouth and vocal tract and therefore approximately 30 scans will be required to cover this area. As we are investigating a minimum of 3 vowels, this could total approximately 90 scans (at 10-15 seconds for each scan).

5.3. What is it like being in an MRI scanner?

The scanner makes quite loud noises while it operates. For your comfort, you will be provided with ear plugs or ear defenders.

The design of an MRI scanner can make a person feel as if they are in an enclosed space. As this is a study for research purposes and not a clinical scan, we want the volunteer to feel as comfortable and relaxed as possible. For this reason, we have listed claustrophobia as part of our exclusion criteria for this study.

If at any stage, including shortly before or during your scan you become worried or wish to ask a question, you will be able to speak to one of the Radiographers who will use an intercom to keep in touch with you.

Please note that we are required to send a routine report of the scan to your Doctor.

6. Confidentiality

All information from the APR and MRI scans are entirely confidential, in the same way as are all other medical records. Pictures gathered from both scans will be stored and processed using computers and, after the study is completed, will be copied onto a permanent record which might be studied again at a later time. Information gathered during your scan will be shared with other medical and scientific researchers. In these cases, people cannot be identified from their scans, and there are strict laws that will safeguard your privacy at every stage.

We would like to take a photographic record of this study and you will be asked for your permission before doing so. All photographs will be of the procedure only and not the volunteer, therefore any photos that would include the volunteer will be altered to protect their identity.

7. Who is organising & funding this research?

This research is being organised by Mr Calum Gray (Chief Investigator for this study) who is a Postgraduate Researcher funded by the Engineering Physical Sciences Research Council (EPSRC) at the School of Physics, University of Edinburgh.

Funding for this study is from the School of Physics, University of Edinburgh.

8. Who has reviewed the study?

This study has been reviewed by the following groups and committees:

- School of Physics, University of Edinburgh
- Lothian Research Ethics Committee (LREC)
- SHEFC Brain Imaging Research Centre for Scotland (SBIRCS) Scientific Committee
- Lothian University Hospitals NHS Trust – Research & Development

9. What will happen to the results of this Study?

This research will form part of the Chief Investigator's PhD research and results will be published in his thesis, scientific journals and conference proceedings.

The work is expected to be published in mid 2005.

The Chief Investigator will notify volunteers from this study when results are published and if requested, will be provided a copy of this work.

10. Withdrawing from the Study

You do not have to take part in this study, and **you may withdraw from it at any time**. We are very grateful to you for offering to help us with this research.

11. Further Information & Contacts

Further information on this study and acoustic pulse reflectometry (APR) is available from Mr Calum Gray (Chief Investigator for this study). The contact details are on the first page of this Patient Information Sheet.

Further information on magnetic resonance imaging (MRI) is available, if you require it, from Dr. Colin Turnbull, the Patient Services Director for Radiology in the Lothian University Hospitals NHS Trust (Tel: 0131 537 2042). Dr Turnbull is not involved in this study, and so will be able to give you independent advice.

Otherwise, the Brain Imaging Research Centre's Facility Co-ordinator or one of the Radiographers will be happy to try to answer any other questions that you might have. They can be contacted at the address shown on Page 8 of this Patient Information Sheet.

12. Remuneration

For participating fully in this study, a volunteer will receive a payment of £150. This figure is intended to cover all travel expenses.

In accordance with University guidelines, the volunteer will be required to sign a claim form on the last day of the research study. This form will be sent to the University's Finance Department who will send a cheque for the above sum, payable to the volunteer within 4 weeks.

The Chief Investigator expresses his appreciation to the volunteer for their involvement in this work.

SCHOOL OF PHYSICS
COLLEGE OF SCIENCE & ENGINEERING
THE UNIVERSITY OF EDINBURGH

James Clerk Maxwell Building
The King's Buildings
Mayfield Road
Edinburgh EH9 3JZ

Tel: 0131 650 5273
Fax: 0131 650 5902
Email: info@ph.ed.ac.uk
Web Site: http://www.ph.ed.ac.uk/

VOLUNTEER CONSENT FORM

ACOUSTIC PULSE REFLECTOMETRY & MAGNETIC RESONANCE IMAGING CONSENT FORM FOR HEALTHY VOLUNTEERS

Research Project

COMPARATIVE STUDY BETWEEN ACOUSTIC PULSE REFLECTOMETRY (APR) & MAGNETIC RESONANCE IMAGING (MRI) FOR VOLUMETRIC MEASUREMENT OF THE VOCAL TRACT DURING VOWEL PRODUCTION

Principal Investigator

Mr Calum Gray
Postgraduate Researcher – School of Physics, University of Edinburgh

- I have read the Information Sheet that has been provided to me, and this Consent Form, and have been given the opportunity to ask questions about them. I am satisfied that I have all the information that I need to provide **informed consent**.

- As a volunteer, I understand that I am not being scanned at the request of a Doctor for any specific medical condition.

- I understand that my Doctor will be informed of my participation in this study, and know that he/she will be provided with a routine clinical report.

- I know of no reason why I should not undergo Magnetic Resonance Imaging or take part in the study.

- I am not suffering any respiratory (breathing) disease, and am currently in good health

- I know that I am under no obligation to take part in the study and **I can withdraw at any time**.

- I understand and agree that medical images obtained during my scan will be stored and processed using computers and that after the study is completed, these may be copied onto a permanent record which might be studied again at a later time.

- I have been informed of and agree to a photographic record being taken of this study. I understand that all photographs will be of the procedure only, therefore any photographs that would include me will be altered to protect my identity.

- I understand and agree that information gathered during my scan will be shared with other medical and scientific researchers, subject to strict laws and University of Edinburgh policies intended to safeguard my privacy.

- **I agree to participate in the study**.

Signature of volunteer

Name of volunteer (please print in block capitals)

Witnessed by (signature)

Name of witness (please print in block capitals)

Date

Name of Volunteer's GP

GP's address

CN Number (for use by SBIR Centre staff)

Three copies of this form will require to be signed before scanning commences:

- *One copy for the Chief Investigator (School of Physics, University of Edinburgh)*
- *One copy for the Centre Coordinator (Brain Imaging Research Centre, Western General Hospital)*
- *One copy for the volunteer*

Postgraduate Researcher (PhD)
Rm.4201 - James Clerk Maxwell Building
The King's Buildings
Mayfield Road
Edinburgh EH9 3JZ

Tel: 0131 650 5257
Fax: 0131 650 5902
Email: cgray@ph.ed.ac.uk
Web Site: http://www.ph.ed.ac.uk/fluids/

STUDY PROTOCOL

Version 2   Date: 06/09/04

LREC Ref: 04/S1101/22

COMPARATIVE STUDY BETWEEN ACOUSTIC PULSE RELFECTOMETRY (APR) & MAGNETIC RESONANCE IMAGING (MRI) FOR VOLUMETRIC MEASUREMENT OF THE VOCAL TRACT DURING VOWEL PRODUCTION

Chief Investigator:
Mr Calum Gray
School of Physics
The University of Edinburgh.

Email: cgray@ph.ed.ac.uk
Contact Tel No. 0131 6657645

This study is divided into two parts:

(a) APR measurements - To be performed in an acoustics lab in James Clerk Maxwell Building, The King's Buildings, Mayfield Road, Edinburgh.

(b) MRI measurements – To be performed at SHEFC Brain Imaging Research Centre for Scotland, Western General Hospital, Edinburgh.

One volunteer is required to participate in both parts of the study.

It is expected that the volunteer will be required for up to 8 hours in total which will be spread over four days (not necessarily consecutive).

The study aims to attract the participation of a voice professional (singer, linguist, phonetician, public speaker, voice technologist, musician (brass/woodwind player) etc.) as the volunteer is required to be able to hold their breath comfortably for short periods of time and have good control over voice, articulators (lips & tongue) and breathing.

A full list of volunteer inclusion criteria can be found in the Volunteer Information Sheet.

(a) APR Measurements

APR has been used with humans over many years in both clinical and research fields. It is deemed a very safe technique as it only uses audio signals created by a loudspeaker with the loudness of the signal kept to within comfortable listening levels.

Procedure for each vocal-tract/vowel measurement is as follows:

- The participant will lie supine on a bed and will be asked to phonate the vowel sound for a sustained period and then halt phonation but hold the phonatory position (vocal tract and articulators).
- The participant will be asked to hold the mouthpiece of the reflectometer (positioned next to them) against their lips and then remain still, holding their breath for approximately 10-15 seconds as a measurement is taken.

It is intended that between 3 and 5 vowels will be measured with a minimum of 5 measurements for each vowel (one measurement takes 10-15 seconds).

As this could be quite tiring for the subject, the duration of this section of the study will span two days, allowing for the possibility of repeat measurements if necessary.

(b) MRI Measurements

MRI is a safe and proven technique using powerful magnets and radio waves which requires the participant to remain still whilst images are taken. For this reason, the participant will be required to either hold their breath for a short period or shallow breathe (approximately 10 seconds) whilst scanning is taking place.

Procedure for each vocal-tract/vowel measurement is as follows:

- The participant will lie supine on the MRI scanner table and will be asked to phonate the vowel sound for a sustained period and then halt phonation but hold the phonatory position (vocal tract and articulators).

- The participant will be asked to remain still and hold their breath for approximately 10 seconds as the scan is initiated. (A trial with the participant shallow breathing instead of holding their breath is also planned).

- This procedure will produce 1 image slice of the vocal tract. To encompass the whole of the vocal tract, it is envisaged that approximately 30 images will be required per vowel.

It is intended that between 3 and 5 vowels will be investigated, therefore a minimum of ~90 scans each lasting 10 seconds will be required. Again, as this could be quite tiring for the subject, the duration of this section of the study will span two days, allowing for the possibility of repeat measurements if necessary.

The participant will be asked to use a mouthpiece during the scan time in order to recreate the conditions from previous APR experiments (to gain an accurate comparison of measuring the vocal tract using MRI and APR). Extra rests/repositioning periods will be added into the scanning procedure.

Post processing of MRI data will be performed at the Image Analysis Core, Wellcome Trust Clinical Research Facility, Edinburgh.

ACOUSTICS & FLUIDS RESEARCH GROUP
SCHOOL OF PHYSICS – COLLEGE OF SCIENCE & ENGINEERING
THE UNIVERSITY OF EDINBURGH

Mr Calum Gray
Postgraduate Researcher (PhD)
Rm.4201 - James Clerk Maxwell Building
The King's Buildings
Mayfield Road
Edinburgh EH9 3JZ

Tel: 0131 650 5257
Fax: 0131 650 5902
Email: cgray@ph.ed.ac.uk
Web Site: http://www.ph.ed.ac.uk/fluids/

6th September 2004

Information for General Practitioner

Dear General Practitioner,

Volunteer: ***(Volunteer Name)***

Research Study: **COMPARATIVE STUDY BETWEEN ACOUSTIC PULSE RELFECTOMETRY (APR) & MAGNETIC RESONANCE IMAGING (MRI) FOR VOLUMETRIC MEASUREMENT OF THE VOCAL TRACT DURING VOWEL PRODUCTION**

The above named has taken part in a research study into vocal tract measurement.

This study involved Acoustic Pulse Reflectometry (APR) and Magnetic Resonance Imaging (MRI) scans of the vocal tract.

Please find attached a copy of the Patient Information Sheet for this study which explains fully the procedures involved.

The results of the MRI scan will be forwarded to you by the Brain Imaging Research Centre for Scotland (BIRCS), Western General Hospital, as per normal patient scan procedures.

If there is any further information you may require, please do not hesitate to contact me.

Yours faithfully,

Mr Calum Gray
Chief Investigator

enc.

Appendix B

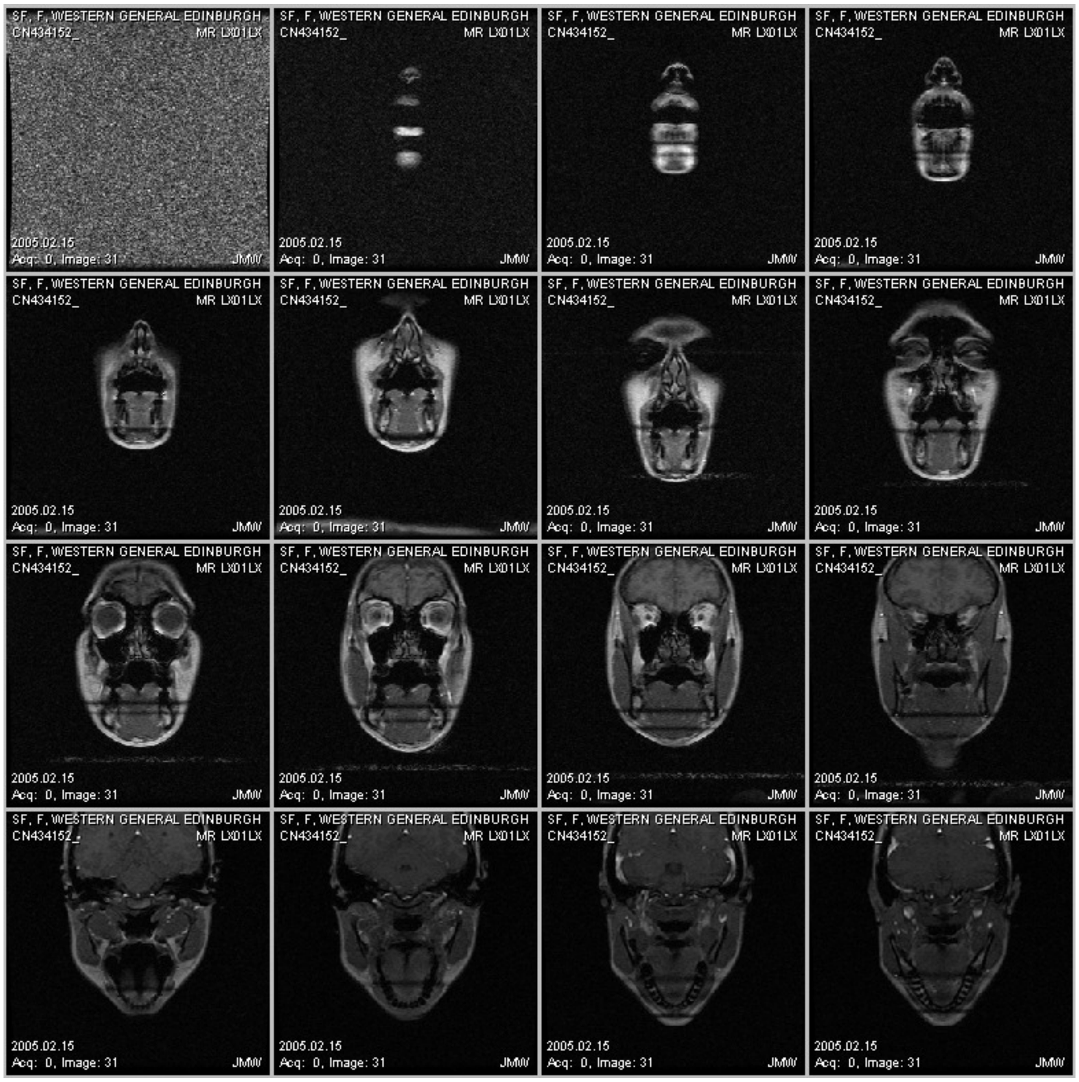
MR images for production of vowels

B.1 MRI Data Set: Vowel [a]

Midsagittal ‘profile’ image for vowel [a]

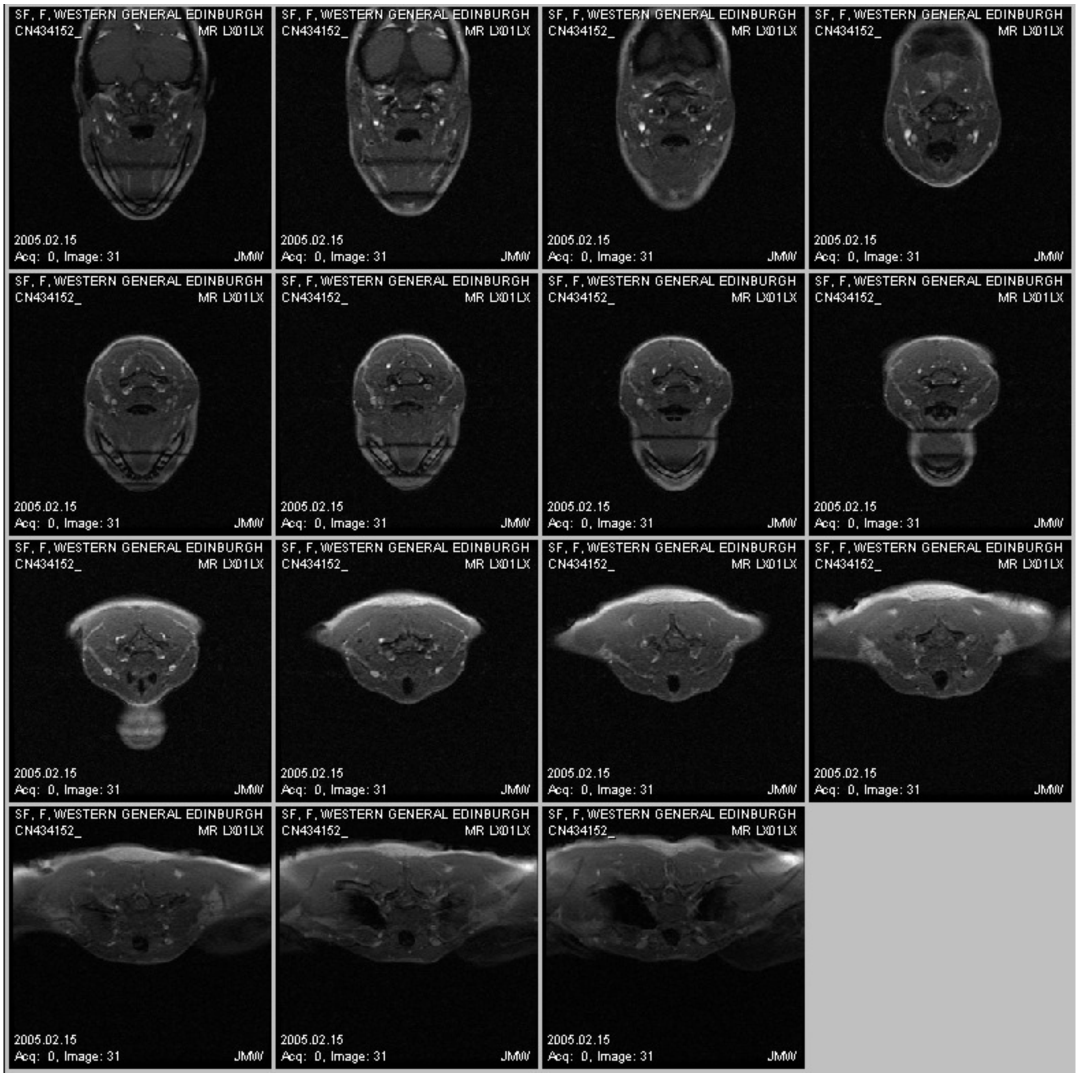
MRI Data Set: Vowel [a]

Images 1 to 20 of 31



MRI Data Set: Vowel [a] cont.

Images 21 to 31 of 31

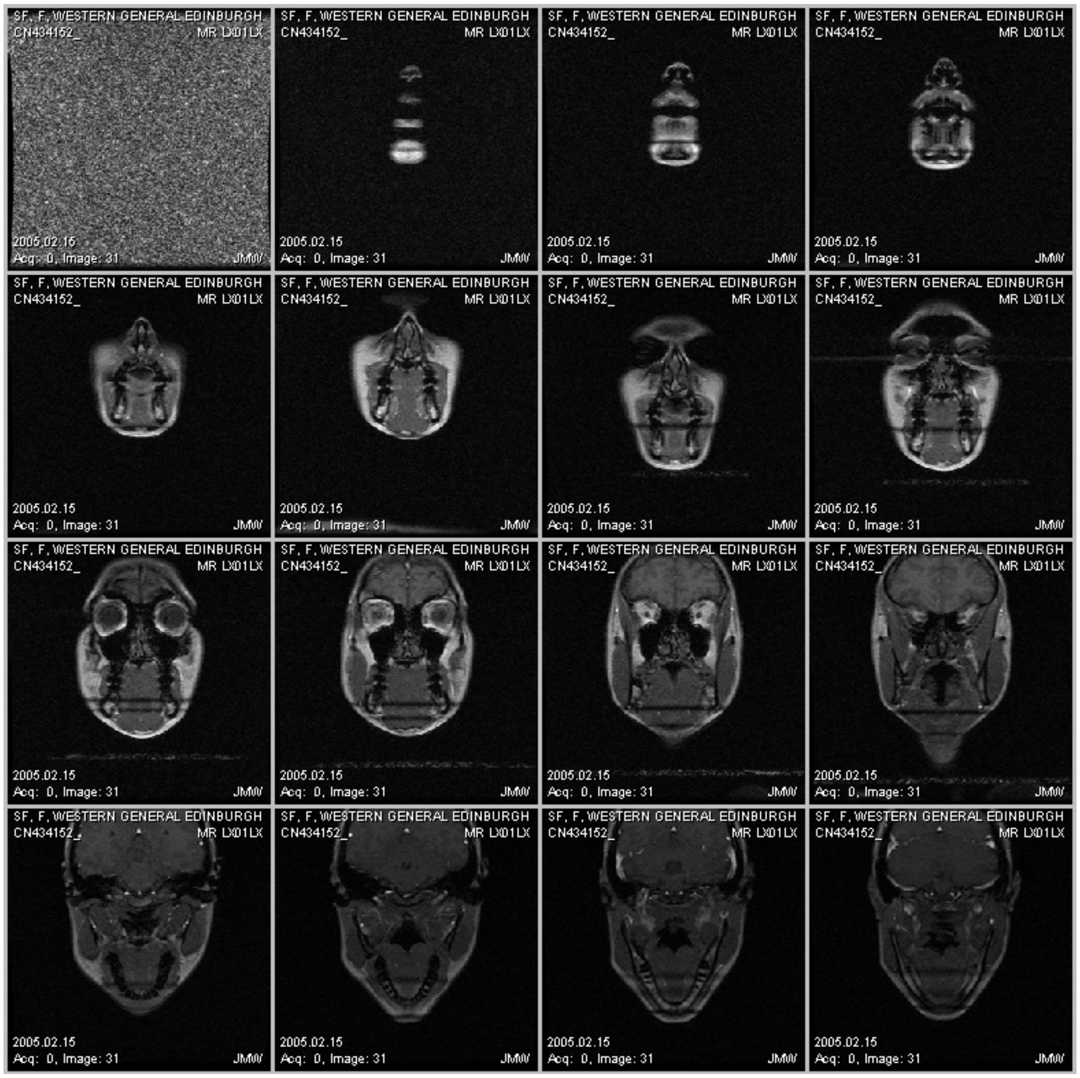


B.2 MRI Data Set: Vowel [i]

Midsagittal 'profile' image for vowel [i]

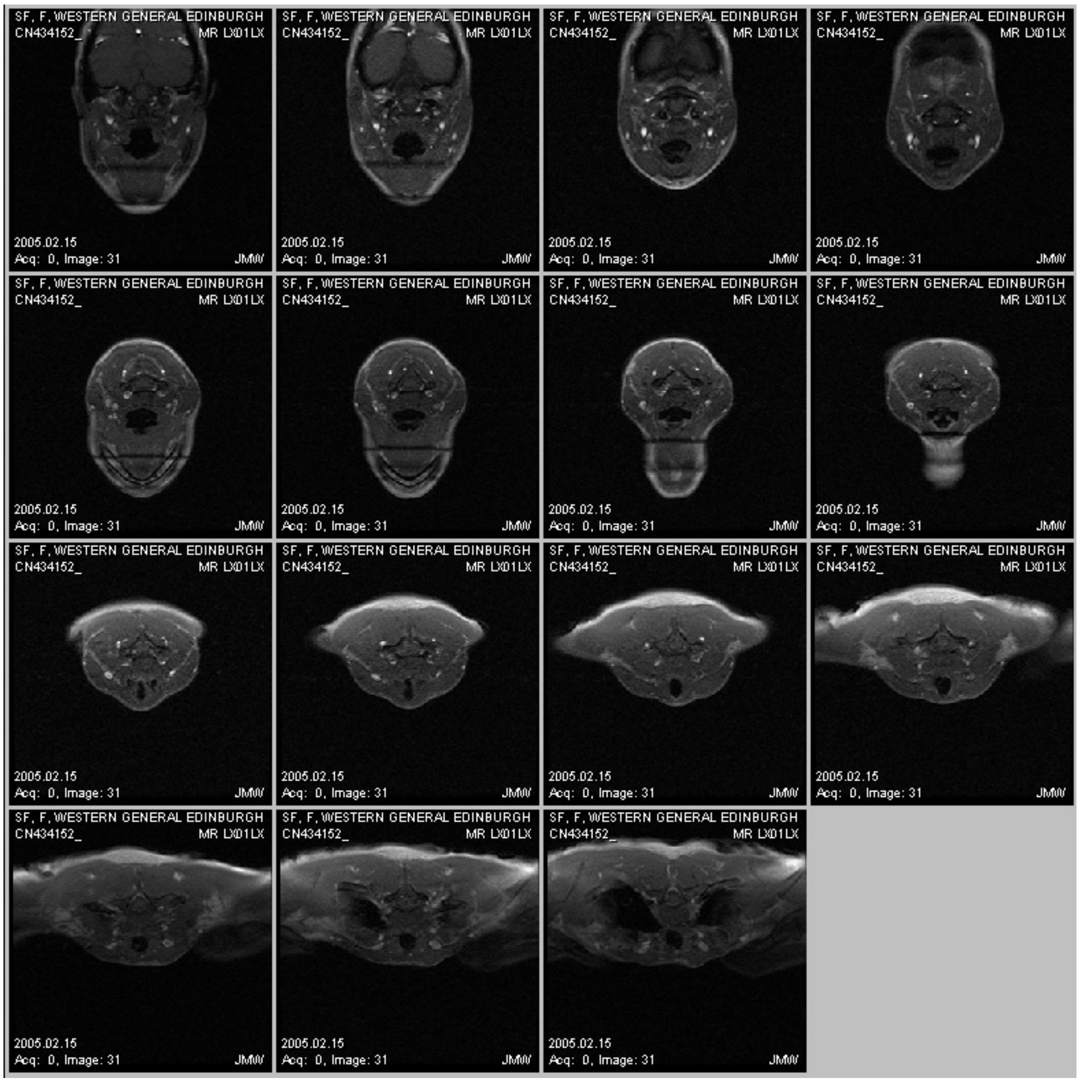
MRI Data Set: Vowel [i]

Images 1 to 20 of 31



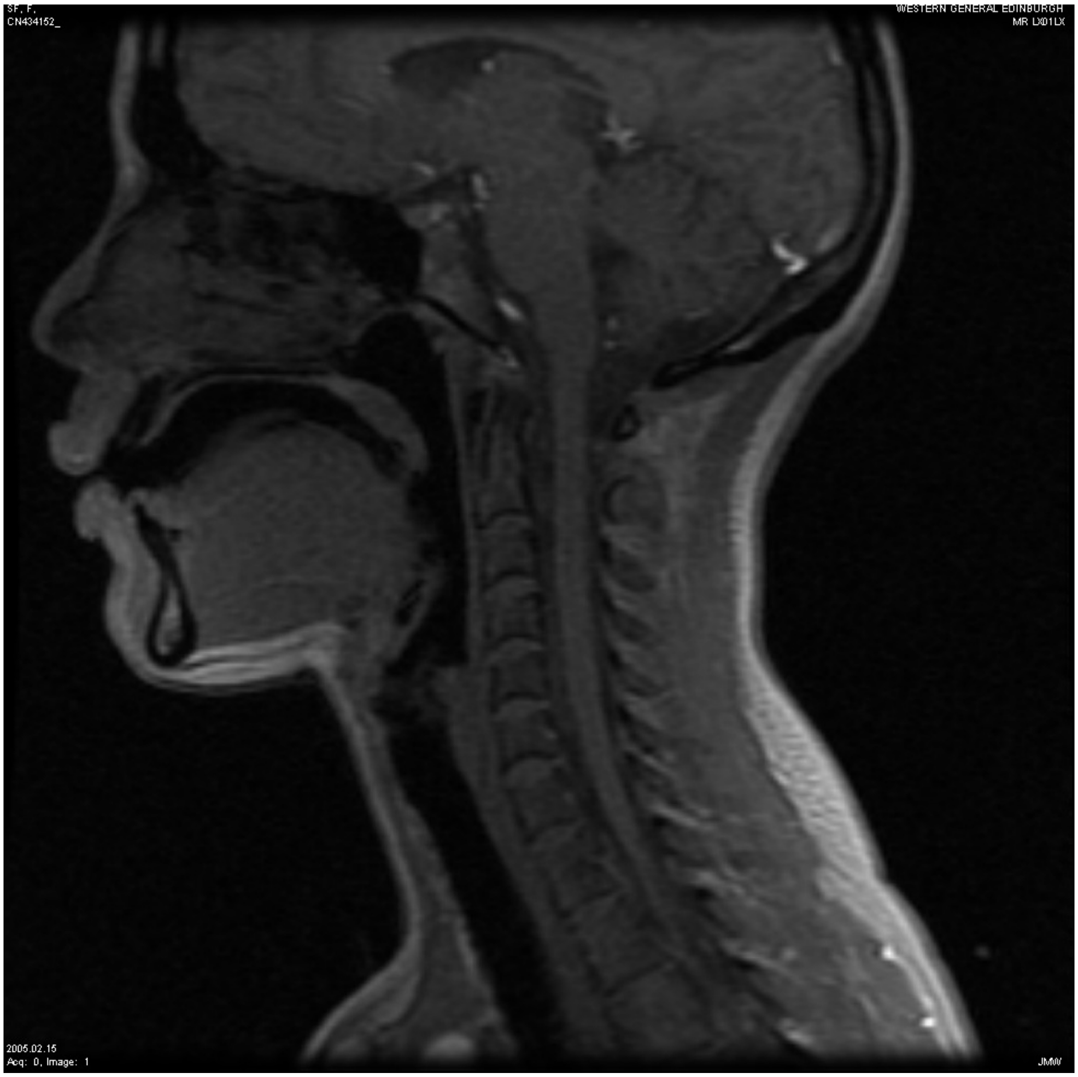
MRI Data Set: Vowel [i] cont.

Images 21 to 31 of 31



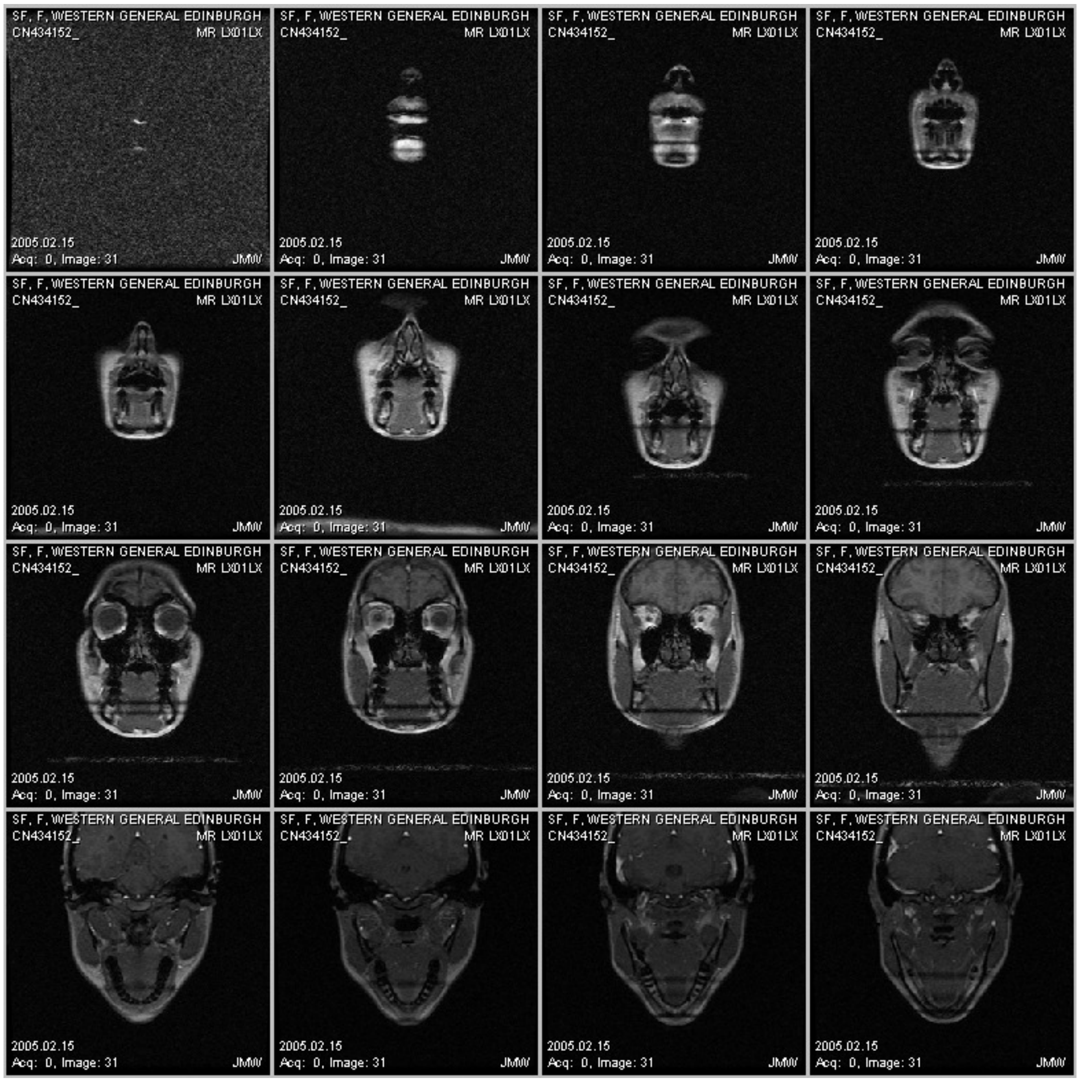
B.3 MRI Data Set: Vowel [u]

Midsagittal 'profile' image for vowel [u]



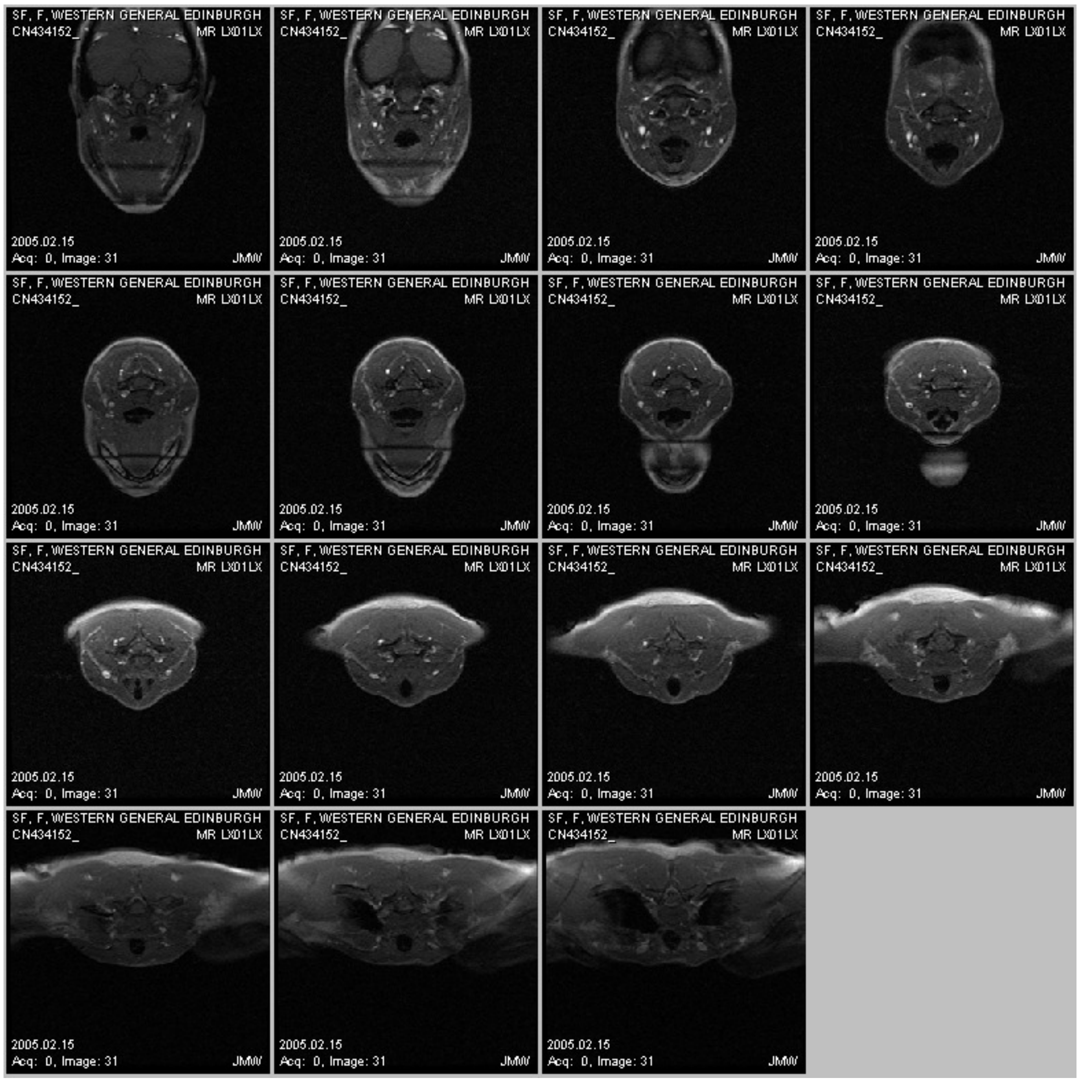
MRI Data Set: Vowel [u]

Images 1 to 20 of 31



MRI Data Set: Vowel [u] cont.

Images 21 to 31 of 31



Bibliography

[1] Jespersen, O., *Language, Its Nature, Development and Origin*. Unwin Hyman. 1968

[2] Clark, R.A.J. *Generating Synthetic Pitch Contours Using Prosodic Structure*. PhD thesis (2003), The Department of Linguistics, University of Edinburgh.

[3] Bailly, G., Campbell, N., and Möbius, B. *ISCA Special Session: Hot Topics in Speech Synthesis*. In Proc. *EUROSPEECH 2003*. Geneva, Switzerland, 2003.

[4] Shadle, C.H. and Damper, R.I. *Prospects for articulatory synthesis: A position paper*. In Proc. *4th ISCA ITRW on Speech Synthesis*. Blair Atholl, Scotland, 2001.

[5] Kob, M. *Physical Modeling of the Singing Voice*. PhD thesis (2002), Institute of Technical Acoustics, Faculty of Electrical Engineering and Information Technology, Aachen University.

[6] Story, B.H. and Titze, I.R., *Voice simulation with a body-cover model of the vocal folds.* J Acoust Soc Am, 97(2): p. 1249-60,1995.

[7] Marshall, I. *Acoustic reflectometry for airway measurement*. PhD thesis (1992), University of Edinburgh.

[8] Douglas, N.J., *Upper airway imaging.* Clin Phys Physiol Meas, 11 Suppl A: p. 117-9,1990.

[9] Mermelstein, P., *Determination of the vocal-tract shape from measured formant frequencies.* J Acoust Soc Am, 41(5): p. 1283-94,1967.

[10] Schroeder, M.R., *Determination of the geometry of the human vocal tract by acoustic measurements.* J Acoust Soc Am, 41(4): p. Suppl:1002-10,1967.

[11] Sondhi, M.M. and Gopinath, B., *Determination of vocal-tract shape from impulse response at the lips.* J Acoust Soc Am, 49(6): p. 1867-73,1971.

[12] Ware, J.A. and Aki, K., *Continuous and discrete inverse-scattering problems in a stratified elastic medium. I. Plane waves at normal incidence.* J Acoust Soc Am, 45: p. 911-921,1969.

[13] Sharp, D.B. *Acoustic pulse reflectometry for the measurement of musical wind instruments*. PhD thesis (1996), Department of Physics and Astronomy, University of Edinburgh.

[14] Jackson, A.C., Butler, J.P., Millet, E.J., Hoppin, F.G., Jr., et al., *Airway geometry by analysis of acoustic pulse response measurements.* J Appl Physiol, 43(3): p. 523-36,1977.

[15] Jackson, A.C. and Olson, D.E., *Comparison of direct and acoustical area measurements in physical models of human central airways.* J Appl Physiol, 48(5): p. 896-902,1980.

[16] Fredberg, J.J., Wohl, M.E., Glass, G.M., and Dorkin, H.L., *Airway area by acoustic reflections measured at the mouth.* J Appl Physiol, 48(5): p. 749-58,1980.

[17] Brooks, L.J., Castile, R.G., Glass, G.M., Griscom, N.T., et al., *Reproducibility and accuracy of airway area by acoustic reflection.* J Appl Physiol, 57(3): p. 777-87,1984.

[18] Rubinstein, I., McClean, P.A., Boucher, R., Zamel, N., et al., *Effect of mouthpiece, noseclips, and head position on airway area measured by acoustic reflections.* J Appl Physiol, 63(4): p. 1469-74,1987.

[19] Hoffstein, V. and Fredberg, J.J., *The acoustic reflection technique for non-invasive assessment of upper airway area.* Eur Respir J, 4(5): p. 602-11,1991.

[20] Marshall, I., *Impedance reconstruction methods for pulse reflectometry.* Acustica, 76(3): p. 118-128,1992.

[21] Marshall, I., *Acoustic reflectometry with an arbitrarily short source tube.* J Acoust Soc Am, 91(6): p. 3558-3564,1992.

[22] Marshall, I., Maran, N.J., Martin, S., Jan, M.A., et al., *Acoustic reflectometry for airway measurements in man: implementation and validation.* Physiol Meas, 14(2): p. 157-69,1993.

[23] Marshall, I., Rogers, M., and Drummond, G., *Acoustic reflectometry for airway measurement. Principles, limitations and previous work.* Clin Phys Physiol Meas, 12(2): p. 131-41,1991.

[24] Watson, A.P. *Impulse measurements on tubular acoustic systems*. PhD thesis (1989), University of Surrey.

[25] Watson, A.P. and Bowsher, J.M., *Recent progress in time domain work on brass instruments.* Proc. Institute of Acoustics,1987.

[26] Watson, A.P. and Bowsher, J.M., *Impulse measurements on brass musical instruments.* Acustica, 66(3): p. 170-174,1988.

[27] Amir, N., Rosenhouse, G., and Shimony, U., *A discrete model for tubular acoustic systems with varying cross-section - the direct and inverse problems. Part 1 and Part 2: Theory and Experiment.* Acustica, 81: p. 450-474,1995.

[28] Amir, N., Rosenhouse, G., and Shimony, U., *Losses in tubular acoustic systems - theory and experiment in the sampled time and frequency domains.* Acustica, 82(1): p. 1-8,1996.

[29] Sharp, D.B. and Campbell, D.M., *Leak detection in pipes using acoustic pulse reflectometry.* Acustica, 83(3): p. 560-566,1997.

[30] Sharp, D.B., Myers, A., Parks, R., and Campbell, D.M. *Bore reconstruction by pulse reflectometry and its potential for taxonomy of brass instruments*. In Proc. *15th International Congress on Acoustics*. Trondheim, Norway, 1995.

[31] Kemp, J. *Theoretical and experimental study of wave propagation in brass musical instruments*. PhD thesis (2003), University of Edinburgh.

[32] Campbell, D.M. and Greated, C.A., *The Musician's Guide to Acoustics*. Oxford University Press. 1994

[33] Fant, G., *Acoustic Theory of Speech Production*. 2nd edition ed. Mouton, Paris. 1970

[34] Titze, I.R., *Principles of voice production*. 2nd ed.: National Center for Voice and Speech. 2000

[35] Titze, I.R. and Strong, W.J., *Normal modes in vocal cord tissues.* J Acoust Soc Am, 57(3): p. 736-49,1975.

[36] Ishizaka, K. and Flanagan, J.L., *Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords.* Tech. J., 51: p. 1233--1268,1972.

[37] Titze, I.R., *The human vocal cords: a mathematical model. I.* Phonetica, 28(3): p. 129-70,1973.

[38] Titze, I.R., *The human vocal cords: a mathematical model. II.* Phonetica, 29(1): p. 1-21,1974.

[39] Ladefoged, P., *Elements of Acoustic Phonetics*. 2nd ed.: University Of Chicago Press. 1995

[40] Smith, J.O., *Physical modeling using digital waveguides.* Computer Music Journal, 16: p. 74-91,1992.

[41] Kelly, J.L. and Lochbaum, C.C. *Speech Synthesis*. In Proc. *Fourth International Congress on Acoustics*. Copenhagen, Denmark, 1962.

[42] Välimäki, V. and Karjalainen, M. *Improving the Kelly-Lochbaum vocal tract model using conical tube sections and fractional delay filtering techniques*. In Proc. *Int. Conf. Spoken Language Processing*. Yokohama, Japan, 1994.

[43] Flanagan, J.L., *Analysis Synthesis and Perception of Speech*. 2nd ed. Springer-Verlag, Berlin. 1972

[44] Mott, F.W. *The Brain and the Voice in Speech and Song*. [EBook #13111] 2004 [cited 2005 July]; Available from: www.gutenberg.org.

[45] Macon, M.W., Jensen-Link, L., Oliverio, J., Clements, M.A., et al. *A Singing Voice Synthesis System Based On Sinusoidal Modeling*. In Proc. *International Conference on Acoustics, Speech, and Signal Processing*. Munich, Germany, 1997.

[46] Kaiser, J.F. *Some observations on vocal tract operation from fluid flow point of view*. In Proc. *Conference on Physiology and Biophysics of Voice*. Iowa City, IA, 1983.

[47] Bloothooft, G. and Plomp, R., *The sound level of the singer's formant in professional singing.* J Acoust Soc Am, 79(6): p. 2028-33,1986.

[48] Sundberg, J., *Articulatory interpretation of the "singing formant".* J Acoust Soc Am, 55(4): p. 838-44,1974.

[49] Cook, P.R. *Identification of control parameters in an articulatory vocal tract model, with application to the synthesis of singing*. PhD Thesis (1990), Stanford University.

[50] Ziskin, M.C., *Basic physics of ultrasound*, in *The principles and practice of ultrasonography in obstetrics and gynaecology*, R.G. Sader and A.E. James, Editors., Appleton Century Crofts, New York. p.1-23. 1980.

[51] Kamal, I., *Acoustic Reflectometry of the Nose and Pharynx*. Boca Raton, Florida: Universal Publishers. 2004

[52] Kinsler, L.E., Coppens, A.R., and Sanders, J.V., *Fundamentals of Acoustics*. John Wiley & Sons. 2000

[53] Morse, P.M. and Ingard, K.U., *Theoretical Acoustics*. Princeton University Press. 2000

[54] Li, A. *Improvements to the acoustic pulse reflectometry technique for measuring duct dimensions*. PhD thesis (2004), Department of Environmental and Mechanical Engineering, Faculty of Technology, The Open University.

[55] Keefe, D.H., *Acoustical wave-propagation in cylindrical ducts - Transmission-line parameter approximations for isothermal and non-isothermal boundary conditions.* J Acoust Soc Am, 75(1): p. 58-62,1984.

[56] Stahnke, W., *Primitive binary polynomials.* Mathematics of Computation, 27: p. 977-980,1973.

[57] Forbes, B.J., Sharp, D.B., Kemp, J.A., and Li, A., *Singular system methods in acoustic pulse reflectometry.* Acustica, 89: p. 743-753,2003.

[58] Chu., W.T., *Impulse-response and reverberation-decay measurements made by using a periodic pseudorandom sequence.* Applied Acoustics, 29: p. 193-205,1990.

[59] Hsu, K.C. *Simulation of room acoustics*. Undergraduate thesis (1996), University of Queensland, Australia.

[60] Schroeder, M.R., *Integrating-impulse method measuring sound decay without using impulses.* J Acoust Soc Am, 66(2): p. 497-500,1979.

[61] de-Salis, M.H.F., Movchan, N.V., and Oldham, D.J., *Characterising holes in duct walls using resonance frequencies.* J Acoust Soc Am, 111: p. 2583-2593,2002.

[62] MacWilliams, F.J. and Sloane, N.J.A., *Pseudo-random sequences and arrays.* Proceedings of the IEEE, 64(12): p. 1715-1729,1976.

[63] Wright, M.C.M., ed. *Lecture notes on the mathematics of acoustics*. Imperial College Press: London. 2004.

[64] Borish, J. and Angell, J.B., *An efficient algorithm for measuring the impulse response using pseudorandom noise.* Journal of the Audio Engineering Society, 31(7): p. 478-487,1983.

[65] Kemp, J.A., Buick, J.M., and Campbell, D.M. *Practical improvements to acoustic pulse reflectometry: the virtual dc tube method and source reflection cancellation*. In Proc. *International Symposium on Musical Acoustics (ISMA).* Perugia, Italy, 2001.

[66] Sondhi, M.M. and Resnick, J.R., *The inverse problem for the vocal tract: numerical methods, acoustical experiments, and speech synthesis.* J Acoust Soc Am, 73(3): p. 985-1002,1983.

[67] Gray, C.D., Greated, C.A., and Campbell, D.M. *Acoustic pulse reflectometry for vocal tract measurement*. In Proc. *Stockholm Music Acoustics Conference*. Stockholm, Sweden, 2003.

[68] Titze, I.R., *Speaking vowels versus singing vowels.* The Journal of Singing, 52(September): p. 47,1995.

[69] Sundberg, J., *The science of the singing voice*. Northern Illinois University Press. 1987

[70] Peterson, B. and Barney, H., *Control methods used in a study of the vowels.* J Acoust Soc Am, 24: p. 175-184,1952.

[71] Nair, G., *Voice-Tradition and Technology: A State of the Art Studio*. San Diego, CA: Singular Publishing Group. 1999

[72] Appelman, D.R., *The science of vocal pedagogy, theory and application.* Bloomington, IN: Indiana Universit Press. 1967

[73] Martin, G.C. *CG Imagery Development*. 1996 [cited 2005 January 2004]; Available from: http://www.garycmartin.com/.

[74] Fletcher, N.H. and Rossing, T.D., *The physics of musical instruments*. Springer-Verlag New York Inc. 1999

[75] Espy-Wilson, C. *Articulatory strategies, speech acoustics and variability*. In Proc. *From Sound to Sense*. MIT, 2004.

[76] Louis, B., Glass, G., Kresen, B., and Fredberg, J., *Airway area by acoustic reflection: the two-microphone method.* J Biomech Eng, 115(3): p. 278-85,1993.

[77] Story, B.H., Titze, I.R., and Hoffman, E.A., *Vocal tract area functions from magnetic resonance imaging.* J Acoust Soc Am, 100(1): p. 537-54,1996.

[78] Story, B.H., Titze, I.R., and Hoffman, E.A., *Vocal tract area functions for an adult female speaker based on volumetric imaging.* J Acoust Soc Am, 104(1): p. 471-87,1998.

[79] Goldstein, U. *An articulatory model for the vocal tracts of growing children*. PhD thesis (1980), MIT.

[80] Clement, P.A.R. and Gordts, F., *Consensus report on acoustic rhinometry and rhinomanometry - (Standardisation Committee on Objective Assessment of the Nasal Airway, I.R.S. and E.R.S.).* Rhinology, 43: p. 169-179,2005.

[81] Blink, E. *Basic MRI Physics*. 2004 [cited 2005 January 2005]; Available from: www.mri-physics.com.

[82] Hornak, J.P. *The Basics of MRI*. 2003 [cited 2005 January 2005]; Available from: www.cis.rit.edu/htbooks/mri/.

[83] Shellock, F.G. *MRI Safety*. 2005 [cited 2005 January 2005]; Available from: www.mrisafety.com/.

[84] Baer, T., Gore, J.C., Gracco, L.C., and Nye, P.W., *Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels.* J Acoust Soc Am, 90(2 Pt 1): p. 799-828,1991.

[85] Engwall, O. *A revisit to the application of MRI to the analysis of speech production - testing our assumptions*. In Proc. *6th International Seminar on Speech Production*. Sydney, 2003.

[86] Engwall, O. and Badin, P. *An MRI study of Swedish fricatives: coarticulatory effects*. In Proc. *5th Speech Production Seminar*. München, Germany, 2000.

[87] Narayanan, S.S., Alwan, A.A., and Haker, K., *Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals.* J Acoust Soc Am, 101(2): p. 1064-77,1997.

[88] Engwall, O. and Badin, P., *Collecting and analysing two- and three-dimensional MRI data for Swedish.* TMH-QPSR, 4,1999.

[89] Lindblom, B.E. and Sundberg, J.E., *Acoustical consequences of lip, tongue, jaw, and larynx movement.* J Acoust Soc Am, 50(4): p. 1166-79,1971.

[90] Mermelstein, P., *Articulatory model for the study of speech production.* J Acoust Soc Am, 53(4): p. 1070-82,1973.

[91] Narayanan, S., Alwan, A., and Haker, K., *An articulatory study of fricative consonants using magnetic resonance imaging.* J Acoust Soc Am, 98: p. 1325-1347,1995.

[92] D'Urzo, A.D., Rubinstein, I., Lawson, V.G., Vassal, K.P., et al., *Comparison of glottic areas measured by acoustic reflections vs. computerized tomography.* J Appl Physiol, 64(1): p. 367-70,1988.

[93] Sundberg, J., *The acoustics of the singing voice.* Sci Am, 236(3): p. 82-4, 86, 88-91,1977.

[94] Jan, M.A., Marshall, I., and Douglas, N.J., *Effect of posture on upper airway dimensions in normal human.* Am J Respir Crit Care Med, 149(1): p. 145-8,1994.

[95] Martin, S.E., Marshall, I., and Douglas, N.J., *The effect of posture on airway caliber with the sleep-apnea/hypopnea syndrome.* Am J Respir Crit Care Med, 152(2): p. 721-4,1995.

[96] Martin, S.E., Mathur, R., Marshall, I., and Douglas, N.J., *The effect of age, sex, obesity and posture on upper airway size.* Eur Respir J, 10(9): p. 2087-90,1997.

[97] Shen, H., Huang, J., Yamanouchi, K., Sakurai, S., et al., *Effect of posture on pharyngeal area measured by acoustic reflection technique.* Nihon Kyobu Shikkan Gakkai Zasshi, 29(5): p. 554-9,1991.

[98] Yildirim, N., Fitzpatrick, M.F., Whyte, K.F., Jalleh, R., et al., *The effect of posture on upper airway dimensions in normal subjects and in patients with*

the sleep apnea/hypopnea syndrome. Am Rev Respir Dis, 144(4): p. 845-7,1991.

[99] ASSA. *Treatment options for adults with obstructive sleep apnea*. [cited 2005 January 2005]; Available from: http://www.sleepapnea.org/resources/pubs/treatment.html.

[100] Marcus, C.L., McColley, S.A., Carroll, J.L., Loughlin, G.M., et al., *Upper airway collapsibility in children with obstructive sleep apnea syndrome.* J Appl Physiol, 77(2): p. 918-24,1994.

[101] Schwartz, A.R., Eisele, D.W., and Smith, P.L., *Pharyngeal airway obstruction in obstructive sleep apnea: pathophysiology and clinical implications.* Otolaryngol Clin North Am, 31(6): p. 911-8,1998.

[102] van-Walstijn, M. and Campbell, D.M. *Large-bandwidth measurement of acoustic input impedance of tubular objects*. In Proc. *Institute of Acoustics Spring Conference*. Salford, UK, 2002.