Resonance Tuning in Professional Operatic Sopranos

Rebecca Rose Vos

PhD

University of York
Electronic Engineering

August 2018

Abstract

Soprano singers are capable of singing at pitches exceeding 1000 Hz, where the spacing of the harmonics means that the vocal tract resonances are not fully utilised. Sopranos therefore move the articulators, to "tune" the resonances of the vocal tract near to harmonics of the voice source, improving the efficiency of sound production. Although resonance tuning has been observed in soprano singers, it is not yet understood how this phenomenon is achieved and which articulators play the most significant roles in altering the vocal tract resonances. A preliminary experiment explored the use of broad band noise excitation of the vocal tract to observe resonance tuning behaviour in girl choristers. A second experiment extended this procedure to include MRI to observe the vocal tracts of 6 professional soprano opera singers and investigate how the articulators affect vocal tract resonances. The effects of MRI measurement conditions on singers were also investigated to establish whether measurements obtained during MRI are representative of normal singing. Finally, a perceptual test was conducted to study the perception of different methods of resonance tuning. As expected, considerable R_1:f_0 tuning, and some R_2:$2f_0$ tuning was observed in both groups. MRI revealed some links between resonances and articulators, however no consistent patterns in production were observed across subjects. The results showed strong differences in resonance production between different vowels and subjects, suggesting that resonance tuning production is not only a complex and context-specific topic, but also highly individual.

Contents

List of Figures

List of Tables

Acknowledgements

This research would not have been possible without my wonderful supervisors, Helena Daffern, Damian Murphy, and formerly David Howard, for their incredible patience, guidance and support throughout this PhD research.

I would also like to thank Chris Mellor, of the Maths Skills Centre at the University of York, for his patience and invaluable help with statistics, Andrew Chadwick, for extensive technical help and help in building equipment, and York Neuro-Imaging centre, for their assistance with the MRI scans.

Thanks also to Nathalie Henrich and her team at the Speech and Cognition Department, GIPSA-lab, Grenoble, France for their guidance with the research on the perception of Resonance tuning, and John Smith and Joe Wolfe of the Acoustics Lab at the University of New South Wales, who developed and wrote the software for measuring vocal tract resonances using broad band noise excitation.

Lastly, thanks to my Parents, Lynne and Hans Vos, for their support and my wonderful education, and Jim, without whom this thesis would never have been finished. A final thanks to all my other friends, relatives and colleagues, for their help, support and friendship over the last few years.

Funding

This work was generously supported by:

- The Audio Engineering Society of America (Educational Foundation grant for graduate studies)
- The University of York Department of Electronic Engineering (Departmental Scholarship)
- The British Federation of Women Graduates (Beryl Mavis Green prize for Academic Excellence).

Declaration

I hereby declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this or any other University. All sources are explicitly stated and referenced.

I also declare that some parts of the work in this thesis have been presented previously, at conferences and in journals, in the following publications:

Differences in vocal tract resonances introduced by MRI conditions in a male and female singer (pilot study), R R Vos, H Daffern, and D M Howard, presented at the 11th Pan-European Voice Conference (PEVOC-11), Florence, Italy, 31st August - 2nd September, 2015.

Resonance tuning in three girl choristers, R R Vos, H Daffern, and D M Howard, Journal of Voice, vol. 31, no. 1, pp. 122-e1, 2017.

The Perception of Formant Tuning in Soprano Voices, R R Vos, D T Murphy, D M Howard, and H Daffern, Journal of Voice, vol. 32, no. 1, pp 126-el, 2018.

Determining The Relevant Criteria For 3D Vocal Tract Characterisation, R R Vos, H Daffern, and D M Howard, presented at the 45th Annual Symposium of the Voice Foundation, Philadelphia, USA, 1st June - 5th June, 2016.

Determining the Relevant Criteria for Three-Dimensional Vocal Tract Characterization, R R Vos, D T Murphy, D M Howard, and H Daffern, Journal of Voice, vol. 32, no. 2, pp. 130-142, 2018.

Differences between Speaking and Singing in different Positions, R R Vos, H Daffern, and D T Murphy, presented at the 12th Pan-European Voice Conference (PEVOC-12), Ghent, Belgium, 30th August - 1st September, 2017.

Chapter 1

Introduction

> The study of the mechanism of the human voice, very instructive for the physiologist, can also have some undeniable advantages for the singer. Nothing, in fact, can be more valuable to him than to know by what procedures the vocal instrument manages to produce the vibrations, to what operation of the organs we owe the range of the voice, the registers, the timbres, the ring of the tones, their intensity, their volume, the rapid succession of the notes.
>
> ---
>
> Manuel P Garcia, A Complete Treatise on the Art of Singing, 1984 [1]

Singing is a highly complex activity requiring precise movements by the singer of parts of their body largely invisible and inaccessible to them or their teachers. To become a professional singer, an individual must spend years of dedicated study, cultivating an ability to control to a fine degree the breathing apparatus, larynx, and vocal tract. Understanding this intricate and nuanced activity provides an insight into the most fundamental method of human communication: the voice.

There are many different genres and styles of singing, including pop, church music, musical theatre and contemporary musical theatre within Western music, and even within classical music singing there are sub-genres such as consort singing, lieder, and opera. Opera represents one of the most technically demanding styles, as the singers not only have to perform to audiences of thousands without amplification, but must also act at the same time, and the best singers train for decades to hone their abilities.

Musical pedagogy is not a modern concept. Plato (c. 427-437 BC) discussed the concepts of musical education in children, and recommended three years of compulsory study on the lyre [2]. Modern (classical) singing pedagogy aims to teach singers the skills of vocal technique, including audience communication, dramatic interpretation, linguistic exactitude, and artistry and musicianship [3]. However, language used by singers and singing teachers is often based on sensation and uses imagery which can be confusing. Instructions such as "imagine you are winding a golden ribbon" can be difficult to translate into physical gestures (as described by Miller [3] p4). All this must happen with apparent ease, and in the case of opera and other musical genres, alongside acting.

Physiologically, singing is a precise and complex art, and research on the vocal tract during singing has advanced greatly in recent years [4]. The development of new techniques such as articulography, and the improvement and wider availability of existing methods such as magnetic resonance imaging (MRI) [5, 6, 7, 8] and laryngoscopy [9], have allowed researchers to gain more detailed insight into the inner workings of the *voice organ*: the breathing apparatus, larynx and vocal tract.

In male voices, with a low fundamental frequency ($\sim$ 80 - 250 Hz) the harmonics are closely spaced, and it is relatively straightforward to extract formant information from the voice spectrum. Much of the existing research examining the acoustic spectrum of singing voice has therefore focussed on male voices (for example, [10]). Until recently however, very little research existed on the female voice, as in female voices the fundamental frequencies are significantly higher than male voices (up to $\sim$ 1000 Hz), meaning that the harmonics are very widely spaced, and it is therefore more difficult to extract information about the vocal tract from the acoustic spectrum.

The wide spacing of harmonics in the female voice also means that at high pitches (exceeding 1000 Hz) sopranos (the voice type with the highest frequency range) gain little from techniques commonly observed in classically trained males, such as the Singer's Formant Cluster [11]. This is a technique whereby male singers alter the frequencies of formants 3, 4, and 5, to increase the acoustic power produced within a certain frequency range. For a soprano, there may be few or no harmonics within the appropriate frequency range [12], however sopranos are able to alter their vocal tract shape, to move the resonances closer to a nearby harmonic [13]. This has the effect of amplifying the harmonic, and boosting the acoustic energy produced by the singer with little additional energy from the singer. This technique is known as resonance tuning, and has been found to be employed by sopranos across a wide frequency range [14, 15].

This research investigates the female singing voice, specifically the production and perception of resonance tuning in soprano singing. To identify vocal tract resonances at high fundamental frequencies (nearing 1 kHz), a method of measuring the vocal tract resonances using external noise excitation is investigated. Magnetic resonance imaging will be employed to investigate the role of different articulators during singing. As the conditions experienced by subjects during MRI scans are not conducive to normal behaviour, the implications of using MRI in singing voice research will also be considered.

Finally, to contextualise the results obtained about resonance tuning in soprano singing, the listeners' perception of different methods of resonance tuning will be investigated using a subjective test.

1.1 Hypothesis

The Hypothesis tested in this PhD research is:

Noise excitation to measure vocal tract resonances, and magnetic resonance imaging to observe the articulators can be used to understand the production of resonance tuning methods employed by professional soprano singers, which vary across different vowels and pitches as well as between singers.

1.1.1 Description of Hypothesis

Noise excitation to measure the vocal tract resonances An experiment will be conducted to test the method of measuring vocal tract resonances in girl choristers using broad band noise.

Magnetic resonance imaging to observe the articulators MRI will be used to collect images of the vocal tract during singing. Comparisons will be made between resonance measurements, speech and singing recordings, and recordings of sung vowels produced under normal and MRI conditions, to verify that measurements collected by MRI are representative of normal singing production.

The production of resonance tuning Statistical analysis will be carried out to investigate the relationships between the vocal tract articulators and vocal tract resonances.

Vary across different vowels and pitches Resonance tuning will be explored for three different vowels, for six professional opera singers, across each singer's entire vocal range.

Resonance tuning varies between singers The resonance tuning methods used by both girl choristers and professional soprano opera singers will be investigated.

1.1.2 Objectives

The main objective of this PhD research is therefore:

1. To understand the methods by which resonance tuning is produced by soprano singers.

To achieve this aim, two sub-objectives have also been identified:

2. To better understand the purpose of resonance tuning by studying its perception by listeners.
3. To understand the effects of MRI measuring conditions on singers, and verify the usefulness of MRI in singing research.

1.2 Structure of Thesis

Chapter 2 begins with a Literature review, presenting an overview of voice production and different methods of analysing the voice. This is followed by a summary of resonance tuning techniques used in both low and high voices, and the perception thereof.

Chapter 3 investigates the *perceptual effects* of different resonance tuning techniques, in order to consider the effects of using resonance tuning to improve acoustic efficiency. This was achieved by asking listeners to complete a perceptual test rating synthetic singing samples using different methods of resonance tuning. This will allow the results of Chapter 6 (investigating the articulatory mechanisms used in the production of resonance tuning) to be viewed in the context of listener perception.

Chapter 4 describes a preliminary experiment carried out to measure vocal tract resonances in girl choristers, using broad band noise excitation. The data collected will allow the equipment and methods used in the main (MRI) experiment to be tested, as well as providing an insight into resonance tuning strategies used by girl choristers, a group for which this has not previously been studied.

Chapter 5 introduces the main experiment at the core of this thesis, which aims to understand which articulators are important in affecting the vocal tract resonances. It involves speaking and singing tasks in three different situations; firstly, MRI conditions, then "normal" conditions (standing in a normal performance pose in an anechoic chamber), and finally "simulated MRI" conditions (lying supine in an anechoic chamber with MRI noise played over headphones). This chapter also details the the analysis undertaken to confirm that the measurements taken in MRI conditions are comparable to "normal" conditions. This includes a comparison of resonance measurements standing and lying in the anechoic chamber, long-term average spectra (LTAS) of speaking and singing, both standing and lying, in the anechoic chamber, and comparison of the audio recorded in the MRI machine with the anechoic chamber.

Chapter 6 presents the results obtained from the experimental method introduced in Chapter 5 pertaining to the production of resonance tuning, using measurements of vocal tract resonances and measurements of articulators (obtained by MRI) in professional soprano opera singers. This chapter also discusses some of the conclusions that can be drawn from this data.

Finally, Chapter 7 provides a conclusion for this thesis, an overview of the experiments and analysis carried out, a consideration of the impact of this work, and directions for future research.

1.2.1 Novel Contributions to field

Contributions to the field, resulting from this PhD research are:

- Measurement of the type and extent of resonance tuning used by girl choristers, a group for which this has not previously been studied.

- Analysis of the effects of MRI measuring conditions on professional singers during speaking and singing, by comparing resonance measurements, long-term average spectra, and recordings of sung vowels.
- Measurement of the type and extent of resonance tuning in professional soprano opera singers (at national/international principal level). A group of this size and level has not previously been studied.
- An improved algorithm for generating three-dimensional area functions from MRI scans of the vocal tract.
- Measurement of articulators in singing using three-dimensional magnetic resonance imaging, in professional soprano opera singers, on three vowels, across their entire vocal range. Previous studies involving 3D MRI have not included multiple vowels or the full range of pitches.
- Perceptual study on the perception of different methods of resonance tuning in synthesised soprano voices by listeners. This is inspired by a similar study involving synthesised male voices.

Chapter 2

Literature Review : Voice production and Analysis

This chapter provides an overview of voice production, covering the respiratory system (the lungs, chest wall and diaphragm), the phonatory system (larynx or vocal folds), and the articulatory system (the vocal tract). It also discusses the various theoretical models available for predicting the behaviours of the vocal folds and vocal tract.

This thesis focusses specifically on the female singing voice, so special attention is paid to the specific methods of voice production employed by female singers, and in particular the technique of *resonance tuning*.

The literature review was conducted using online resources such as researchgate [16] and google scholar [17], as well as books and papers on voice production (including exploring the bibliographies of these works). Key terms used for searches included: "female voice", "female singing", "resonance tuning", "formant tuning", and "MRI of voice".

2.1 Voice Production

Voice production is a highly complex process that requires many precise movements of a large number of muscles to be executed in a particular way. Sound produced by the human voice can either be *voiced* (in which the vocal folds are vibrating) or *unvoiced*, and may or may not include a *fricative* component (turbulent noise). The voice apparatus can be divided into three major subsystems according to Kent and Read [18]. These are the respiratory system (the lungs, chest wall and diaphragm), the phonatory system (larynx or vocal folds), and the articulatory system (the vocal tract), and are shown in Figure 2.1

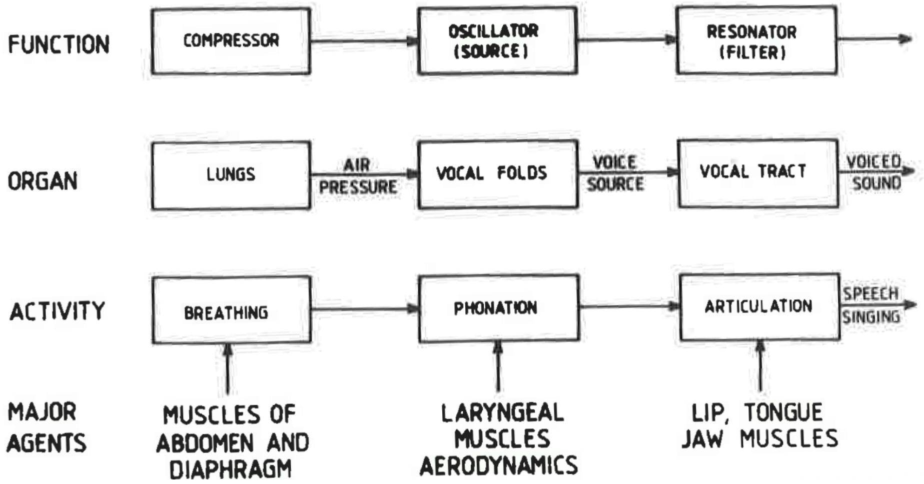


Figure 2.1: The three major subsystems involved in voice production, from [19].

The respiratory system is responsible for breathing, and can be thought of as the “power source” of the voice [19], pushing a steady stream of air through the phonatory and articulatory subsystems.

In the next part of the system this stream of air is then converted into harmonic sound (sound containing frequency components of integer multiples). This is achieved by the phonatory subsystem, where the vocal folds open and close repeatedly to regulate the airflow and convert the stream of air into harmonic sound.

A measure of the air flow through the glottis (the gap between the vocal folds) is shown in Figure 2.2. The cycle repeats over its period T, whereby the flow is at its maximum when the glottis is the most open, and minimum when the glottis is the most closed.

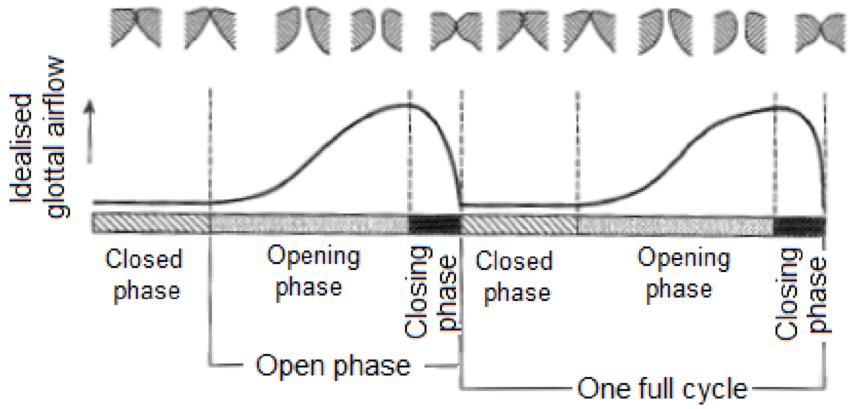


Figure 2.2: Shows a graph of the measure of the air flow through the glottis as a function of time, and the corresponding position of the vocal folds. (Reproduced with permission from [20])

A useful parameter of the glottal cycle is the open quotient (OQ), a calculation of the fraction of the cycle during which the glottis is open, as it can indicate the laryngeal mechanism used [21]. Conversely the closed quotient (CQ) is the fraction of the cycle during which the glottis is closed. A low CQ produces a *breathy* voice quality, whereas a high CQ can result in a *pressed* quality [22].

The spectrum of the signal produced by the glottis is approximated in Figure 2.3. Note that it consists of equally-spaced frequency components, with an overall sloping envelope shape of approximately -12dB per octave [23].

The final part of the voice production system is the articulatory system; the vocal tract approximates a tube (from the glottis to the lips) (see Figure 2.4), however, its shape is not fixed, due to the range of possible positions of the tongue, soft palate, lips, and many other smaller parts. Moving these components alters the shape of the vocal tract, so that when a signal (in this

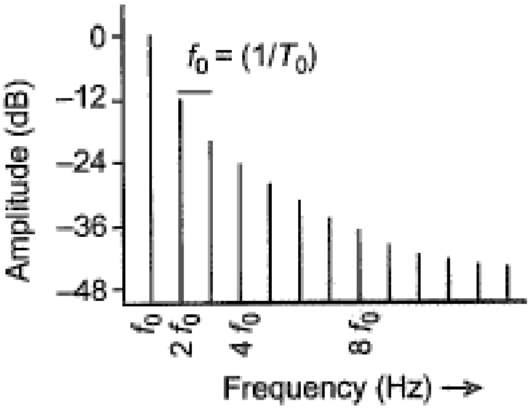


Figure 2.3: The spectrum of the glottal signal produced, consisting of the fundamental frequency and integer multiples thereof. (Reproduced with permission from [20])

case the sound produced by the glottis) is passed through it, the different frequency components are amplified or attenuated according to the properties of the vocal tract, changing the quality of the sound produced [19, 24]. Different vowel sounds, tones and timbres are produced by the filtering of the sound source by the moving vocal tract.

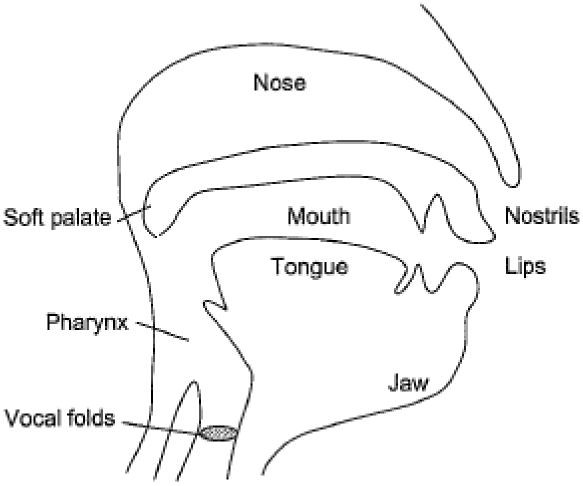


Figure 2.4: The basic structure of the vocal tract. (Reproduced with permission from [20])

2.1.1 The source-filter model

The previous descriptions of the effects of the vocal tract on the glottal source are based on the so-called source-filter model [25], which is a set of assumptions and conditions used in speech processing and synthesis, allowing the individual subsystems of the human voice to be processed and manipulated separately and independently. From this, detailed models of the glottal source have been created such as the Liljencrants-Fant Model (LF model) [26], which is based on two assumptions: linearity of the vocal tract, and independence of the voice source and vocal tract.

If the vocal tract is assumed to be a linear acoustic system, the properties of superposition and scaling apply, so the relationship between the input and output signals are known and can be modelled mathematically. A hard-walled system of tubes without sharp bends, tight constrictions or sharp projections is a linear acoustic system for sounds of reasonable amplitudes (within the dynamic range of the human voice). The vocal tract can therefore be considered as a linear system, if vibration of the softer walls is neglected. For fricative sound production, the constrictions used to produce turbulent air flow mean that the system can no longer be classed as linear, however for vowel sounds this is not an issue.

The assumption of independence between the voice source and the vocal tract can be examined in two ways; the acoustic interaction [27], which refers to how the acoustic properties of the vocal tract affect the airflow at the glottis, and the physiological interaction, which refers to how the position of the vocal tract affects the vibration of the vocal folds [28].

The ability to model the behaviour of the voice source and vocal tract separately, makes it relatively easy to reproduce them artificially, which can aid our understanding of voice production. Changing the parameters of the LF model [29] changes the properties of the voice source, and can be used to produce different qualities in the voice, for example *breathy*, *pressed*, or *ringing* voice.

It is possible to extract information on the glottal flow as a function of time $g(t)$, its time derivative $\frac{dg}{dt}$, and information about the effects of the vocal tract, from the acoustic output of the voice using inverse filtering techniques. This involves taking a speech signal and using adaptive filtering techniques where the error is repeatedly calculated and minimised, to obtain accurate models for both the glottal flow model and the vocal tract effects [30].

In many cases, since the source-filter model allows the voice source and vocal tract to be examined independently, it can be assumed correct, and provides a useful tool for voice analsis. In truth, however, the assumption that they can be treated independently is not entirely correct, as there is some acoustic and physiological interaction between them [31]. This interaction is not yet fully understood, and so the source-filter model will be assumed to be largely accurate for the purposes of this research, as it provides a convenient platform for analysing the vocal tract effects without taking into account the behaviour of the voice source.

2.1.2 The Sound Source

Having established the three subsystems that allow voiced sound to be produced, this section examines the phonatory subsystem in more detail. The different pitches that can be produced by the larynx are addressed, as well as methods of altering the quality of this sound (in isolation from the articulatory subsystem).

The vocal folds are fleshy folds inside the larynx, which have a complex structure. They adduct (close) and abduct (open) very rapidly, the number of cycles per second corresponding to the fundamental frequency (f_0) of the sound source.

One of the earliest and simplest explanations of vocal fold vibration relies on the Bernoulli force [19]. This is the force generated when a stream of fluid passes around an object, so that some layers of the stream are forced

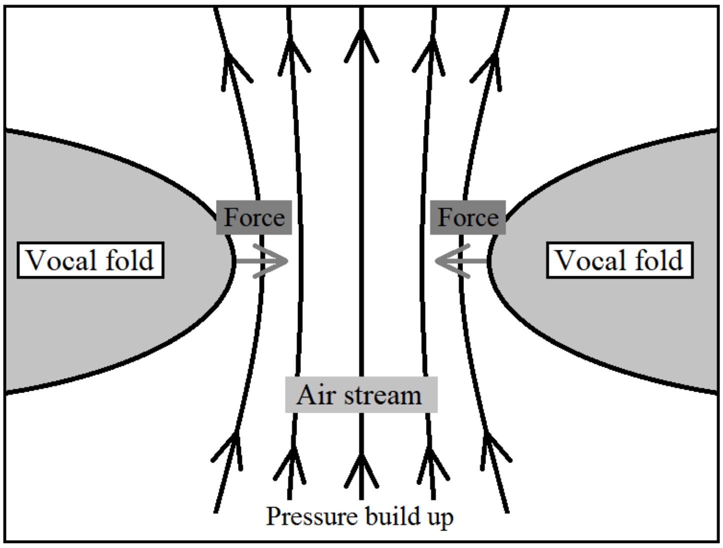


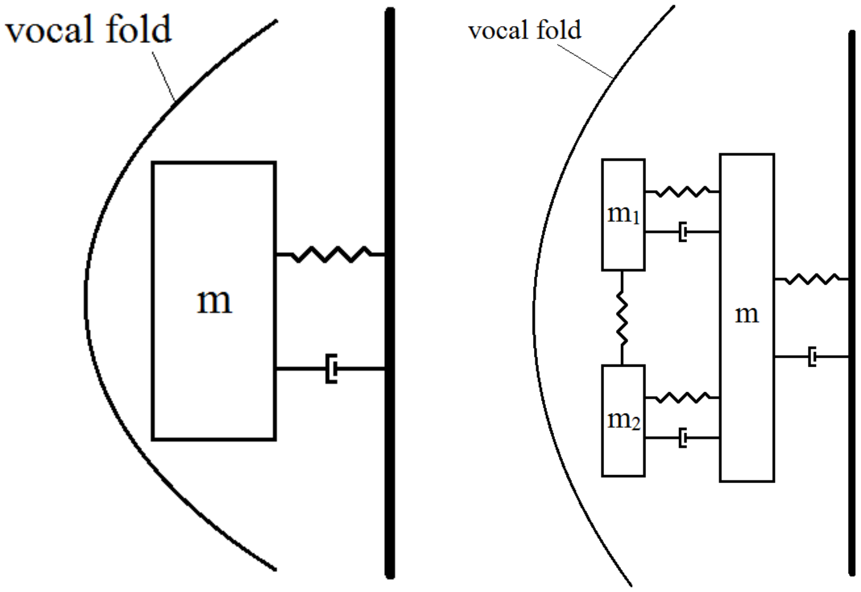
Figure 2.5: Shows how the pressure difference between layers of air creates the Bernoulli force, pulling the vocal folds together.

to travel a longer distance than the others. Since the layer that travels the furthest from the object must cover a larger distance than the closer layers, it has a greater velocity and so a lower pressure relative to the other layers. This creates a pressure difference between layers, exerting a force on the object which pulls it into the stream. In the case of a stream of air passing over the vocal folds, the layers of air closest to the folds (at the edges of the stream) have to travel the furthest, so they have the lowest pressure and the vocal folds are pulled towards the centre of the glottis, closing it. This is illustrated in Figure 2.5, where the horizontal arrows indicate the force on the vocal folds.

Once the vocal folds have been closed, the pressure builds up below them, eventually forcing them open and allowing a stream of air to pass through. At this point the elasticity of the vocal fold tissue limits how far outwards they are pushed, and pulls them back to a neutral position, ready for the whole cycle to start again. This is known as the *myoelastic aerodynamic* theory of phonation [32].

This model was previously accepted by theorists [33]; however, it does not provide a full explanation for the mechanism by which the vocal folds vibrate, as the damping effect of the vocal fold tissue would soon cause the vocal folds to stop moving: the air stream would not be able to provide enough energy to sustain vibration. The *neurochronaxic theory* states that the method by which the vocal folds produce sound is the rapid movement of muscles to control the adduction (closing) and abduction (opening) of the vocal folds, however, repeating this cycle at the frequency required for a female singer (up to 1000 times per second) is beyond the capability of these muscles [19, 34].

More recent models [35] therefore take into account another component; an acoustic tube to represent the vocal tract. When the vocal folds are pulled together by Bernoulli forces, the air stream continues to move up the tube, creating a low pressure area just above the glottis. This low pressure area, combined with the increased pressure below the glottis (from the air stream continuing to flow up from the lungs) helps to pull the glottis open from above while it is also pushed open from below. This is known as the *one-mass model*, and was proposed by Flanagan and Landgraf in 1968 [36]. It modelled vocal fold vibration with a single mass-spring oscillator driven by airflow from the lungs, and produced self-sustained oscillations as long as the vocal tract load was inertive, unlike the myoelastic-aerodynamic model (without the tube). However, it is still not an entirely accurate representation of vocal fold vibration. The one-mass model is shown in Figure 2.6(a), with a spring to model the elastic properties of a vocal fold and a damper to model the absorption of the tissue.



(a) The one mass model of the vocal folds with a spring and damping component (after Titze, [37])

(b) The two mass model of the vocal folds, which models the behaviour of both the muscle and covering layer of the vocal folds.

Figure 2.6: The one and two-mass models of the vocal folds.

The main issue with the one-mass model is that the vocal folds do not move as single masses. Observing the movement of the vocal folds during phonation reveals that they move in a wave-like motion, with the bottom edge opening first and closing first (vertical phase difference) [38]. To allow for this wave-like motion, the *body-cover* or *two-mass* model was proposed by Ishazaka and Flanagan in 1972 (see Figure 2.6(b)) [39], which consisted of a large mass representing the vocalis muscle (main body of the vocal fold, see Figure 2.8) and two smaller masses, representing the outer layers of the vocal folds. This model was able to sustain oscillation with or without a vocal tract and provided the degrees of freedom necessary to produce the vertical phase difference. This model has been widely used as a simple, low-dimensional model of the vocal folds, although models involving up to 16 masses are in use [40].

Vocal folds in Singing

In singing, being able to accurately sing at the desired pitch is of paramount importance. According to Sundberg [19], the fundamental frequency of the sound produced by the vocal folds is affected by two major factors. The most significant of these is the *laryngeal musculature*, or more specifically, the length of the vocal folds; the more they are stretched, the longer, thinner and tenser they become and the higher the frequency produced.

Altering the length of the vocal folds is achieved by the cricothyroid muscle contracting and tilting the cricoid cartilage (see Figure 2.7), which decreases the distance between the thyroid and cricoid cartilages, and increases the distance between the thyroid and arytenoid cartilages [19].

In addition to the length of the vocal folds, the fundamental frequency of the note can also be affected by the pressure in the lungs. It has been shown by Van den Berg et al. [33], that increasing subglottal pressure, in addition to increasing the intensity of phonation can cause a small increase in phonation frequency, which must be actively controlled by singers in order to preserve constant f_0. Other factors are also thought to contribute, for example Sataloff [32] adds that contracting the cricothyroid muscle can also increase the f_0 of the note produced, by increasing the stiffness of the vocal fold cover. A complex system, the length of the vocal folds is the most influential mechanism for controlling the phonation frequency.

There are many factors that can affect the behaviour of the vocal folds; professional singers reduce the breathiness of their voice by increasing glottal closure [41] which in turn increases the vocal efficiency and the volume of sound produced. Commonly in training, a “trial and error” approach is used, or exercises such as lip trills, tongue trills, bilabial fricatives, humming, and phonation into tubes or straws [28], which are intended to increase the vocal tract interaction with the source, and improve the vocal economy [42].

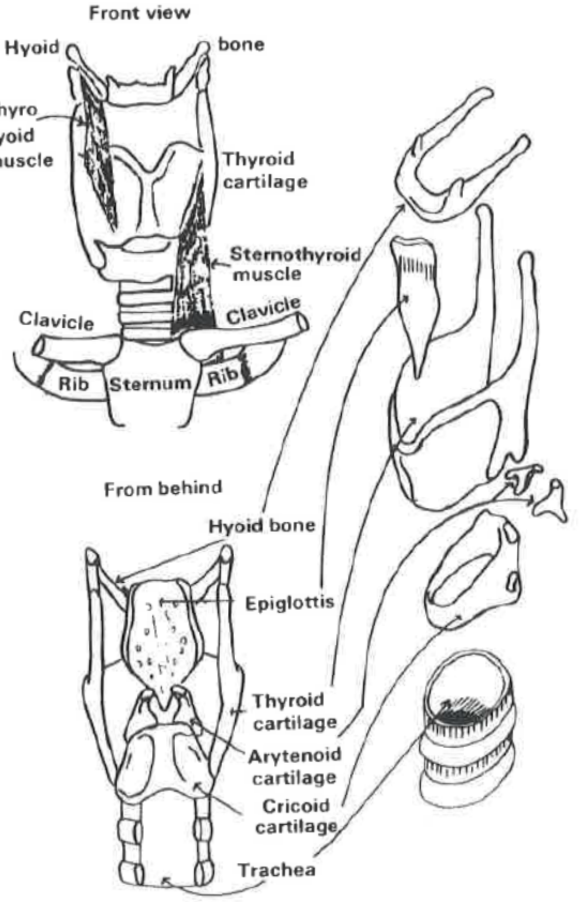


Figure 2.7: The detailed inner structure of the larynx (from [19]).

Voice quality

The exact properties of each individual larynx and vocal tract are unique, and therefore the precise qualities of each voice are particular to the singer, including for example, the extent and duration of vocal fold closure. If the vocal folds do not close fully, a gap in the vocal folds called a *chink* can remain, the shape of which can vary between subjects. Poor closure can result in the voice having a *breathy* or *airy* tone and is generally discouraged in classical singing training [43, p. 41].

Females tend to have a more "breathy" voice than males. The causes of this are thought to be both psychological and (particularly in females) physical

[44].

Varying the use of the vocal folds produces different effects in singing. One example is *pressed* voice, which is thought to occur as a result of excessive adduction of the vocal folds, and is characterised by high harmonic content (low spectral slope). Another is *ring*, which is thought to be related to the area of the epilaryngeal tube, skewing quotient and possibly the fundamental frequency, and is defined by perceptual experiment by Bergan [45] as a boost of energy around 3000 Hz, which is acoustically linked to clustering of F_3-F_5. "Ring" is however a highly ambiguous term, with Howard et al. [46] describing a voice with ring as having more energy in the "region around 4 kHz, which is more than 1000 Hz higher than what is observed in [male] adults; and in the region around 7.5 - 11 kHz".

The behaviour of the vocal folds can also be affected adversely by factors beyond the singer's control. For example, dehydration has been found to increase the phonation threshold pressure - the minimum pressure the lungs must produce to make the vocal folds produce a pitched sound [47].

Differences in vocal fold vibratory characteristics are also observed between sexes. Sulter et al. [41] found that females were more likely to have a *chink* in their glottal closure, exhibit horizontal phase differences (the back of the glottis opening and closing before the front) and have a larger amplitude of vibration. They also observed a slight difference between trained and untrained singers, with the trained singers exhibiting more complete glottal closure. This supports Howard's [22] findings for a group of 26 singers singing at high pitches, where trained singers were found to have higher closed quotient than untrained singers for pitches above 494 Hz (B4), however, the untrained singers were found to have higher closed quotient at pitches below 294 Hz (D4). This study also found that the gradient [CQ/log(f_0)] tended to correlate positively with the number of years singing training/experience. Barlow et al. [48] also observed differences in CQ between singers singing in different genres, with singers exhibiting higher glottal closure for musical theatre styles than classical.

The length of the vocal folds increases with age [49], and generally reach a larger size in males. The length of vocal folds can contribute to a singer's voice classification, with Roers et al. [50] finding mean vocal fold lengths of 14.9, 16.0, 16.6, 18.4, 19.5 and 20.9 mm for sopranos, mezzo-sopranos, altos, tenors, baritones and basses, respectively.

Registers

The vocal folds consist of several layers of muscle (see Figure 2.8), which allows the vocal folds to vibrate with different masses, depending on which layers are vibrating. The transitions between laryngeal vibratory mechanisms are commonly marked by pitch "breaks" and changes in the quality of note produced.

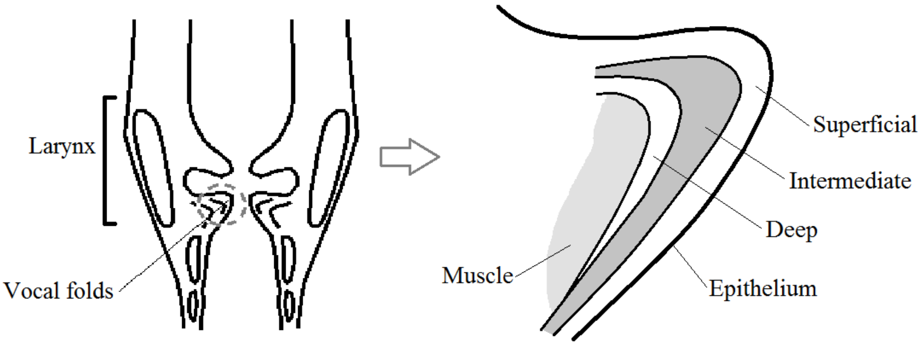


Figure 2.8: The structure of the vocal folds. (After [51])

In 1880 this was seen by the physiologist and voice production teacher Emil Behnke and the throat surgeon Lennox Browne using a laryngoscope to view the vibrating glottis [52]. They noticed that:

> *"During the lowest series of tones the vocal ligaments vibrated in their entire thickness. During the next series of tones the vocal ligaments vibrate only with their thin inner edges, and during the highest series of tones a portion of the vocal chink is firmly closed, and only a small part of the vocal ligaments vibrates." ([52], p86)*

These different modes of glottal vibration correspond to different singing *registers*, defined by Garcia [1] as:

> *"A series of consecutive and homogeneous tones going from low to high, produced by the same mechanical principle, and whose nature differs essentially from another series of tones equally consecutive and homogeneous produced by another mechanical principle. All the tones belonging to the same register are consequently of the same nature, whatever may be the modifications of timbre or of the force to which one subjects them." ([53], p68)*

Garcia claimed that the human voice is composed of different registers: *poitrine* (chest), *fausset* (falsetto), *head* (tête), and *contre-basse* (counter bass) [54]. There is, however, considerable disagreement on the terminology and definition of registers, with different naming schemes proposed by various authors. For example Behnke [52] described three registers for the male voice (*lower thick, upper thick and upper thin*) and five registers for the female voice (*lower thick, upper thick, lower thin, upper thin and small*), based on his observations of the vocal folds.

Roubeau et al. [55] detected four laryngeal mechanisms using electroglottography. Representing the lowest f_0[1], M0 is also called the *creak* voice or *vocal fry*. The vocal folds vibrate aperiodically and this produces a broad band spectrum with no clear pitch. Slightly higher in f_0 is mechanism M1, known as the *modal* voice or *chest* voice, which is usually used for speech. Vocal fold vibration is highly periodic and the relatively rapid closure of the vocal folds results in a spectrum rich in harmonics. Mechanism M2 produces higher f_0s, and is known as the *falsetto* voice in men and *head* voice in women, in which only the ligament and mucosa covering the vocalis muscle vibrate. Vocal fold vibration is still highly periodic, but the higher harmonics are usually rather weaker than those of M1. The highest singing register is M3, used mainly by sopranos specialising in the highest ranges and typically starting somewhere

---

[1] Since the M0 register can involve aperiodic phonation it does not necessarily have a f_0.

near 1046 Hz (C6). It is known as the *whistle* register, and the physiology and acoustics of it are not yet well understood [56].

Sopranos generally transition from laryngeal mechanism M1 to M2 at around 340 Hz (∼E4–F4) [9, 55, 57]. The next transition in the soprano range occurs at around 500–700 Hz (∼C5–G5), and is thought to be related to vocal tract tuning [57], as this range corresponds approximately to the pitch range where the fundamental frequency is equal to or higher than the first vocal tract resonance, above which sopranos tune the frequency of their first vocal tract resonance (R_1) to the fundamental frequency [13, 58, 59]. A third transition is commonly reported in the top range of the soprano voice, occurring somewhere in the broad range 660 Hz (E5) to 1570 Hz (G6) [52, 55]. This break is associated with a transition to the highest vocal register, commonly known as the whistle register (also known as flageolet, flute, bell, small, and pipe [55]). In this very high range, some studies [9, 55] report significant differences in laryngeal behaviour compared to the M2 laryngeal mechanism, as the vocal folds are thin and more tensed, and have a smaller amplitude of vibration. The vocal fold contact is also reduced, and in some cases there may be no contact at all. This is based on electroglottograph data [55], as well as inspection via endoscope [9].

2.1.3 The Vocal Tract

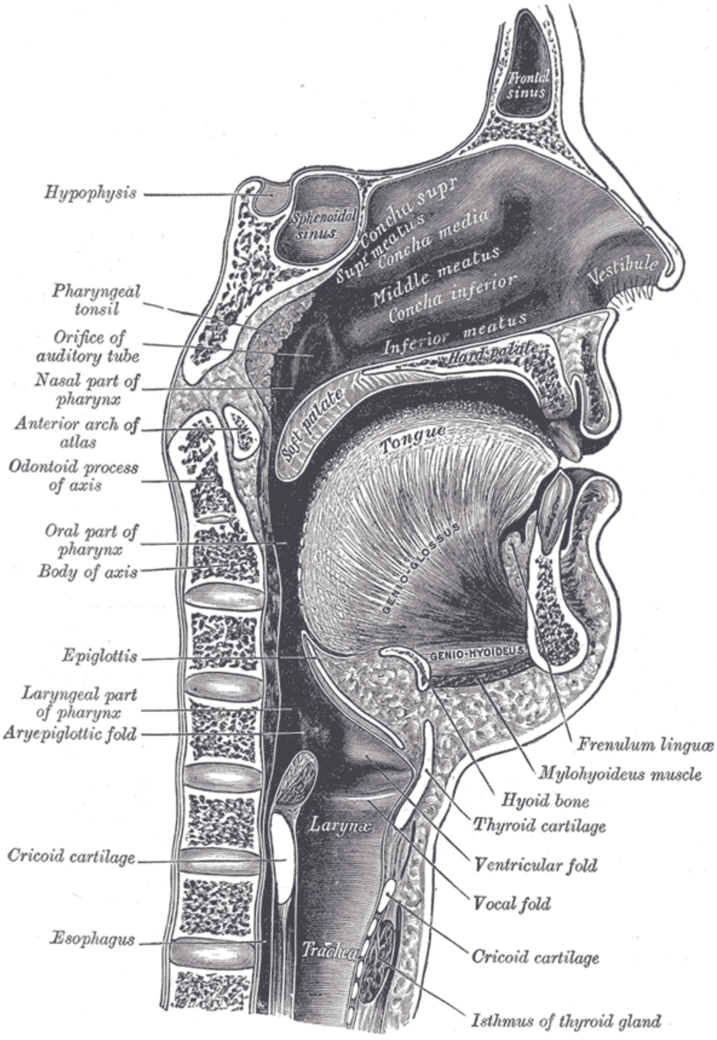


Figure 2.9: The detailed structure of the vocal tract. [60]

The third subsystem; the articulatory system, consists of the vocal tract, which is roughly the shape of a cylindrical tube open at one end (the mouth) and closed at the other (the glottis).

The vocal tract begins at the glottis; the air gap between the vocal folds, which is situated within the larynx. The back of the throat is known as the *pharynx* and is made up of three parts - the lower part is the *hypopharynx*, the middle part is the *oropharynx*, and the upper part (inside the nasal cavity) is the *nasopharynx*, which is separated from the oral cavity by the *soft palate*. The oral cavity is bounded by the *hard palate* at the top, the tongue at the bottom, and the teeth or cheeks at the sides. Finally the lips control the entrance to the mouth, and the jaw can alter the shape of the whole vocal tract.

The acoustic behaviour of the vocal tract is dominated by its tube-like properties. A rigid pipe, open at one end, will produce reflections of the acoustic wave from the closed end. This causes standing waves to form, with a velocity node (minimum) at the closed end, and a velocity antinode (maximum) at the open end. The lowest resonance of the tube is the frequency at which a quarter of a wavelength is equal to the length of the tube (this type of tube is also known as a quarter-wave resonator). Multiples of the fundamental frequency are also generated, as shown in Figure 2.10 [61].

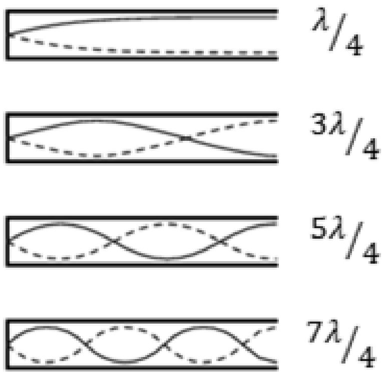


Figure 2.10: The velocity of air moving in a rigid tube open at one end for the first four resonances. [62]

An equation can be derived for the relationship between the wavelengths of the resonant frequencies of a simple tube, open at one end, and the length of the tube:

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

Where L is the length of the tube, and n is an integer.

This rearranges to give the equation for the tube's resonant frequencies:

$$f_n = \frac{(2n-1)c}{4L} \tag{2.2}$$

If the frequency response of a rigid tube is plotted, it can be seen that there are peaks in amplitude at the resonant frequencies, which occur at regular intervals (see Figure 2.11). A plot of this nature is known as a *transfer function*, and shows the ratio of the output of a system to its input, as a function of frequency.

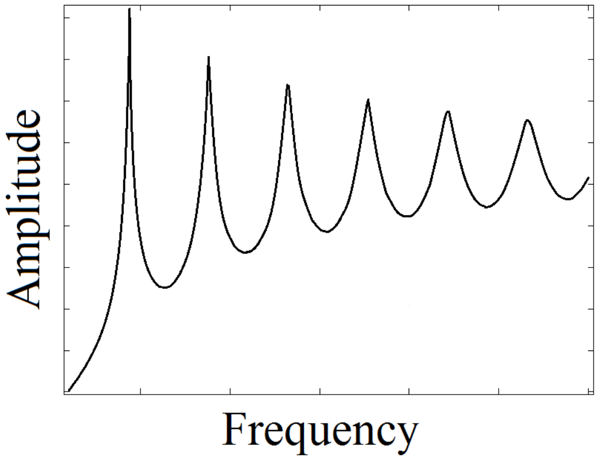


Figure 2.11: Plot of frequency response of a rigid tube, closed at one end and open at the other. The resonances are regularly spaced and of consistently-decreasing amplitudes.

One of the key properties of the vocal tract (compared to a simple rigid tube) when considering resonances, is that the cross-sectional area of the vocal tract can be increased or reduced at different points along its length. Constriction of the vocal tract at a place where the standing wave of a resonance exhibits minimum-amplitude pressure causes that resonance to drop in frequency, whereas expansion of the tract at those same places raises the frequency of the relevant resonance [63]. This allows singers to shift the frequencies of the resonances of the vocal tract by altering parts of the tract known as *articulators* or *sound modifiers*: the soft palate, jaw, tongue, etc.

Similar to a simple rigid tube, the frequency response of the vocal tract also shows peaks in amplitude at the resonant frequencies of the vocal tract (see Figure 2.12), however, unlike the simple tube, their spacing is not necessarily regular. The vocal tract resonances form the envelope of the voice signal, with the fundamental frequency and its harmonics as the carrier function.

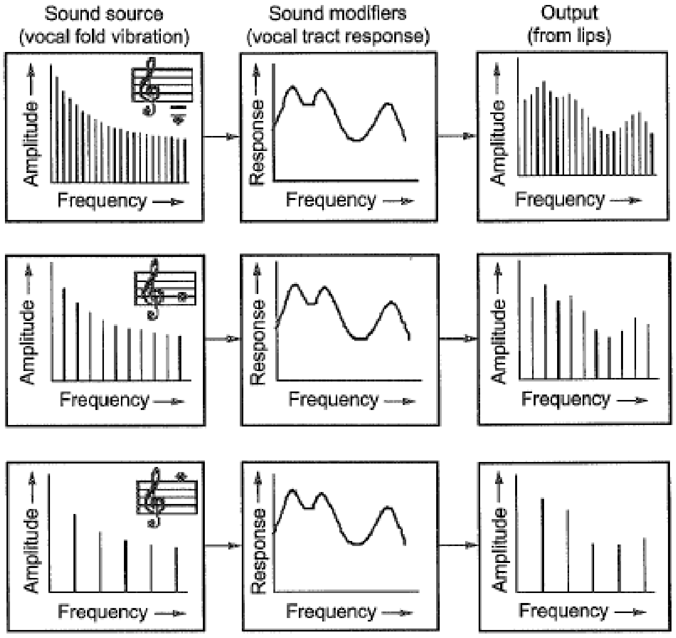


Figure 2.12: Shows how a vowel is produced on three different fundamental frequencies (top - lowest, bottom - highest). The left plot shows the glottal signal, the middle plot shows the transfer function of the vocal tract, and the right plot shows the resulting acoustic output. (Reproduced with permission from [20])

The series of broad peaks in the voice spectrum are known as vowel formants (F_n), defined by Fant [64] as "spectral peaks of the sound spectrum of the voice" and arise from the effect of the resonances of the vocal tract (R_n) on the glottal source spectrum [58]. There is some confusion in later literature as to the distinction between the terms *formant* and (vocal tract) *resonance.* In this thesis, a *formant* is defined as a broad peak in the spectrum of an acoustic signal, arising from a resonance of the vocal tract, whereas a *vocal tract resonance* is a resonant frequency of the vocal tract. In older works, the term formant has been used to mean both formant and vocal tract resonance interchangeably [65].

It is generally accepted that the first 3-5 formants are responsible for the perception of vowels [66] and their frequencies determine which vowel is perceived by a listener. The formant frequencies can be changed through movements of the vocal tract to give changes in the accent, timbre, quality or even the vowel perceived. For example, movements of the jaw have been found to alter the position of the first vocal tract resonance [67], whilst movements of the body of the tongue alter the frequency of the second vocal tract resonance, and the position of the tip of the tongue affects the position of the third vocal tract resonance [63].

2.2 Analysis of The Voice

Advances in computing and research technology means that there are many different methods of analysing the human voice and these can be divided into three broad categories; methods that analyse the *acoustic output* of the voice, methods of analysing the *voice source* (larynx and vocal folds) and methods that investigate the properties of the *vocal tract* and any physiological changes made to it.

2.2.1 Analysis of the Acoustic Output

Fourier Transforms

The Fourier series allows any waveform to be synthesised by adding sine waves of integer multiples of a fundamental frequency. A Fourier analysis is the inverse process, which examines a harmonic signal and decomposes it into its frequency components. This can be carried out on a sound recording to examine the spectrum (frequency content) of the sound. In voiced sounds (when the vocal folds are phonating), this allows the harmonics and their relative amplitudes to be detected. In low voices (below approx. 300 Hz), when the harmonics are closely spaced, the positions of the formants can be detected by simply examining the envelope shape of the acoustic spectrum.

Examples of this can be seen in Figure 2.13, which shows a spectrum of a male voice singing an /ɑ/ vowel (a), and a female voice singing the same vowel (b) (produced using a 2^{14}-point Fourier transform in MATLAB [68]).

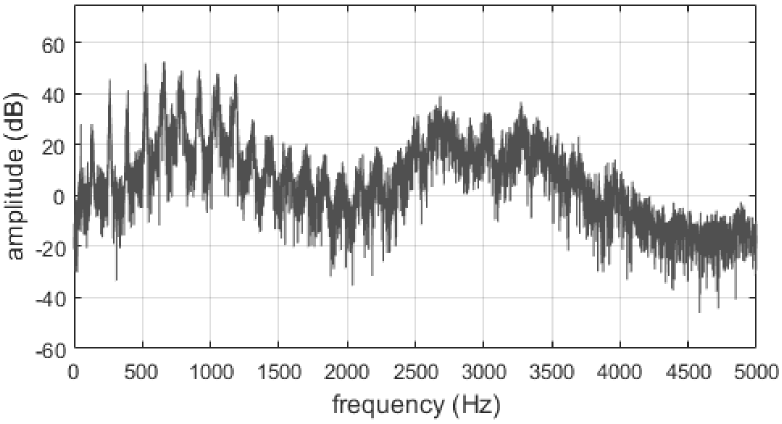
Fourier transforms are performed over a "window" of a defined number of samples. A short (time) window gives good temporal resolution, but poor frequency resolution, whereas a long window gives a good frequency resolution, but poor temporal resolution. Choosing the window length for the Fourier transform is therefore a compromise; the value chosen in this example gives adequate frequency and time resolution.

Spectrograms

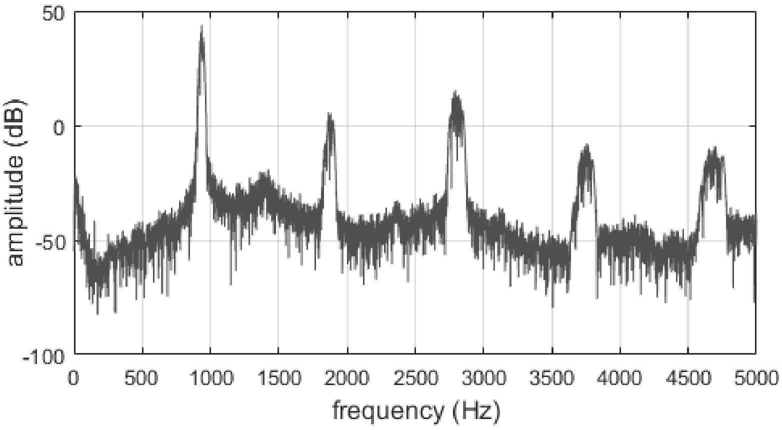
A spectrogram is an extension of the Fourier transform [69], with the added dimension of time. Higher amplitudes are represented by darker shades, so that not only can the frequency components of a single sample be observed, but also their changing behaviour over time. An example spectrogram created in Praat [70] of a female saying "the north wind and the sun were arguing about which of them was the strongest" is shown in Figure 2.14(a) (window length 0.1 seconds, frequency step 20 Hz).

Long-term average spectra

A long-term average spectrum (LTAS) displays the average sound level over time in different frequency bands [71] and can give information on quasi-constant glottal and vocal tract characteristics, such as the singer's formant or speaker's formant [72]. It is often applied to running speech (as in Löfqvist, [73]); however, it also has applications in singing.

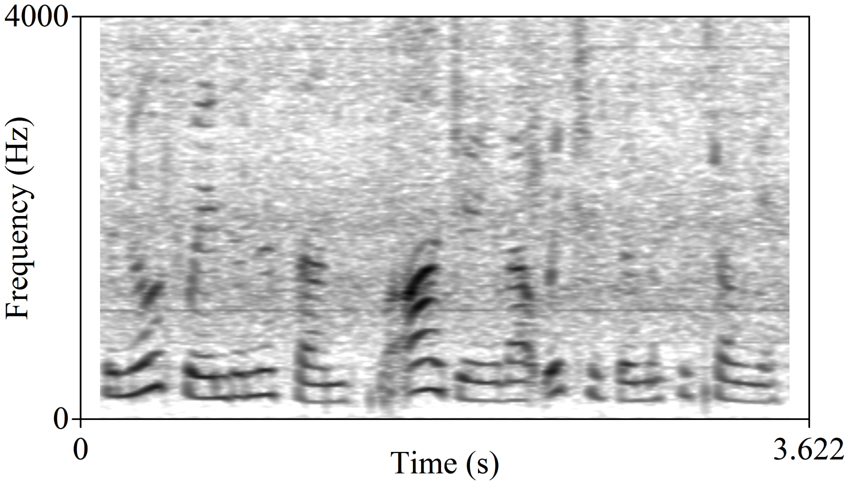


(a) Spectrum of a male voice singing an /ɑ/ vowel [62], produced using a 2^{14}-point FFT.

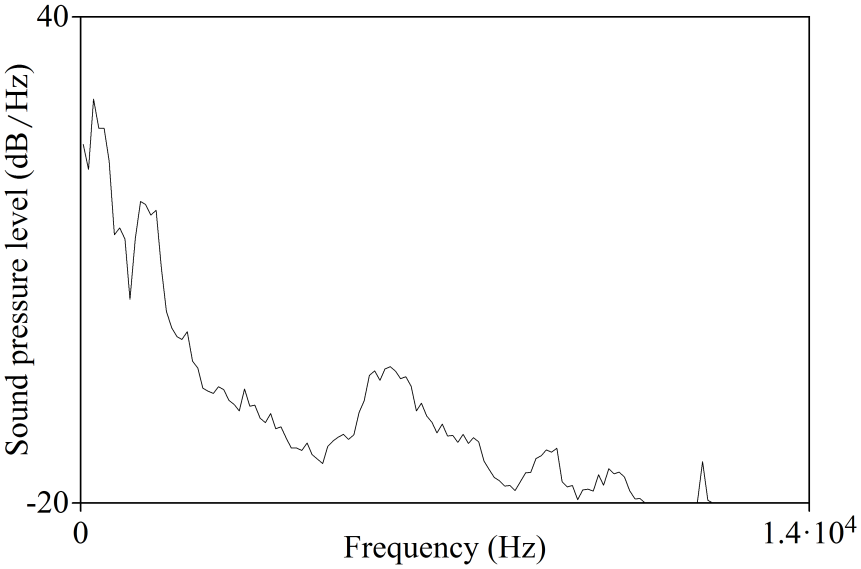


(b) Spectrum of a female voice singing an /ɑ/ vowel (subject 1 from Chapter 5), produced using a 2^{14}-point FFT.

Figure 2.13: The different spectra of singing produced by a low (male) voice and a high (female) voice.



(a) Spectrogram of a female saying “the north wind and the sun were arguing about which of them was the strongest”.



(b) LTAS of a female saying “the north wind and the sun were arguing about which of them was the strongest”.

Figure 2.14: Shows two different examples of analysis of the acoustic output of the voice.

A number of methods have been suggested for quantifying LTAS information, for example by calculating the ratio of power in different frequency regions. The *singing power ratio* (SPR) is calculated by subtracting the amplitude of the strongest harmonic between 2 and 4 kHz from the level of the strongest harmonic between 0 and 2 kHz, and was found by Watts et al. [74] to be significantly higher in non-talented singers than talented singers.

The occurrence of LTAS peaks has been found to be related to voice quality [75, 76], and the sex of the speaker/singer, for example Mendoza et al., observed differences between male and female voices, including differences in aspiration noise and spectral tilt [77]. White et al. [78] also observed differences in male and female child voices using LTAS.

Figure 2.14(b) shows the LTAS for the same audio sample as 2.14(a), created in Praat [70] (bandwidth 100 Hz).

Linear Prediction

Linear prediction [62, 79] is a mathematical technique where future values of a discrete-time signal are predicted as a linear function of a number of previous (known) samples. The most common mathematical representation (from [79]) is:

$$\hat{s}(n) = -\sum_{i=1}^{M} a_i s(n-i) \tag{2.3}$$

Where $\hat{s}(n)$ is the predicted signal value, $s(n-i)$ are the previous observed values, and a_i are the predictor coefficients to be found. The *order* of the linear predictor (M) refers to the number of previous samples that are linearly combined. The minus sign is chosen so that the error is based on a *difference* of two variables [79].

This allows an error to be generated (where s is the actual value of the signal), which becomes the driving function of the linear prediction:

$$e(n) = s(n) - \hat{s}(n) \tag{2.4}$$

The coefficients are chosen to minimise the error between the sample and its predicted value. The solution to the model is found using auto-covariance or auto-correlation, and can be used to describe the original signal using an all-pole filter, or the inverse filter with an all-zero filter.

It is possible to use linear predictive coding (LPC) to estimate the envelope shape of a voice spectrum [80], and once this is obtained, the positions of the formants can be found in two possible ways. The simplest is to use a peak-picker algorithm, which simply finds local maxima, however this only works if all the peaks are true maxima in the envelope shape, and not just points of inflection, so it might occasionally fail to detect peaks that are close together or not high in amplitude. The other method is to use a root-finder, such as the MATLAB "root" function [81], which is slightly more complex, but much more reliable.

```
 % calculate LPC coefficients from signal (12th-order LPC):
[a,g] = lpc(x(:,1),12)
% take FFT of zero-padded coefficients:
Temp = fft([a zeros(1,length(x)-length(a))])
% calculate envelope spectrum:
filter_spectrum=20*log10(abs(1./Temp))-10*log10(g)
```

Figure 2.15 shows the same spectrum as Figure 2.13(a), with the envelope spectrum found using the LPC method (dotted line) and the formant values detected (asterisks). In this case the fifth formant is detected by the root solving method, but not by the peak-picking method, due to the lack of a distinct peak in the envelope function.

Difficulties analysing high voices

As demonstrated in Figure 2.13, the acoustic spectra of low and high voices show distinct differences, leading to several issues with the analysis of very high (soprano) voices. The main problem for researchers stems from the fact that since the fundamental frequency is high (soprano range approx. C4 - C6, or 261 - 1046 Hz), the harmonics are widely spaced and it can be difficult

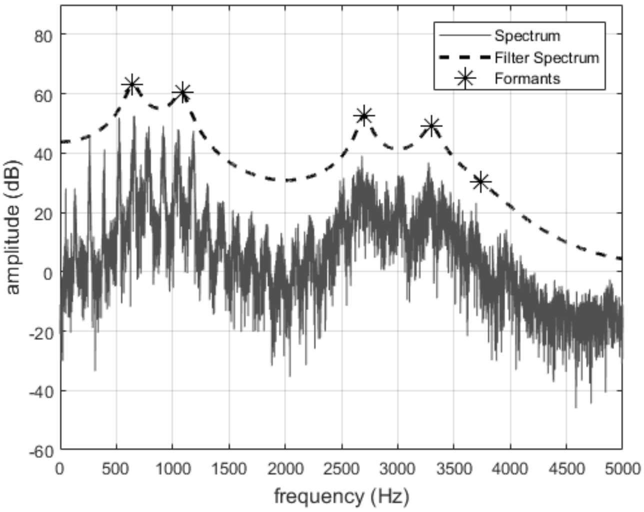


Figure 2.15: The spectrum of a male voice singing an /ɒ/ vowel [62], with the envelope spectrum found using LPC

to detect the positions of the formants accurately using the usual methods of speech analysis such as spectrography, linear predictive coding or (polynomial) root solving, which are generally highly accurate at low frequencies [82]. Neither spectral analysis nor linear prediction (popular in speech analysis) are reliable for detecting resonances at fundamental frequencies above approximately 350 Hz [83].

This difficulty in determining the locations of the formants spectrographically is also reflected perceptually in the difficulty of perceiving vowels accurately at high pitches, with vowels becoming harder to identify with increasing fundamental frequency (discussed further in section 2.4.2).

2.2.2 Analysis of the Voice Source

Analysis of the voice source (separate from the effects of the vocal tract) allows mathematical models of the vocal folds to be created, so that the mechanics of their vibration can be better understood. The vocal folds can also be directly observed by inserting a laryngoscope or endoscope into a singer's nose or mouth to watch the vocal folds vibrating. In 1880 Behnke and Browne [52] used a laryngoscope, a small mirror on a rod, to view the vocal folds vibrating and observed the differences between different laryngeal mechanisms. Modern methods allow more detailed information to be obtained, and involve the use of an endoscope, a small camera on the end of a flexible tube, in conjunction with a stroboscope [41], to observe the exact vibratory patterns of the vocal folds. Alternatively, a high-speed camera could be used and the recordings then viewed in slow motion to obtain the same result. The disadvantage of this method is that it is an invasive and uncomfortable procedure and it may be difficult for the singer to sing naturally with their throat obstructed.

Properties of the vocal folds, such as the open or closed quotient can be monitored more easily than endoscopy or laryngoscopy using electrolaryngography. This technique was developed by Fabre in 1957 [84], and requires the singer to wear a pair of electrodes (or several pairs), attached using an elastic neck-band, then a low-intensity, high-frequency modulated current is passed through the electrodes, and the impedance between them is measured. The impedance is approximately proportional to the vocal fold contact area, and so this allows the vocal fold activity to be monitored. Changes in the laryngeal vibratory mechanisms can be detected by examining the electroglottograph (EGG) signal, and its derivative [55].

More pairs of electrodes allow more information to be obtained about the vocal folds. In 1992 a two-channel EGG was introduced by Rothenberg [85], which was also capable of measuring the height of the larynx. In 2009, Kob et al. proposed an EGG with 6 pairs of electrodes that would give access to EGG signals as well as a two-dimensional larynx tracking signal [86]. In 2012, Hezard et al. [87] proposed an alternative method called electrical impedance tomography (EIT), which consisted of "injecting a high-frequency current inside a body with two electrodes and measuring the resulting potential distribution (with other electrodes) at the body's surface". Although EIT offers the potential for more detailed observation of the vocal folds, it is not yet widely used.

2.2.3 Analysis of the Vocal Tract

The physiological properties of the vocal tract can be analysed using imaging techniques, such as X-ray or magnetic resonance imaging (MRI) data, or by observing the moving parts of the vocal tract using an articulograph. It is also possible to measure the resonances of the vocal tract using external excitation, such as noise excitation at the larynx or injecting a signal into the mouth.

X Ray

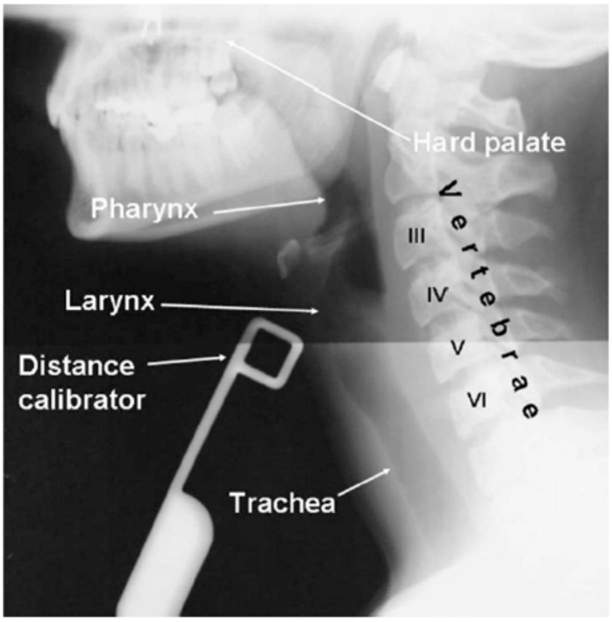


Figure 2.16: An example of an image of the vocal tract obtained using X-ray imaging (from [88]).

In 1951 MacMillan and Keleman X-rayed all standard Russian vowels and consonants (reported by Fant [25]), to examine the configuration of the vocal tract. Although pioneering at the time, the information about the vocal tract that can be obtained from X-rays of the vocal tract is limited, due to poor resolution of tissue and the fact that the images are a two-dimensional representation of a three-dimensional object. Added to this is the deterrent that repeated exposure to X-rays increases a person's risk of developing cancer [89].

Articulography

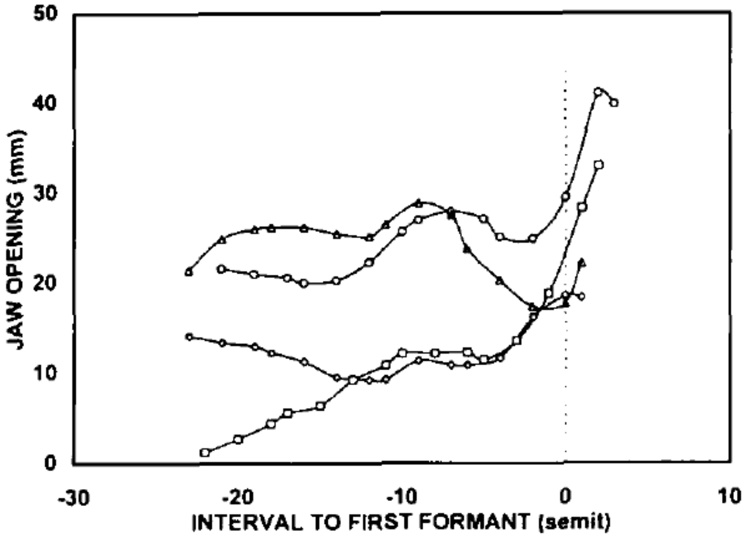


Figure 2.17: Jaw opening for an /a/ vowel, plotted as a function of the pitch interval in semitones to the formant value measured when the subject sang the vowel at a pitch located in the lower part of the pitch range. Different symbols refer to subjects. (from [67]).

One technique that does allow three-dimensional, dynamic data on the vocal tract to be gathered is the use of an articulograph. This requires small connector coils to be positioned on and in the subject's mouth. Each of the coils is a transmitter that produces an alternating magnetic field at different frequencies. This induces an alternating current in the sensors, and allows the user to obtain the distances of each sensor from the nine transmitters. It is then possible to calculate the location of each transmitter in 3 dimensions, and store and display the positions of the sensors [90]. This allows the researcher to obtain specific information about the movements of the tongue, mouth, soft palate, and jaw, which yields information about how they are adjusted to achieve resonance tuning. However, information about the vocal tract resonances cannot be gained in this way, as it does not yield enough information to create a 3D model. An example of articulograph data is shown in Figure 2.17: measurements of jaw opening against the pitch interval in semitones to the nearest formant for two different vowels (from [67]).

Acoustic excitation of the vocal tract

Measuring the frequencies of the vocal tract resonances (R_N) can give important information about the effect of the vocal tract on the voice source. Methods of directly measuring the vocal tract resonances generally involve using a known signal to excite the vocal tract at one end and recording the output from the mouth.

One method of extracting an impulse response of an acoustic system is excitation using a swept-sine signal [91]. This method is successfully used in room acoustics applications [92] and also has applications to voice science. In 1970 Fujimura and Lindqvist [93] measured the resonances of the vocal tract by applying a swept-sine signal to the neck of subjects, just above the glottis, and recording the output. The results obtained for vowel sounds were mostly reliable, and yielded detailed information on the locations and bandwidths of the vocal tract resonances. However, a problem with this method was that the subject was required to close their glottis while the measurements were taken, which a small number of subjects found difficult to do, leading to poor results. The effects of measuring through the neck are also unknown, which could lead to errors in the results.

Excitation at the mouth using swept-sine excitation has also been successfully used to measure the transfer functions of 3D printed vocal tracts [94]. However, since the frequency content of the signal varies with time (unlike noise-like signals that contain all frequencies simultaneously), this method is sensitive to movements of the articulators, which can cause an apparent broadening of the resonance peaks [93], and is therefore not an appropriate measurement method for the vocal tracts of real subjects, who are not capable of remaining perfectly motionless.

Another method of obtaining the vocal tract transfer function, pioneered by Epps et al. [95], and used by others including Garnier et al. [15], Henrich et al. [14], and Joliveau et al. [58], involves exciting the vocal tract at the mouth with a synthesised pseudo-broad-band-noise signal, consisting of harmonics spaced 5.38 Hz apart. The use of harmonic components as opposed to actual broad band noise allows a “calibration” procedure to be carried out, where the subject keeps their mouth closed while the signal is played and recorded at the mouth, and the amplitudes of the different frequency components are adjusted to ensure that the excitation signal produces a flat frequency response when recorded. The measurement is then taken by exciting the vocal tract with this “calibrated” signal and the transfer function calculated. This allows the locations of the resonances to be deduced and provides a more natural result, since the subject can sing normally while the measurement is being taken, rather than being asked to hold an artificial position. This method is also more robust to small movements of the subject than swept-sine excitation, as the frequency content of the signal does not vary with time.

Magnetic resonance imaging

MRI data overcomes many of the problems with X-rays, as it does not involve ionising radiation [96], but instead uses a strong magnetic field to detect radio frequency signals from excited hydrogen molecules. MRI is well suited for investigating the vocal tract, as it allows the articulators to be observed directly, in a way that is relatively non-invasive for the patient. MRI does have its disadvantages however; scans are expensive, loud and tend to require the subject to lie supine in a narrow tube (unsuitable for claustrophobic subjects). They must also satisfy the safety criteria for MRI (not have metal in their body, be pregnant, obese or have tattoos).

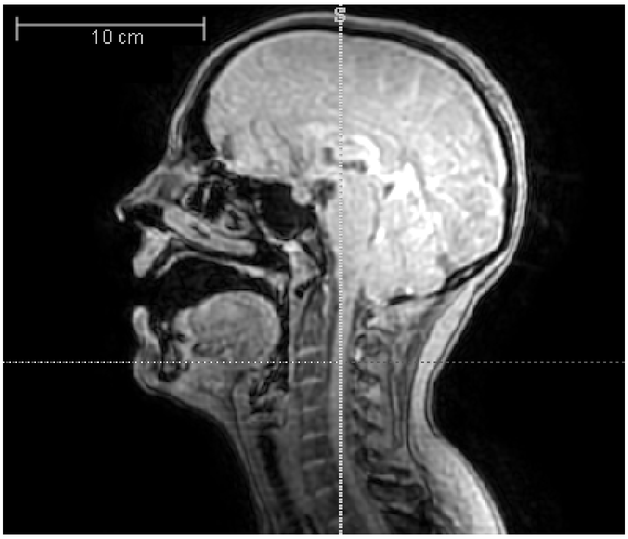


Figure 2.18: an example of MRI of vocal tract.

MRI scans are commonly available either as static 3-dimensional images, or as a dynamic 2-dimensional slice. Static MRI scans capture a series of 2-dimensional image slices through the head of the subject, allowing detailed 3-dimensional information concerning the body part of interest to be obtained and a highly accurate 3D model of the vocal tract to be generated. However, static MRI scans are only able to generate one set of these 2D slices at a time, requiring the subject to maintain a single position for enough time to obtain a single 3D image (usually of the order of 10 seconds). Dynamic MRI scans allow a moving image of a slice of the vocal tract to be captured, with a frame rate up to approximately 10 frames per second [6]. As it is only a single slice, this unfortunately does not allow a full three-dimensional moving model to be obtained, but can shed light on factors such as the behaviour of the tongue, soft palate, jaw, and epiglottis during singing.

The main advantage of 2D MRI is that it allows images to be captured in real-time which is closer to normal voice production. However, images from 3D MRI, although static, allow data in the transverse as well as mid-sagittal plane to be collected over a range of pitches, which can be used to generate more accurate cross-sectional area functions (plots of cross-sectional area of the vocal tract against its length) [97]. It also allows information in the transverse dimension, such as the width of the pharynx and other adjustable parts of the vocal tract (e.g. tongue) and the volume of the vocal tract to be examined over a singer's entire pitch range.

A number of studies have used MRI to observe the vocal tract during both speech and singing. For example Delvaux et al. [98] studied 1 female and 2 male singers to investigate the impact of the piriform fossae on the singing voice and found that they created a spectral trough in the 4-5 kHz region. Similarly, Echternach et al. [99] used MRI to study the vocal tract of a single soprano singing at very high frequencies and found minor modifications of the vocal tract shape, consisting of a decrease of the piriform sinuses, and small changes of tongue position. Echternach et al. [5] also investigated register changes in one tenor and one baritone using MRI and found that the singers made "few and minor modifications of vocal tract shape" when they changed from modal to falsetto and "some clear modifications" with changes in pitch (but not register). Narayanan et al. [100] used dynamic two-dimensional MRI to investigate the vocal tract shape during fluent speech, obtaining 8-9 images per second, showing clear real-time movements of the lips, tongue, and velum. Takemoto et al. [101] used a 3D "cine-MRI" technique to capture speech production, where the subject repeated a phrase 640 times.

There is very little research using 3D MRI to specifically investigate singing techniques in soprano voices. This will be discussed, together with resonance tuning in female voices using MRI, in section 2.3.2.

Visualising teeth in MRI

Apart from the unusual conditions required for MRI (the supine position and loud noise of the scanner), a significant limitation of magnetic resonance imaging is the difficulty in imaging teeth. Due the low hydrogen content of tooth and bone, these materials have the same appearance as air in an MRI image [102], as can be seen from Figure 2.18.

Failing to account for the teeth in segmentation of the vocal tract, can lead to over-estimation of the vocal tract volume [102, 103] (as the volume of the teeth are mis-estimated). A number of studies have attempted to integrate MRI images of the vocal tract with images of the teeth obtained separately, with varying degrees of success and these can be approximately divided into three categories [104].

The first of these involves a dental cast, such as Yang et al. [105], who used a plaster cast, which was then scanned, and Hasegawa-Johnson et al. [106], who also made a dental cast of each participant and submerged them in water. The disadvantage of this method is that bubbles and foams can form on the cast, causing artefacts.

The second type used MR-visible mouthpieces to cover the teeth, such as Kitamura et al. [107], who used a mouthpiece made from a thermoplastic elastomer, or Wakumoto et al. [108], who developed plates for upper and lower dental crowns. The weakness in this method arose from the thickness of the dental plates used (a few mm), which did not allow the air-tooth boundary to be accurately defined.

The final category uses a liquid contrast medium in the oral cavity. Olt et al. [109] asked subjects to fill their mouths with water, which resulted in clearly distinguished dental structures. Similarly, Takemoto et al. [103] asked subjects to hold blueberry juice in their mouths. In both of these methods, the subjects were required to remain in a supine position holding the contrast medium for a long data acquisition time, which was uncomfortable for the subjects and artefacts were caused by flow of the contrast fluid.

2.3 Resonance tuning

In classical singing and particularly in opera, one of the most important objectives for any singer is to be heard over the orchestra or other accompaniment. One method by which this is achieved is by altering the spectrum of the sound produced by adjusting the position of the vocal tract [14] in order to move the position of one or more formants away from their typical values in speech.

This altering of the vocal tract is known as formant tuning, or more correctly, *resonance tuning*, whereby the singer alters the resonances to move one or several of the early formants. This is done to increase the acoustic power transmitted by their voice, hence reducing the amount of energy required from the singer to produce a note of given amplitude and allow them to be heard over an accompaniment more easily.

2.3.1 Resonance Tuning in Male voices

In low male voices (bass/baritone), where the fundamental frequency does not exceed about Hz, the harmonics are closely spaced and there is a large amount of energy in the range of the first few formants. This makes it easy to deduce the positions of the vowel formants from the envelope shape of the sound spectrum of the vocal output [82] (for example, see Figure 2.15).

The closely-spaced harmonics in the frequency range of the low formants allows bass/baritone singers to alter the sound spectrum so that there is more energy in a frequency range of the spectrum not being masked by other noise sources (known as the singer's formant cluster, SFC).

Henrich et al. [14] investigated the resonance tuning methods used by all four voice types (sopranos, altos, tenors and basses), and observed that where the fundamental frequency is around 100 Hz (for low voices such as basses), the harmonics are sufficiently closely spaced that at least one will usually fall near to the normal value of R_1 (the first vocal tract resonance) to give a boost in sound level (see Figure 2.19), meaning that systematic resonance tuning would offer little advantage. The values of R_1 appeared near their values in speech, although may be slightly shifted to give better acoustic efficiency. Similar patterns for R_2 were observed.

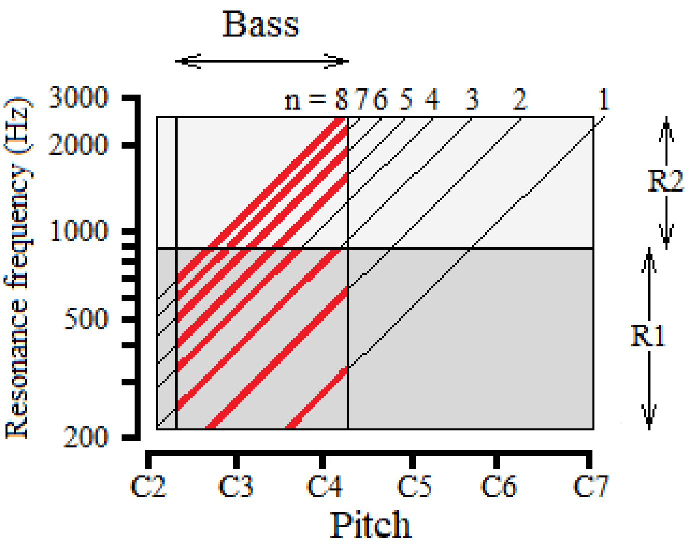


Figure 2.19: The many resonance tuning strategies available to be used by Basses, due to the large number of harmonics falling into the ranges of R_1 and R_2 (after [14]). Diagonal black lines show the first n harmonics, and diagonal red lines show the harmonics that can be tuned to the first or second resonance.

The Singer's Formant Cluster

It is extremely important for opera singers to be heard over large orchestras in concert halls and opera houses, so increasing their perceived loudness and making their singing more "resonant" is essential. One well-understood technique used by male singers and altos is the Singer's Formant Cluster (SFC) (formerly called the *Singer's Formant*), which is a clustering of formants 3, 4, and 5 [11] and is characterised by increased energy in the voice spectrum in the region between 2-4 kHz. It is generally found in operatic singing and some other western classical styles, including musical theatre, and has also been found in theatrical speaking [72]. The SFC is usually achieved by altering the position of the articulators, for example by lowering the larynx or increasing the space in the pharynx, which results in the convergence of formants 3, 4, and 5, which are not crucial for accurate vowel identification [66].

The long-term average spectrum of a whole orchestra produces a sloping spectrum (shown in Figure 2.20), meaning that the acoustic energy in the frequency region of the SFC (2-4 kHz) is relatively low. Therefore when a singer sings using the SFC, they exploit this lack of acoustic energy, allowing the audience to hear the singer over the orchestra. Significantly, the 2-4 kHz frequency region is also the most sensitive region of human hearing [110].

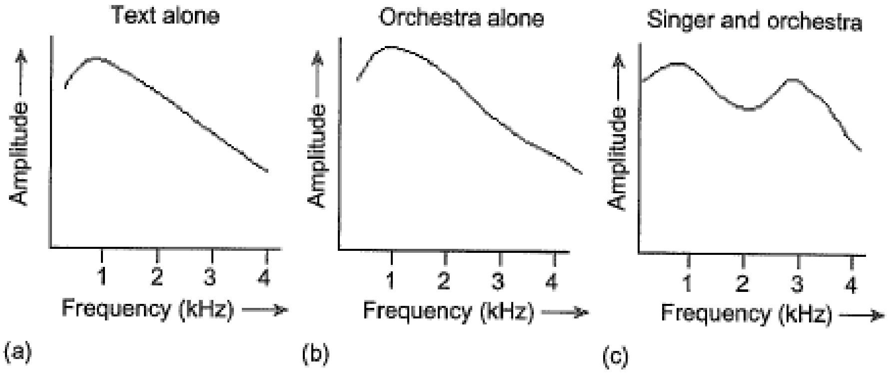


Figure 2.20: A representation of the LTAS of (a) a singer alone, (b) an orchestra alone, and (c) the two performing together (not to scale) (Reproduced with permission from [20]).

The SFC is not the only method of formant tuning used by male singers. They are also capable of tuning one or more of their formants to near a harmonic of their voice. This is known to be employed by several famous singers, including the tenors Luciano Pavarotti and Placido Domingo. Miller [111] found that while Domingo relied on the resonance of the SFC at around 2.8 kHz, Pavarotti used formant tuning to shift the second formant of the vowel /ɒ/ by about 500 Hz, to coincide with the third harmonic of the note, giving the sound an entirely different quality, and Pavarotti his distinctive tone.

For higher male voices such as tenors (f_0 range approx. 131 - 523 Hz), the tuning of individual resonances is also thought to have advantages. Near the upper limit of the tenor range at around 500 Hz, it is possible that the nearest harmonic to R_1 (in speech) might be 250 Hz away. This is especially the case for the /u/ vowel, which has a low value of R_1, and hence R_1:f_0, and R_1:$2f_0$ tuning are possible over the full tenor range, and R_1:$3f_0$ tuning might be useful in the lower part of the range. R_2 could also be tuned to $2f_0$, $3f_0$, or even $4f_0$ [14, 112]. The full range of resonance tuning strategies available to tenors is shown in Figure 2.21.

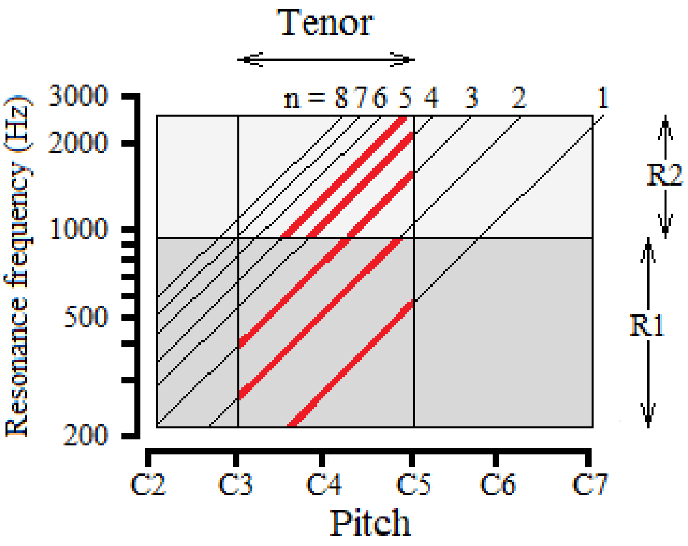


Figure 2.21: The resonance tuning strategies available to be used by Tenors, due to the harmonics falling into the ranges of R_1 and R_2 (after [14]). Diagonal black lines show the first n harmonics, and diagonal red lines show the harmonics that can be tuned to the first or second resonance.

This was investigated by Titze et al. [113] who used an analysis-by-synthesis technique, adjusting the formant frequencies and glottal properties of a source-filter model to match the spectra of 6 tenor voices. R_1 was found to be higher than the fundamental for all vowels except /u/, and the absence of R_1:f_0 tuning was interpreted as due to a desire to maintain a "characteristic male quality" [113].

Both R_1:$2f_0$ and R_2:$2f_0$ formant tuning were investigated in baritone singing by Miller and Schutte [114] using measurements of sub- and supra-glottal pressure. They concluded that the "resonance-enhancing effects of formant tuning appear to be intentionally exploited by the singer in response to the demands of the musical phrase".

Neumann et al. [115] inferred from the sound spectrum that, for male opera singers in the *chest* register, the second harmonic was "resonated by" the first resonance (R_1:$2f_0$ tuning) and the fourth harmonic by the second resonance (R_2:$4f_0$ tuning) with the implication that $R_2 \approx 2R_1 \approx 4f_0$. Across the passaggio (the transition between registers), R_2 often fell near $3f_0$.

2.3.2 Resonance Tuning in Female voices

Unlike male singers, where the SFC helps to boost the amplitudes of the harmonics around 3 kHz, evidence of a true SFC in sopranos is extremely limited. At higher fundamental frequencies, the acoustic possibilities of the SFC become less useful to the singer, as it can only increase the relative amplitude of harmonics in the spectrum between 2-4 kHz. Therefore at very high fundamental frequencies, very few harmonics will fall in the SFC area of the spectrum. Whilst low female voices may make use of the technique to a limited extent, the physiological and acoustic limitations of the soprano voice make the use of alternative techniques necessary.

Since sopranos sing at extremely high pitches there is already a considerable amount of spectral energy in this region due to the presence of high-amplitude "early" harmonics [12], caused by the high fundamental frequencies. For example if a singer sings a high C, at approximately 1000 Hz, then the first, second and third harmonics will be at 2 kHz, 3 kHz, and 4 kHz respectively, so only 3 harmonics will fall in the frequency range usually associated with the singer's formant cluster. Barnes et al. [116], in a study involving 6 sopranos, also found a large amount of energy at high frequencies for successful singers (in the region 2-4 kHz), but no evidence of a singer's formant cluster.

Sundberg [13] proposed that soprano singers could *tune* one or both of the first two vocal tract resonances to near the harmonics of the larynx voice source in order to make full acoustic use of these resonances and increase the acoustic output power without requiring increased effort from the singer.

At the upper end of the soprano range, fundamental frequency can exceed 1000 Hz. It is not only possible, but highly likely, that the fundamental frequency will fall above one or more of the first two formants, as the first and second formants typically lie between 310 and 860 Hz and 920 and 2790 Hz, (D#4 and A5, and A#5 and F7) respectively [117], which are mostly within a soprano's range. This implies that not only will the production of sound be much less efficient, as some resonances of the vocal tract are not being utilised to their full potential, but it is also likely that the vowel will be harder to identify, as there is little or no spectral energy at the formant frequencies.

In the speech of an adult female, for example, the expected frequency of the first formant for an /ɑ/ vowel would be approximately 850 Hz [117]. However, if singing in the upper middle of her range at an A5 (880 Hz), the first formant is redundant as there is no sound energy in this frequency range to amplify. The soprano singer therefore tunes this resonance (again by manipulating the placement of the sound modifiers) near to the fundamental frequency, greatly increasing the relative intensity of the fundamental and increasing the perceived loudness of the sung tone.

An experiment by Garnier and Henrich [15] investigated the resonance tuning strategies used by sopranos across their range. The study involved twelve sopranos (4 non-experts, 4 advanced, 4 professionals) who sustained pitches on /ɑ/ vowels, from A4 (440 Hz) to their highest sustainable note, while the frequencies of the first two vocal tract resonances (R_1 and R_2) were measured by broad band excitation at the mouth [58].

They found that R_1:f_0 tuning was employed below C6 by all the professionals and advanced singers and, to some extent, by the non-expert singers as well. R_2:$2f_0$ tuning was seen in 3 professionals, 2 advanced, and 2 non-expert singers. Six of the singers used R_2:f_0 tuning at very high pitches (above C6), whilst R_1:$2f_0$ tuning was only found in two of the singers (in the lower part of the range investigated).

The resonance tuning strategies available to soprano singers is shown in Figure 2.22. It can be seen that R_1 can be tuned to f_0 throughout the entire vocal range, and $2f_0$ at the lower end of the range. R_2 can be tuned to $3f_0$, or in the upper part of the soprano range, to $2f_0$.

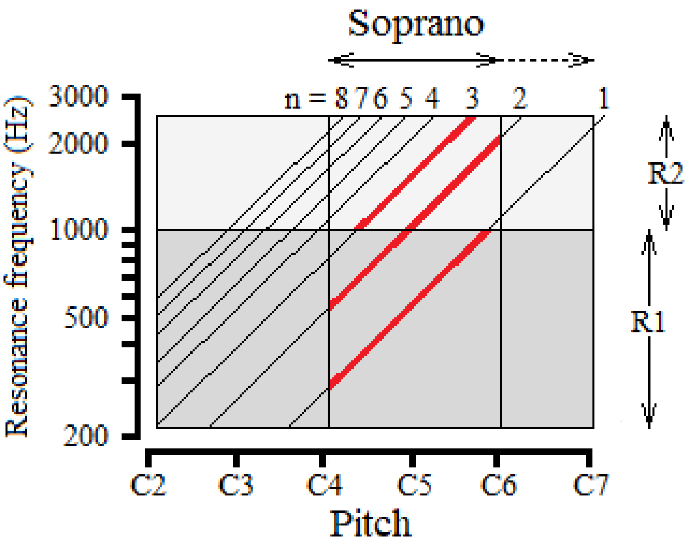


Figure 2.22: The few resonance tuning strategies available to be used by Sopranos (after [14]). Diagonal black lines show the first n harmonics, and diagonal red lines show the harmonics that can be tuned to the first or second resonance.

Production of Resonance Tuning

Although research into the effects of various articulators on *speech* has been ongoing for over 40 years, e.g. [25, 118, 119, 120, 121], it cannot be assumed that the same articulatory techniques are used in singing. Indeed Ventura et al. [122] investigated the differences in the vocal tract between speaking and singing and found differences in the volumes of the whole vocal tract cavity and the oral and the pharynx cavities for some vowels.

The resonance tuning techniques available to professional soprano opera singers are now understood to an extent [13, 14] (limited to very few subjects and vowels), as well as those strategies most used in different pitch ranges of the soprano voice [59]. However, the exact methods used to *produce* these effects are not yet fully understood. Garnier et al. [59], investigated resonance tuning while capturing video of the singer's face to measure effects such as lip opening and spreading, and found that while some singers increased mouth area with f_0 over the whole range of R_1:f_0 tuning, others only showed this strategy on the higher part of this range. Studies investigating male singers, such as Pabst et al. [123] and Daffern [124] found that some singers raise their larynges with increasing fundamental frequency, although this was not always the case.

Previous studies on resonance tuning using MRI include Echternach et al. [6], who investigated registers in both male and female voices using real-time 2D MRI, considering factors including lip opening, jaw opening, tongue height, jaw protrusion, oropharynx width and uvula elevation. Clear vocal tract changes were not found to be associated with either of the register transitions investigated. However, changes in the measured physical parameters were found when f_0 approached the first formant. In a subsequent study, Echternach et al. [99] also used a combination of real-time 2D and static 3D MRI to investigate 3D factors including the tongue shape, the size of the piriform sinuses, and lip and jaw opening at very high fundamental frequencies. The study found only minor modifications of the vocal tract shape, involving a decrease of the piriform sinus as well as small changes of tongue position. Formant frequencies were not very different between C6 and G6 for F_1 and F_3, respectively, and F_2 was only slightly raised for G6. Bresch et al. [7] used real-time 2D MRI to investigate resonance tuning in five sopranos, and although subjects generally showed a more open mouth shape with increasing fundamental frequency, it was suggested that sopranos might not all employ the same generalisable strategies for resonance tuning as had previously been thought.

Studies researching resonance tuning in soprano voices have generally not included as many subjects as would be desirable statistically, due to the difficulties involved in obtaining suitable subjects and suitable measurement facilities (such as the very high cost of MRI scans). For example, in Sundberg et al. [13] and Carlsson et al.'s [125] early work identifying resonance tuning, only one soprano was considered. Similarly, Echternach et al. used MRI to study the vocal tract of a single soprano singing at very high frequencies [99], and register changes in one tenor and one baritone [5]. Miller et al. [126] compared methods of locating formant frequencies in one bass-baritone singer, and Delvaux et al. [98] studied one female and two male singers to investigate the impact of the piriform fossae on the singing voice. Similarly, in a study on speech, Sulter et. al [118] studied a single male subject to compare predicted resonances with measured values. Clément et. al [127] compared vocal tract resonances obtained from recorded speech with those calculated from an area function of the vocal tract acquired using MRI in one male speaker.

2.3.3 Other resonance tuning techniques

While the SFC and soprano resonance tuning techniques are now beginning to be understood, these are by no means the only resonance tuning techniques available to singers. The human voice is an incredibly versatile and adaptable instrument, the product of thousands of years of evolution, and can produce sound in a variety of methods.

The Whistle register

The highest vocal register is commonly referred to as the *whistle* register, and uses laryngeal mechanism M3. It is characterised by a concentration of acoustic power in the two first harmonics [128], reduced power around 3 kHz and enhanced jitter [129]. Perceptually, its voice quality is described as *fluty*. In this frequency range, $R_1 : f_0$ tuning becomes less useful, as above a certain frequency the vocal tract cannot be further lengthened and the jaw reaches its lower limit, which prevents R_1 being raised any higher. Evidence of $R_2 : f_0$ has been found in professional soprano singers above 1046 Hz (C6) [15].

Overtone singing

An extreme example of resonance tuning is Tuvan throat singing [130], where a low fundamental note is sung and one resonance of the vocal tract is varied to play a tune, producing a *whistling* or *fluty* tone. Resonance tuning techniques are also used in the Bulgarian folk singing styles “teshka” and “leka” [131]. These involve tuning the first vocal tract resonance to the second harmonic of the voice, in order to produce a distinctive tone.

2.4 The Perception of Resonance Tuning

Although resonance tuning is advantageous to the singer in terms of acoustic efficiency, it does have disadvantages; the primary function of the lower formants is to convey information about vowels, so it would be expected that changing the position of the formants would change the quality of the vowel, which may be one of the reasons that vowels at high pitches are harder to identify.

This has implications on performance practice and the training of singers, as there is considerable disagreement between singing teachers on the correct approach to singing high notes, many believe that at high pitches it is necessary to “neutralise” vowels to some extent and produce them in a more similar way, whereas others believe it is important to keep the different vowel sounds distinct [3].

2.4.1 Perception of formant/resonance properties

The frequency response of the vocal tract is characterised not only by the locations of the peaks in the transfer function (resonances), but also by properties such as the relative amplitudes of formants/resonances, their Q factor (the centre frequency relative to the bandwidth) and their bandwidths.

Each of these separate factors can be altered to a certain extent before a perceptual difference is noticed by a listener. A study by Flanagan [132] found that for synthesised vowels with a fundamental frequency of 120 Hz, the first formant could be shifted 20 Hz and the second formant 20-35 Hz before listeners could detect a difference. A study by Mermelstein [133], found that the difference limen (in speech) was 50 Hz for F_1, and 142 Hz for F_2. However, this was greater when the vowel was between two consonants than when it was just spoken alone.

The effects of the relative amplitudes of the formants has been investigated by Flanagan [134], who varied the amplitude of the second formant of synthetic vowel sounds and found that 50 % of listeners identified two sounds as "different" when their second formants differed by 3 dB. Kiefte et al. [135] found that the amplitude of the second formant affected whether the listener perceived an /i/ or an /u/ vowel and concluded that this might partly be due to masking effects when the amplitude of a formant was very low relative to neighbouring formants.

Although evidence of resonance tuning in professional soprano singers has now been observed in a number of studies (although it should be noted that not all singers employed the same techniques), there is a lack of research into its perception.

There have been very few studies specifically investigating the perception of resonance tuning. In 1991, Carlsson-Berndtsson and Sundberg published a perceptual study [125] in which synthesised sung vowel sounds were generated to represent a male voice at fundamental frequencies ranging over a descending octave-wide chromatic scale from C4 (261 Hz) to C3 (131 Hz), representing the vowel /ɑ/. These vowel sounds were then treated in one of four ways. In "strategy A" the first formant was tuned to the harmonic closest to 550 Hz. In "strategy B", the second formant was tuned to the harmonic lying closest to 1000 Hz. In "strategy C" either the first or second formant was tuned to the harmonic closest to 550 or 1000 Hz, depending on which option gave the smallest formant frequency deviation from these values. Finally in "strategy D", the formants remained at 550 and 1000 Hz in all vowel sounds. Sounds with tuned formants (using strategies A, B, or C), were presented together with the non-tuned vowel sounds (strategy D) in pairs, and 19 listeners were asked, "Which voice production do you find most correct?". The vowel sounds with unchanged formant frequencies were preferred by all subjects except for one (out of nineteen), however the mere-exposure effect [136] (the psychological phenomenon whereby people prefer stimuli that they are more familiar with) could contribute to these findings, as due to the pairing methods used, subjects heard the sounds with unchanged tuning three times more often than the other tuning strategies.

2.4.2 Perception of vowels at high frequencies

There have been a number of studies on the perception of vowels at high frequencies (nearing 1000 Hz) which show that the likelihood of a sung vowel being misunderstood increases with f_0. Scotto di Carlo et al. [137] showed that in perceptual tests, where listeners were played samples of sung vowels, the vowel was identified incorrectly much more frequently as the pitch of the note increased, for all vowels. 64 % and 62 % of samples were correctly identified in lower and lower-middle registers respectively, but only 18 % and 9 % were correctly identified in upper-middle and upper registers. They also found that incorrectly identified vowels tended to be confused with /ɑ/ the most, and suggested that the reason for this was that, at high pitches, the vocal tract tends to the shape required to produce an /ɑ/ sound. The difference in perception between registers was thought to be caused by different methods of resonance tuning and different use of articulators. The vowels were categorised according to features such as lip rounding and jaw opening; it was found that lip rounding was not conserved at high pitches, as the singer tended to spread the lips, explaining why /i/ and /e/ were relatively well-perceived at high pitches. Jaw opening was also found to be important and closed vowels such as /u/ (see Figure 2.23) were only well-identified at lower pitches.

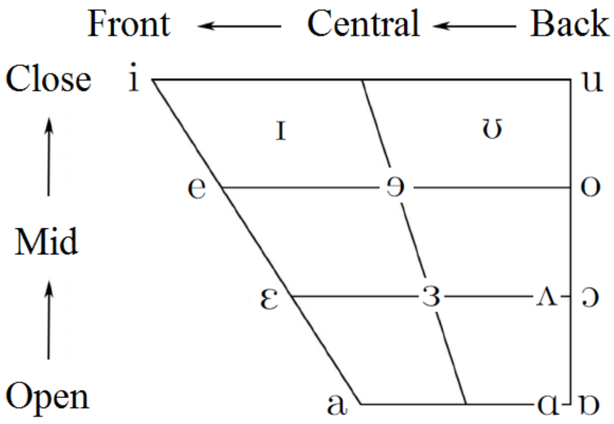


Figure 2.23: The relationship between different IPA vowels and how they can be categorised as *open* to *closed* and *front* to *back*.

A similar study by Benolken et al. [138] recorded a soprano singing 12 vowels over her entire range and asked listeners to identify which vowel it was. This study also found that vowels were more likely to be mis-identified with increasing fundamental frequency and concluded that this was due to the fact that several vowels have similar second formants, but are mainly distinguished by their first formants, so as the fundamental frequency is increased and the first formant raised, these sound increasingly similar.

2.5 Summary

This chapter has presented an overview of voice production and the main mechanisms by which sound is produced by the vocal folds and subsequently modified by the vocal tract to produce the different vowel sounds and timbres that allow production of a wide range of voiced sounds. Measurement techniques for analysing the various aspects of the voice were considered. Spectrograms, LTAS, LPC and Fourier transforms can be employed to analyse the acoustic output of a singer, while laryngoscopy or endoscopy allow direct observation of the vocal folds and EGG can be used to deduce key parameters such as glottal closure, allowing insight into the workings of the vocal folds and larynx. Finally, the structure of the vocal tract itself can be studied using X-ray, magnetic resonance imaging or articulography.

Current literature (and findings) concerning resonance tuning in classical singing have been reviewed; how a singer can modify the vocal tract resonances and the effects of this on the acoustic output have been discussed. Some of the modifications that are commonly applied have been introduced: the singer's formant cluster technique in male voices, possible methods of resonance tuning in higher male voices (including tuning higher formants to harmonics) and finally the reasons for resonance tuning in soprano voices and the observed methods of resonance tuning in professional sopranos.

The main focus of this work will be on resonance tuning in soprano voices. This work will build on findings of Joliveau, Garnier and Henrich et al. [14, 15, 58], which has established the extent and type of resonance tuning by sopranos (section 2.3.2), and work by Echternach and Bresch et al., [7, 99], which has investigated articulator movement in the mid-sagittal plane using MRI techniques.

Chapter 3

The Perception of Resonance Tuning

This chapter details an experiment carried out to investigate listeners' *perception* of different methods of resonance tuning. The chapter sheds light on the purpose and effects of resonance tuning to better understand how it can be applied. Resonance tuning aims to increase the acoustic efficiency of voice production across a range of fundamental frequencies, but it also has an impact on perceptual aspects of the singing voice.

This pilot study investigates the nature of these perceptual effects and informs the main work of this thesis, presented in Chapter 6, in which the type and extent of resonance tuning used by adult soprano singers is investigated, as well as the articulatory mechanisms involved in its production.

This work was first published as a journal paper "The Perception of Formant Tuning in Soprano Voices" in Journal of Voice [139].

3.1 Introduction

As introduced in Chapter 2, when singing at high fundamental frequencies, resonance tuning represents a highly efficient method of increasing the acoustic power produced by the voice and it is now an accepted phenomenon in soprano singing [13, 15, 58]. Adjusting the resonances of the vocal tract ensures that a harmonic will fall within the frequency range of a vocal tract resonance and allows the singer to match the acoustic impedance of the source (glottis) and the filter (vocal tract) [59, 140]. The first and second formants in female speech typically lie between 310 and 860 Hz (D#4 and A5) and 920 and 2790 Hz (A#5 and F7) respectively [117]. The soprano range can extend to above 1000 Hz, there is therefore a wide range of frequencies over which resonance tuning can be used by a soprano singer.

In 1991, Carlsson-Berndtsson and Sundberg [125] published a perceptual study in which synthesised sung vowel sounds were generated to represent a male voice and had the first formant tuned to a nearby harmonic. The samples with unchanged formant frequencies were preferred by all listeners except for one (out of nineteen), so it was concluded that listeners preferred singing without formant tuning. However, the design of the study meant that formants could be tuned to different harmonics in adjacent vowel sounds. Also, listeners heard the untuned samples three times as often as any of the other tuning strategies, possibly introducing confounding effects such as the mere-exposure effect [136].

In addition to the effects of resonance tuning, acoustic theory suggests that vowel recognition greatly diminishes at high fundamental frequencies [137] and there is still some debate as to whether singers should attempt to “neutralise” vowels at high fundamental frequencies. Singers can therefore either choose to focus on the sound quality produced (rather than the perceptual distinction between vowels) or make a special effort to keep vowels distinct, potentially sacrificing some acoustic efficiency and ease of production [3].

This chapter will discuss a study conducted using a perceptual test to investigate the perception of resonance tuning at high fundamental frequencies. In previous works [14, 15], four different methods of resonance tuning have been observed in sopranos:

(i) extensive R_1:f_0 tuning,

(ii) often in conjunction with R_2:$2f_0$ tuning;

And at the upper and lower ends of the frequency range,

(iii) some R_2:f_0 and

(iv) R_1:$2f_0$ tuning.

The two most common tuning conditions reported in previous work are R_1:f_0 and R_2:$2f_0$ tuning. Therefore, these are investigated in this experiment, both together and in isolation, producing four different tuning strategies in total.

The protocol used alters that of Carlsson-Berndtsson and Sundberg [125] to be suitable for the soprano voice and removes the possibly confounding influence of the mere-exposure effect [136]. The properties investigated include which tuning strategies are *preferred*, their *naturalness* and which produce the mostly clearly *identifiable* vowel sounds.

The hypothesis of this experiment is that the strategies used most frequently by sopranos in practice will be those that are preferred by listeners, perceived to be most natural and correctly identified most often. A good performer should aim to not only produce a sound that is pleasing and natural to the audience, but also be understandable and accurately convey the meaning of the text. It might not be possible to achieve all of these aims, so some degree of compromise may be required.

3.2 Method

Similar to the procedure used by Carlsson-Berndtsson et al. [125], synthesised vowel sounds were created to replicate voiced sounds, for which the resonance frequencies could be controlled to represent different resonance tuning strategies. Samples with f_0 typical for a soprano range were synthesised and, as resonance values have been shown to remain approximately constant in singing up to the frequency where $f_0 = F_1$ [125], the average formant values in speech for women's voices were used for the baseline resonance values (as defined by Peterson and Barney [117]). These are shown for the three vowels investigated in Table 3.1. As in [125], four resonance tuning strategies were tested:

- In "strategy A" no resonance tuning was used, so the vowel resonances remained constant at the average values for the vowel.
- In "strategy B", the first resonance was tuned to the fundamental (R_1:f_0), while the second and third resonances (R_2 and R_3) were kept constant at the average values for the vowel.
- In "strategy C", the second resonance was tuned to the second harmonic (R_2:$2f_0$), while the first and third resonances (R_1 and R_3) were kept constant at the average values for the vowel.
- In "strategy D", the first resonance was tuned to the fundamental (R_1:f_0), and the second resonance was tuned to the second harmonic (R_2:$2f_0$), while the third resonance (R_3) was kept constant at the average value for the vowel.

3.2.1 Synthesised Signal

Synthetic vowel sounds were used in this study, as they offered control over the parameters to be investigated (i.e. resonances). Synthetic vowels also

Vowel	F_1	F_2	F_3
/ɑ/	850 Hz (G#5)	1220 Hz (D6)	2810 Hz (F7)
/u/	370 Hz (F#4)	950 Hz (A#5)	2670 Hz (E7)
/i/	310 Hz (D#4)	2790 Hz (F7)	3310 Hz (G#7)

Table 3.1: The first three formant values for three vowels, when spoken by female voices [117].

eliminate the variations that occur in natural vowel sounds, which could potentially introduce confounding effects in the results. The synthesised sounds were generally not perceived to be authentic, but they nevertheless permitted the assessment of relative changes in perceptual attributes, such as preference, naturalness and vowel identification. This is useful for informing the design of the main study in Chapter 6.

Glottal Signal

The synthesised vowel sounds were produced in MATLAB [68] using a Liljencrants-Fant (LF) glottal flow model to create a model of the voice source (glottal signal). Typical parameter values for a female were used from [26] (setting $R_d = 1$) (full parameters in Appendix I):

$$F_a = 400Hz, \quad R_k = 0.30, \quad R_g = 1 \tag{3.1}$$

Where F_a is the cut-off frequency (accounting for the degree of spectral tilt), R_k specifies the relative duration of the falling branch from the peak at time T_p to the discontinuity point T_e, and R_g is a parameter which increases with a shortening of the rise time T_p.

Vibrato was also added to the voice source to increase naturalness, and make it sound sung rather than spoken. This consists of a 6 Hz [141] sinusoidal modulation of the fundamental frequency, with an extent of 60 cents [141].

Vocal tract effects

The resonances of the vocal tract were treated as a series of connected single-peak IIR filters using the `iirpeak` function in MATLAB [68] (all synthesis code in Appendix N). The glottal signal was passed through each filter in turn. A study investigating formant bandwidth [142] used averaged data from Fujimura and Lindqvist [93] and Fant [143]. It was found that the bandwidth remains approximately constant, at around 50 Hz, for formant frequencies between 300 and 2000 Hz. In light of this, the values used for the resonances here are the formant values shown in Table 3.1 [117]. The resonances and formants can be considered approximately equivalent, with the bandwidths fixed at 50 Hz.

To make the synthesised voice sound more natural, and to prevent transient effects due to a sudden onset and offset of the sound, an amplitude window the same length as the sample was applied. The window consisted of the rising and falling halves of a Hanning window in the first and last quarter of each sample, respectively. This simulated a “flat”, sustained vowel sound, with an onset, constant period, and offset [144]. The amplitude window was defined as a proportion of the sample length, but since all the samples had the same length, this equated to a constant onset duration, as typically occurs in natural vowels.

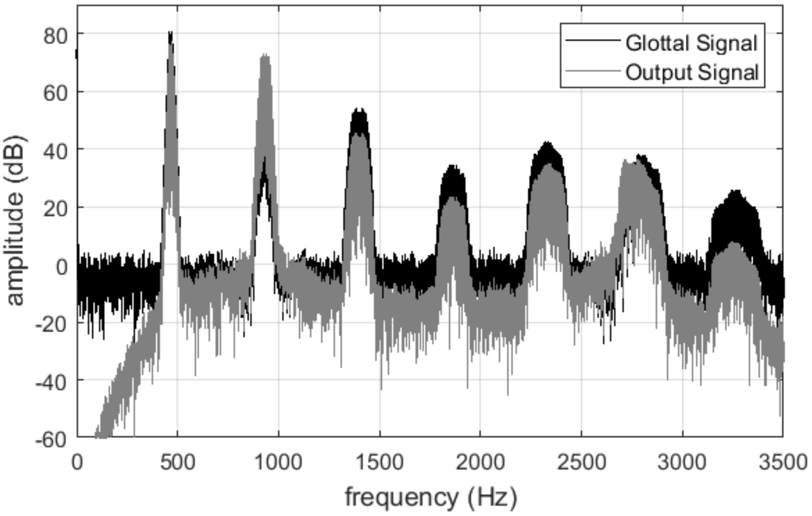


Figure 3.1: The glottal signal (black) synthesised using the LF model, and the output signal including the vocal tract effects for an /u/ vowel (grey).

The resulting synthesised signal was then de-emphasised (as recommended by Hermes [145]) to improve the naturalness. This was done by low-pass filtering the signal to produce a resultant spectral slope of approximately -12 dB per octave, so that the relative resonance amplitudes more closely resembled the human voice [19]. The fundamental frequencies were chosen to be around the frequency of the first resonance, as shown in Table 3.2. Examples of the spectrum of the glottal signal and resulting output signal (including the vocal tract effects) are shown in Figure 3.1.

The samples were each 2 seconds in length, with bit depth 16 bits and a sampling frequency of 44.1 kHz. The samples were generated as a single channel, but played over two channels in dual mono.

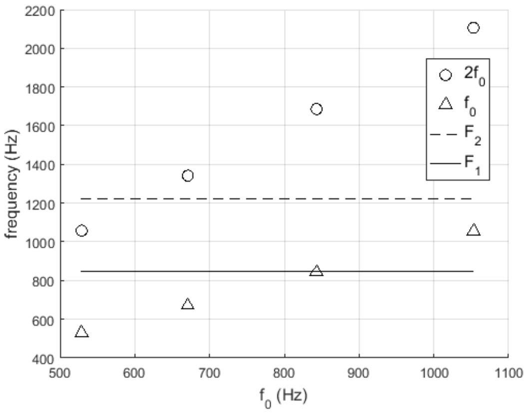
In practice, a vocal tract resonance at a frequency just *above* a harmonic produces an inertive reactance, causing the vocal tract to assist the vibration of the vocal folds, which results in an increased acoustic power output [27]. Conversely, when a vocal tract resonance is slightly *below* a harmonic, there is a compliant reactance and the vocal tract no longer assists the vibration of the vocal folds, resulting in a reduced acoustic power output [146]. There-

Pitch no. Vowel	1	2		3	4
/ɑ/	529 Hz (C5)	671 Hz (E5)	($F_1 = 850$ Hz)	843 Hz (G#5)	1053 Hz (C6)
/u/	233 Hz (A#3)	294 Hz (D4)	($F_1 = 370$ Hz)	370 Hz (F#4)	472 Hz (A#4)
/i/	220 Hz (A3)	277 Hz (C#4)	($F_1 = 310$ Hz)	349 Hz (F4)	440 Hz (A4)

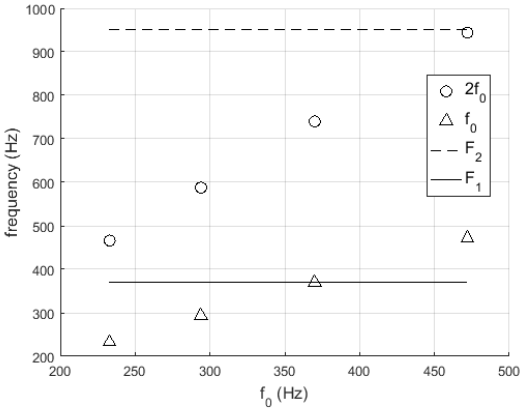
Table 3.2: The fundamental frequencies of the four synthesised vowel sounds for each vowel sound (12 in total). These were generated for each resonance tuning strategy.

fore, to maximise the impact of resonance tuning, vocal tract resonances are tuned to just above the relevant harmonic frequencies (5 cents above in this experiment).

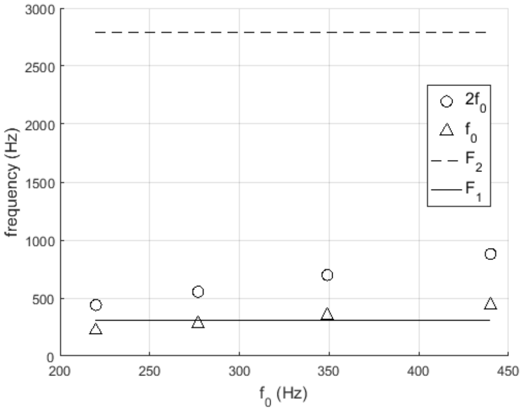
The relationship between the resonances and harmonics can be seen in Figure 3.2, where the harmonics are plotted against fundamental frequency and the formant values in speech (the untuned values for R_1 and R_2) are represented by horizontal lines (from [117]).



(a) /ɑ/ vowel.



(b) /u/ vowel.



(c) /i/ vowel.

Figure 3.2: The values of the first and second formants in speech (solid and dashed lines respectively), for each vowel, and the values of f_0 and $2f_0$ (triangle and circle, respectively) for each of the four pitches.

3.2.2 Participants and Distribution

A perceptual test was designed to test the hypothesis and distributed via email and social media using the online survey software Qualtrics [147]. The audio files for each vowel, pitch and tuning strategy were converted to mp3 files and uploaded to Qualtrics. Before commencing the perceptual test, participants[1] first answered a questionnaire to ascertain demographic information, their level of vocal ability, singing training and their music listening habits. This captured the listener's own singing ability, as well as their experience of listening to professional singing. Nine listeners had some singing training (four of whom had professional training). None of the participants in this experiment had participated in the other experiments of this thesis (the preliminary experiment in Chapter 4, or the main experiment in Chapters 5 and 6).

45 participants took part, but results from 15 of these were discarded, either because they did not complete the entire test, or because they reported serious hearing problems. Of the remaining 30 participants, 20 were male, and 8 female (2 chose "other/prefer not to say"). They were aged 20-75, with a mean age of 33.7 years. The time taken (including breaks) varied from 13 minutes to 73 minutes (discounting 2 outliers), with an average time of 32 minutes.

Participants were able to take the perceptual test on their own devices (excluding mobile devices). 15 participants used closed-back headphones, 7 used open-backed headphones, and 7 used earbuds. Participants were instructed to take the test in a quiet environment with no distractions and not to adjust the audio level on their computer after starting the test.

There may have been slight differences in audio quality between participants, but internet distribution allowed a greater number and variety of partici-

---

[1] The terms "participant" or "listener" have been used to refer to people who participated in the listening test described in this chapter. To distinguish between the different types of involvement, "subject" is used to refer to the singers studied in Chapters 4, 5 and 6.

pants to participate in the test, so was considered worthwhile. Schoeffler et al. compared laboratory and web-based results of an auditory experiment and found no significant differences [148], suggesting that this can be an acceptable distribution method.

Prior ethical approval was gained from the Physical Sciences Ethics Committee at the University of York.

3.2.3 Perceptual test design

The perceptual test consisted of comparisons between sets of four vowel sounds using sliders. Each set contained four synthesised vowel sounds, created as outlined in section 3.2.1, with the same f_0 and vowel, but treated with the four different tuning strategies A, B, C and D.

The design was similar to a MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) [149] test design, where the untuned sample could be considered as a reference. The participants could press the buttons to play the vowel sounds as many times as they wished. Each set of four vowel sounds was presented in a random order and the order of vowel sounds presented in each question was also randomised to minimise the effects of program-dependence. Participants were asked to rate preference and naturalness on continuous sliding scales from 0 to 100, with 100 indicating the highest preference or naturalness.

For the vowel identification, listeners were presented with each sample individually. They were asked “which vowel does this sample sound the most like?” and given a choice of 12 different vowel sounds, presented as short words.

Examples of the test graphical user interface (GUI) used to present comparisons to participants for the (a) preference, (b) naturalness, and (c) vowel identification are shown in Appendix A.

This design ensured that each resonance tuning strategy was heard by the listeners an equal number of times, reducing the possibility of confounding effects due to the mere-exposure effect [136].

3.3 Results and Analysis

3.3.1 Data Processing

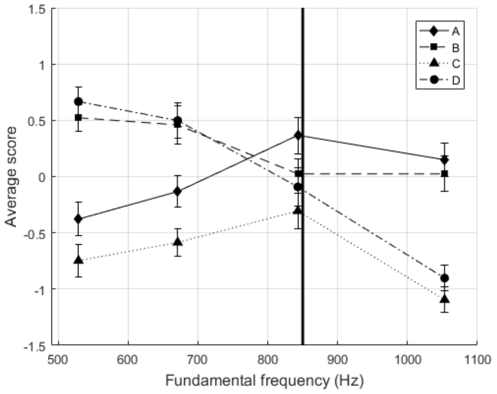
Data collected from the questionnaire, together with the perceptual test answers were collected in Excel and then imported into MATLAB for analysis. Raw results are included in Appendix N.

The scores for preference and naturalness were first normalised to have a mean of zero and a standard deviation of 1 across each participant, to reduce inter-participant variability. The mean score and the standard error of the mean across all participants were then calculated for each vowel, f_0 and tuning strategy, so that the average normalised score could be plotted against f_0 for each vowel. The results for preference and naturalness are shown in Figures 3.3 and 3.4 respectively.

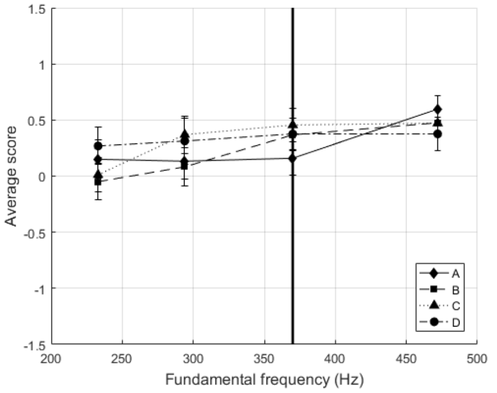
The question on vowel identification was analysed by calculating the percentage of participants that chose the correct vowel sound for each sound. These values are shown in Figures 3.5(a)-(c)(i) for each vowel, and the most commonly chosen vowel sound (correctly or incorrectly) is shown in Figures 3.5(a)-(c)(ii).

The continuous scales for the questions on preference and naturalness allowed Analysis of Variance to be carried out on these results. This is discussed in section 3.3.6.

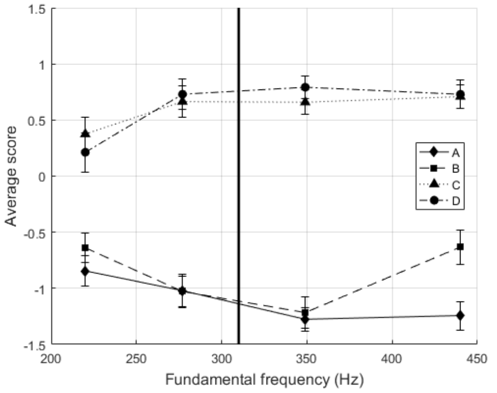
3.3.2 Results



(a) /ɑ/ vowel.

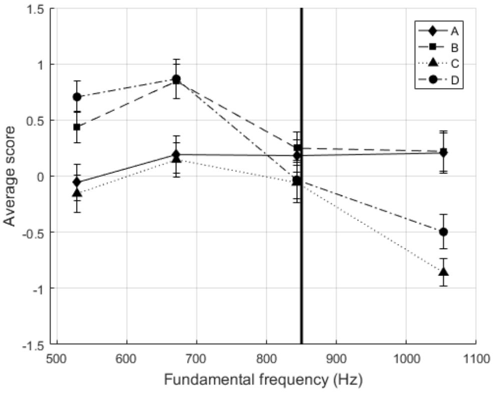


(b) /u/ vowel.

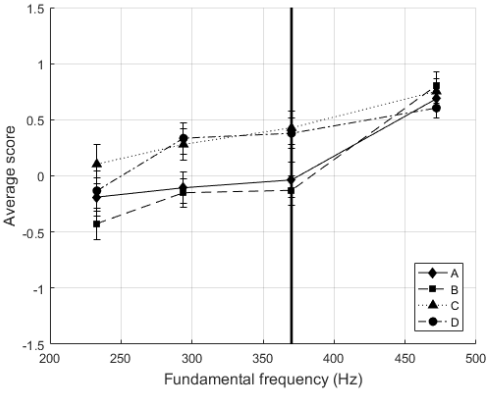


(c) /i/ vowel.

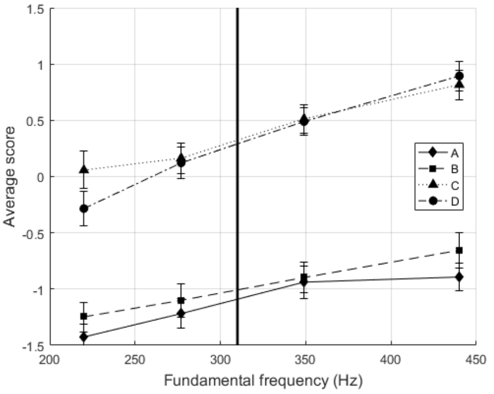
Figure 3.3: The average scores for the different tuning strategies A-D (see section 3.2.3) investigated for *preference*, for each vowel. The standard error of the mean is shown by error bars. The thick vertical line shows the frequency of the first formant in speech, from [117].



(a) /ɑ/ vowel.

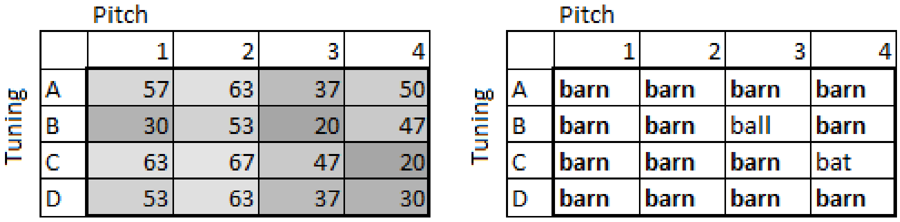


(b) /u/ vowel.



(c) /i/ vowel.

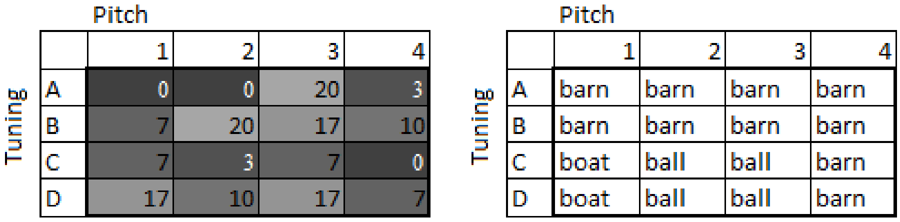
Figure 3.4: The average scores for the different tuning strategies A-D (see section 3.2.3) investigated for *naturalness*, for each vowel. The standard error of the mean is shown by error bars. The thick vertical line shows the frequency of the first formant in speech, from [117].



	Pitch			
Tuning	1	2	3	4
A	57	63	37	50
B	30	53	20	47
C	63	67	47	20
D	53	63	37	30

	Pitch			
Tuning	1	2	3	4
A	**barn**	**barn**	**barn**	**barn**
B	**barn**	**barn**	ball	**barn**
C	**barn**	**barn**	**barn**	bat
D	**barn**	**barn**	**barn**	**barn**

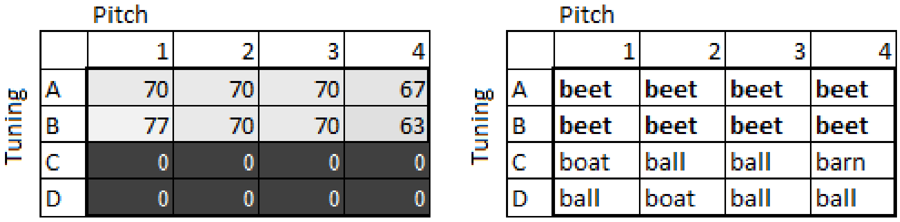
(a) /ɑ/ vowel (“barn”). (b) /ɑ/ vowel (“barn”).



	Pitch			
Tuning	1	2	3	4
A	0	0	20	3
B	7	20	17	10
C	7	3	7	0
D	17	10	17	7

	Pitch			
Tuning	1	2	3	4
A	barn	barn	barn	barn
B	barn	barn	barn	barn
C	boat	ball	ball	barn
D	boat	ball	ball	barn

(c) /u/ vowel (“boot”). (d) /u/ vowel (“boot”).



	Pitch			
Tuning	1	2	3	4
A	70	70	70	67
B	77	70	70	63
C	0	0	0	0
D	0	0	0	0

	Pitch			
Tuning	1	2	3	4
A	**beet**	**beet**	**beet**	**beet**
B	**beet**	**beet**	**beet**	**beet**
C	boat	ball	ball	barn
D	ball	boat	ball	ball

(e) /i/ vowel (“beet”). (f) /i/ vowel (“beet”).

Figure 3.5: Vowel identification results for all three vowels. Figures (a), (c) and (e) show the percentage of vowel sounds correctly identified for each pitch and tuning strategy. Lighter cell shading indicates a higher percentage. Figures (b), (d) and (f) show the most commonly chosen vowels (correct choice in bold) for each pitch and tuning strategy.

3.3.3 /ɑ/ vowel

The results for the /ɑ/ vowel are similar for preference and naturalness, with strategies with R_1 tuning (B and D) scoring highest at f_0 values below R_1, but strategies without R_2 tuning (A and B) scoring highest at higher fundamental frequencies. There is no clear relationship between tuning strategy and vowel identification. The results for the vowel identification for the /ɑ/ vowel show that at f_0 below R_1 strategy C (R_2 tuning only) scored the highest, with strategies A and D (no tuning and both resonances tuned) scoring slightly lower. Strategy B (R_1 tuning) was the most commonly mis-identified. At f_0 values above R_1 no tuning (A) was correctly identified most frequently and R_2 tuning (C) least frequently.

3.3.4 /u/ vowel

The results for the /u/ vowel do not appear to show a clear difference between the different tuning strategies for preference. However, there is some separation for naturalness with strategies with R_2 tuning (C and D) scoring highest in the middle of the f_0 range investigated. The vowel identification was generally very poor for this vowel (only 9 % correct on average). There did not appear to be a clear pattern in these results, although tuning strategies involving R_2 tuning (C and D) scored a little lower than those without (A and B) at most f_0 values. Figure 3.6 shows how this vowel was commonly mis-identified.

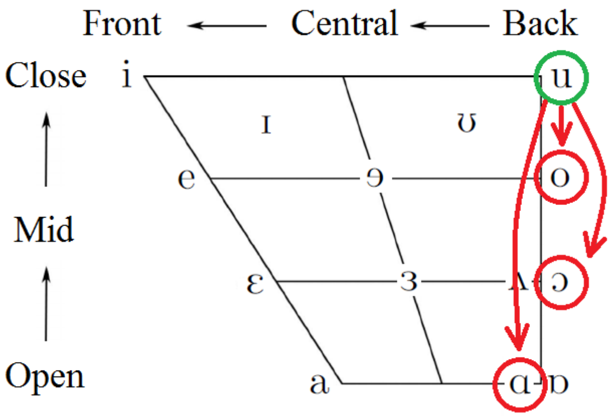


Figure 3.6: A simplified map of the IPA monophthong vowels, and the ways in which the /u/ vowel (top right) was most commonly mis-identified.

Even the untuned vowel sounds were mostly incorrectly identified for the /u/ vowel. However, the most often chosen vowel sounds were similar to the intended vowel (adjacent on the IPA diagram - Figure 3.6). Where sounds were not identified as the intended vowel, the results for preference and naturalness are still valuable, as the listener was not told the intended vowel, simply asked to choose which sound they preferred/found the most natural. Considering these results compared to the other vowels seems to suggest that the /u/ vowel (the most closed and back vowel) is unusual and perhaps fundamentally more difficult to identify or synthesise.

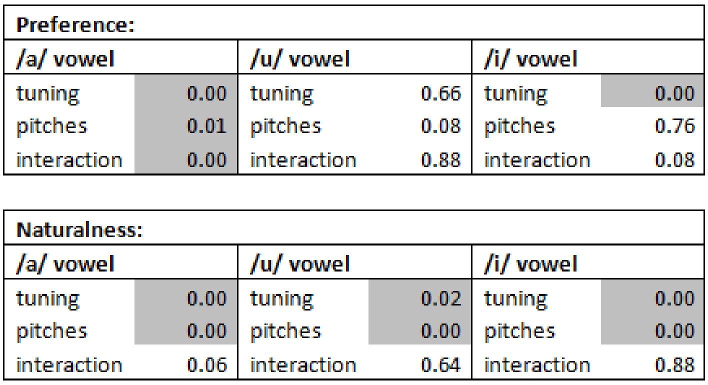
3.3.5 /i/ vowel

The results for the /i/ vowel show clearer responses than the other vowels, with strategies with R_2 tuning (C and D) scoring much higher than strategies without R_2 tuning (A and B) for both preference and naturalness. However, this effect is reversed for the vowel identification, with approximately 70 % of the vowel sounds without R_2 tuning correctly identified, but none of the vowel sounds with R_2 tuning.

3.3.6 Analysis of Variance

The results for the questions on *preference* and *naturalness* were split by vowel, and Analysis of Variance (ANOVA) was carried out in MATLAB (code included in Appendix N). The variables considered were tuning strategy (A, B, C or D) and fundamental frequency. An interaction model was used, to determine whether the variables interacted significantly.

Figure 3.7 shows the p-values for each vowel, for both preference and naturalness questions. The chosen significance level was 5 % ($p \leq 0.05$), and significant results are highlighted in grey.



Preference:					
/a/ vowel		**/u/ vowel**		**/i/ vowel**	
tuning	0.00	tuning	0.66	tuning	0.00
pitches	0.01	pitches	0.08	pitches	0.76
interaction	0.00	interaction	0.88	interaction	0.08

Naturalness:					
/a/ vowel		**/u/ vowel**		**/i/ vowel**	
tuning	0.00	tuning	0.02	tuning	0.00
pitches	0.00	pitches	0.00	pitches	0.00
interaction	0.06	interaction	0.64	interaction	0.88

Figure 3.7: The p-values from the analysis of variance (ANOVA) results for preference and naturalness questions. Significant results are highlighted in grey.

The ANOVA results for the questions on *preference* show that there was a significant difference between the results for different tuning strategies as well as different f_0 values for the /ɑ/ vowel. There was also a significant interaction between these two variables, meaning that the listeners' preference for the sounds depended on a combination of both of these attributes. For the /u/ vowel no significant results were seen, which supports what can be observed in Figure 3.3(b), that is, that there is no clear pattern in the results. For the /i/ vowel there was a significant difference between tuning strategies, but not f_0 values (and no interaction). Again, this supports the results illustrated in Figure 3.3(c), where there is a clear difference between the different tuning strategies, but no great variation in the results across fundamental frequencies.

For the naturalness results, no interaction between the variables was seen for any vowel, so the effects of tuning strategy and f_0 can be considered separately. The results for all three vowels were the same: all three showed a significant difference in naturalness, both between tuning strategies and fundamental frequencies.

Based on the results of the ANOVA, no conclusions can be drawn as to the nature of the effects of tuning strategy and f_0 on the perception of synthesised singing sounds for *preference* and *naturalness*. However, this analysis does indicate that both these variables have a significant effect on perception, although the exact relationship varies between vowels.

3.4 Discussion

In this section, the results for each vowel will be discussed together, first in respect to the preference questions, then naturalness and finally for vowel identification.

3.4.1 Preference

From Figure 3.3(a), it can be seen that for the /ɑ/ vowel, at the lower two f_0 values, strategies with R_1 tuning (B and D) were preferred above strategies without R_1 tuning (A and C). The four tuning strategies all scored similarly when f_0 was equal to R_1. However, when R_1 was above f_0 the results differed, with strategies without R_2 tuning (A and B) preferred over those with R_2 tuning (D and C). R_1 tuning only (B) scored highly across the whole range of f_0 values, which is indeed the method used most often by sopranos in this range [15]. R_2 tuning only (C) scored the lowest across the whole range of f_0 values, indicating that it was the least preferred tuning strategy. This is not surprising at lower fundamental frequencies, because R_2 tuning is rarely observed in that region. However, above the normal range of R_1 tuning, R_2 tuning has been observed, although rarely in isolation [15].

Interestingly, the results for the /u/ vowel (Figure 3.3(b)) show no significant difference in preference scores between the four tuning strategies used. There is a slight increase in score with increasing f_0 for all tuning strategies, which could simply indicate that the listeners preferred the higher-pitched sounds, or that difficulty identifying vowel sounds might play a part. The ANOVA results in Figure 3.7, support this, indicating that for preference, neither tuning nor fundamental frequency had a significant effect.

For the /i/ vowel (Figure 3.3(c)), strategies with R_2 tuning (C and D) were preferred over those without it (A and B) across all f_0 values. The second formant for this vowel is very high in speech (2790 Hz) compared to that of the other two vowels investigated (1120 Hz and 950 Hz for /ɑ/ and /u/, respectively). Therefore, when R_2 is tuned to either the first or second harmonic, this represents a considerable increase in the amount of energy in the lower part of the spectrum compared with an untuned R_2. The very high scores in preference for tuning strategies with R_2 tuning (C and D) indicate that this increase in low-frequency energy was preferred by listeners, which suggests that, in practice, listeners would prefer singers to lower the second resonance to frequencies similar to those in the other vowels.

This preference for tuned second resonances may explain why this technique is employed at very high fundamental frequencies [15] and why "sympathetic" composition takes this into account, using vowels with lower formant values at high frequencies such as an /ɑ/ vowel over an /i/ vowel [150].

3.4.2 Naturalness

From Figure 3.4(a), as for preference, it can be seen that for the /ɑ/ vowel, strategies involving R_1 tuning (B and D) were considered the most natural at f_0 values below R_1. However as f_0 rose above R_1 the perceived naturalness of strategy D (R_1 and R_2 tuning) decreased, while strategy A (no tuning) remained roughly constant, so that at higher f_0, strategies without R_2 tuning (A and B) were perceived as more natural than those with R_2 tuning (C and D). These results are surprising as they do not reflect the resonance tuning methods known to be used by singers for this vowel [15].

One possible explanation for the unexpected results for this perceptual attribute could be explained by Smith et al. [150], who suggests that listeners who often listen to a certain type of vocal production, for example classical singing, may learn to use a different "formant map" for sopranos, giving them their own categorisation of the vowel plane. Therefore, since the participants in this study were not highly trained singers or listeners and were inexperienced with opera, they found the usual resonance tuning techniques used in opera (e.g. $R_1 : f_0$) unnatural in general. The use of synthesised samples could also have a confounding effect on this question, although this is consistent between all the samples, so the *relative* naturalness of the tuning methods can still be considered. In addition to this, "naturalness" is of course a subjective term, and in this experiment the participants were left to decide for themselves what it meant, so there may have been some variation in this between participants.

For naturalness, as for preference, all four tuning strategies scored similarly for the /u/ vowel (Figure 3.4(b)). There was, however, some separation for the middle two f_0 values, with strategies involving R_2 tuning (C and D) scoring a little higher than those without (A and B). This is supported by the ANOVA results (Figure 3.7), which show that, for naturalness, both tuning and fundamental frequency had a significant effect.

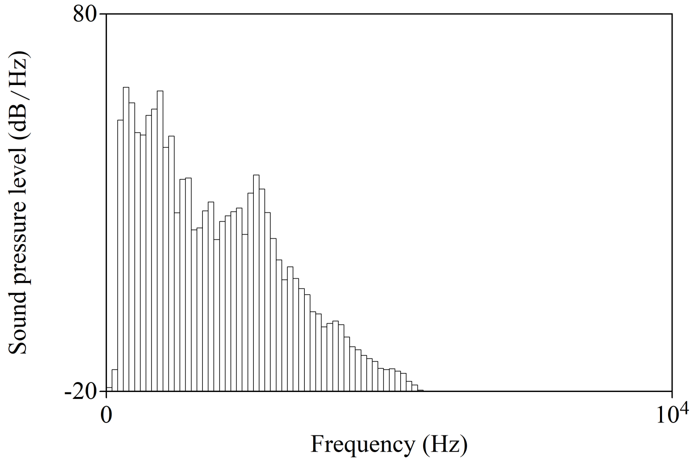
The results for both the preference and naturalness questions for the /i/ vowel are somewhat unexpected, considering that R_2 tuning in isolation at these fundamental frequencies has not often been observed [15, 151]. However, these results must be considered in conjunction with the vowel identification results, in that the participants were simply asked how natural the sounds were, but not told which vowel sounds they represented. It seems that the participants found the sounds with R_2 tuning more preferable and natural than those without, but had difficulty identifying them as an /i/ vowel. This may perhaps be due to the participants' lack of experience with opera and singing in general.

For the /i/ vowel (Figure 3.4(c)), tuning methods involving R_2 tuning (C and D) consistently scored the highest, followed by those without (A and B). The average scores for naturalness remained fairly stable at all f_0 values and, again, a general increase in naturalness with f_0 was seen. As for preference, these results suggest that lowering the high second resonance has the greatest effect on naturalness, irrespective of whether R_1 is tuned.

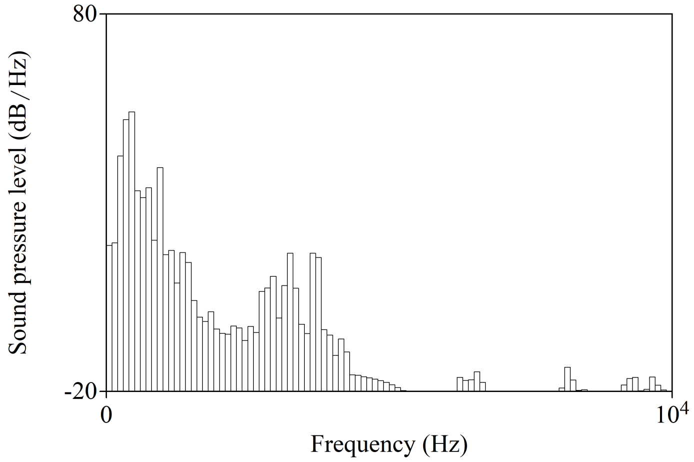
3.4.3 Vowel Identification

In terms of vowel identification, the results for the /ɑ/ vowel (Figure 3.5(a)) show that at f_0 values below R_1, strategy C (R_2 tuning) scored the highest, with A and D (no tuning and both resonances tuned, respectively) just below. Strategy B (R_1 tuning) was the most commonly mis-identified. At f_0 values above R_1 this pattern changed to a completely different order (similar to preference and naturalness) with A the most correctly identified and C the least. The average percentage of sounds correctly identified across all f_0 values and tuning strategies was 46 % (with a standard deviation of 16 %).

The results for the /u/ vowel (Figure 3.5(b)) show that this vowel was correctly identified much less frequently than the /ɑ/ vowel (only 9 % correct on average, with a standard deviation of 7 %). There did not appear to be a clear pattern in these results, although tuning strategies involving R_2 tuning (C and D) scored a little lower than those without R_2 tuning (A and B) at most f_0 values. This could be due to the importance of the position of the second formant in distinguishing this vowel, meaning that at all f_0 values, tuning of R_2 distorted the vowel sound. Tuning strategies A and B were most commonly identified as an /ɑ/ vowel across all f_0 values, whereas strategies with R_2 tuning (C and D) were most commonly identified as /o/ (as in "boat") at the lowest f_0, /ɔ/ (as in "ball") at the middle two f_0 values and /ɑ/ at the highest f_0. This suggests that tuning R_2 causes the vowel to sound more open (see Figure 3.6). However, the poor identification of even the untuned sample suggests that there may have been issues with the synthesis of this vowel sound. Figure 3.8(a) shows a long-term average spectrum (LTAS) for the untuned synthesised /u/ vowels (top). Figure 3.8(b) shows the LTAS of subject 1 from the main experiment (Chapters 5 and 6) singing the same vowel sound on a similar fundamental frequency. It can be seen that the real sample includes more high-frequency energy, which may explain the poor identification of vowels in this study. However, spectral accuracy was not the primary focus of this thesis.



(a) LTAS of the synthesised untuned /u/ vowel



(b) LTAS of subject 1 singing an /u/ vowel at a similar fundamental frequency

Figure 3.8: LTAS of both the synthesised /u/ vowel, and a real singer singing the same vowel. Averaged over all four synthesised pitches.

The results for the /i/ vowel (Figure 3.5(c)) show a very clear pattern, where strategies without R_2 tuning (A and B) were correctly identified in around

70 % of vowel sounds (with a standard deviation of 4 %), however, strategies with R_2 tuning (C and D) were never correctly identified. One explanation of this might be provided by Benolken [138], who suggests that some vowels which have similar first formant frequencies, like the /i/ and /u/ vowels (only 60 Hz apart), are differentiated by their second formants, so altering the second formant results in a dramatic loss in identifiability. The sounds with R_2 tuning (C and D) were most commonly identified as /ɔ/ (as in "ball"), /o/ (as in "boat") or /ɑ/ (as in "barn"), implying that the perceived vowel sound changed from closed to open (see Figure 3.6).

3.4.4 Overall impressions

There were marked and unexpected differences between the results for the three vowels for the three perceptual attributes investigated. The /i/ vowel produced the most notable differences across tuning strategies for all three perceptual attributes, with strategies involving R_2 tuning scoring the highest for both preference and naturalness, but the lowest for vowel identification.

Based on the findings of previous studies [14, 125, 152], it was predicted that the strategy with no resonance tuning (A) would score the highest for all three of the perceptual attributes investigated at fundamental frequencies *below* the first resonance, as there is little evidence of singers using resonance tuning within this frequency range. However, the opposite of this was found: at f_0 values below R_1, strategy A was generally one of the lowest scoring, whereas strategy D (both resonances tuned) scored highly for both preference and naturalness. The results therefore suggest that for certain vowel sounds, if physically possible, it might be perceptually beneficial to employ resonance tuning over a wider range of fundamental frequencies than had previously been thought. At fundamental frequencies below the first resonance, *lowering* R_1 slightly to coincide with the fundamental would increase the acoustic power transmitted, therefore reducing the effort required by a singer to communicate effectively to an audience. This has not yet been observed in practice, however, which suggests either that it is not possible (e.g.

physiologically), or that it is undesirable.

At fundamental frequencies above R_1, it was expected that R_1:f_0 tuning (strategy B) would score highly for all three perceptual attributes, as this is the most commonly observed in practice [14, 15] and R_2:$2f_0$ tuning (strategy C) would score the lowest, as it is rarely observed in isolation [14]. Indeed, Wolfe [140] suggests that R_2 tuning might be unintentional, based on the theory that as the fundamental frequency rises, R_1 is tuned to the fundamental by increasing the opening of the mouth, and as both R_1 and R_2 rise with increased mouth opening, R_2 is raised as a side effect of raising R_1. This would suggest that R_2 tuning in isolation (C) should score quite low for both preference and naturalness. However, for some vowels and f_0 values this was not the case. For example, for preference, R_2 tuning (C) scored highly for the /i/ vowel.

An interesting pattern seen in the results is that the strategies seemed to “pair up” for most of the perceptual attributes, with strategies without R_2 tuning (A and B) behaving similarly, and strategies with R_2 tuning (C and D) also behaving similarly. This seems to suggest that the presence or absence of R_2 tuning had the greatest influence on the listeners’ perception of the sounds. Further investigation is required to fully understand this result, perhaps involving spectral analysis to compare the synthesised vowel sounds to real vowels.

Although most previous studies have focussed on single vowels (most commonly /ɑ/), this study found that the rankings of different tuning strategies are highly dependent on the vowel, as extremely different patterns are observed across the three vowels investigated, /ɑ/, /i/, and /u/. In addition to this, resonance tuning (by any of the three strategies investigated here) does not necessarily improve the *preference*, *naturalness* or *vowel identification*, as in some cases strategy A (no tuning) scored the highest, even at fundamental frequencies above R_1. For example, for the /i/ vowel, no tuning (A) scored lower than the other tuning strategies for naturalness and preference, but improved the *vowel identification.* In addition to this, some tuning strategies might improve one perceptual quality, whilst having little effect on or detracting from another quality. For example, R_1 tuning alone (B) scored poorly for both preference and naturalness for the /i/ vowel, but resulted in good vowel identification.

This suggests that choosing the most appropriate resonance tuning techniques is a balancing act for the singer, as they must tailor the resonances of their vocal tract according to their performance aims and decide whether to prioritise a pleasing voice quality over the clarity of the text in a particular situation, or perhaps sacrifice some naturalness to achieve a higher SPL in another. Deciding when and how to use resonance tuning is therefore an exercise in compromise in terms of performance for the ease of the singer and what the listener might hear. The practical implications of the findings of this experiment, however, hinge on the assumption that singers are capable of controlling their vocal tract resonances with great precision: an interesting point worthy of further research.

Limitations

The compression (using mp3) of the audio samples presented online would have slightly reduced the quality of the samples. However, since this was the same for all samples, the comparison between different tuning methods is not substantially affected. Ideally, the participants would have listened to uncompressed audio in identical listening conditions. However, as discussed in section 3.2.2, the decision to distribute the test online made this impossible. The conditions were, however, consistent between samples for each participant, allowing individuals to appropriately compare the different resonance tuning strategies.

Another point worthy of consideration is the possible variation in regional accents in the participants, which may have affected their perception of the synthesised vowel sounds. No guidance was given on the intended accent of the samples; participants were instructed to “try to remember that the voice is meant to be singing, not speaking, and that it is imitating a classical/opera type voice." There is a generally-accepted pronunciation in opera. However, not all the participants were familiar with opera, so it is possible that this may have introduced some variation into the results.

As is often the case, it would have been desirable to have had more participants in the listening test, to improve the statistical robustness of the findings presented. The listeners were also mostly male and non-experts, which may have impacted on their perception of the singing samples. It would have been particularly interesting to have had more expert listeners, to compare their perception of different resonance tuning techniques to that of the non-experts.

3.5 Conclusion

The experiment that forms the basis of this chapter investigated the impact of specific resonance tuning techniques on perception through a perceptual test which compared synthetic vowel sounds. Although recorded samples would provide better context for the listeners, the main advantage of synthesised sounds is that this allowed the resonances to be directly manipulated and controlled.

The hypothesis that the strategies used most frequently in practice, such as $R_1{:}f_0$ and $R_2{:}2f_0$ tuning, would be preferred by subjects and perceived as more natural and more accurately identified, does not seem to be supported by these results. Rather, the results show no clear patterns or trend, suggesting that the perception of resonance tuning is a highly complex issue, which must take into account a variety of acoustic factors.

The results for the three vowels show very different trends; this was expected to an extent due to the differences in formant values between the different vowels. Previous work [14] has shown that the resonance tuning strategies used by singers varied considerably between different vowels. Since the use of resonance tuning is so different between the different vowels, it is unsurprising that the perception of different resonance tuning strategies is also highly dependent on vowel.

The fact that the results for perception and naturalness for the /i/ vowel did not show dependence on fundamental frequency also ties in with the pattern of resonance tuning observed in Henrich et al [14], which showed that subjects used the same resonance tuning strategies over (nearly) the whole frequency range investigated.

The results from the vowel identification part of this experiment support theories that vowels sung at high fundamental frequencies tend perceptually towards an /ɑ/ vowel, as vowels were most commonly mis-identified as the /ɑ/ vowel.

These findings bring to light some of the complex relationships between different resonance tuning strategies and perceptual attributes, and the different requirements of different vowels. The results show no general patterns for the perception of the different tuning strategies investigated, which appears to be highly dependent on the vowel synthesised. This suggests that, in practice, resonance tuning is likely an exercise in compromise for a singer, as employing a certain resonance tuning strategy might improve one perceptual attribute whilst worsening another.

Chapter 4

Resonance Tuning in Girl Choristers

This chapter presents a preliminary experiment carried out to investigate resonance tuning in girl choristers, a group for which this has not previously been studied. Vocal tract resonances are measured using pseudo-broad band noise excitation. The results of this preliminary experiment inform the methodology of the main experimental work of this thesis, where this measurement technique is used to investigate resonance tuning in adult soprano singers.

The subjects in this experiment were three girl choristers recruited from local cathedral and church choirs. This group of subjects was chosen as their vocal ranges are very close to those of adult sopranos, and they were readily available to participate in the experiment.

Wide-band excitation at the subjects' mouths was used to measure their vocal tract resonances during singing, and vowel formant values in speech were extracted from recordings of spoken text. Measured resonance values were cross-referenced with first and second harmonics for sung vowels across the subjects' ranges to identify the resonance tuning techniques employed. These results were compared with those previously observed by others in professional adult classical singers.

4.1 Introduction

Both male and female children have similar vocal ranges to adult females; Moore [153] found that children had vocal ranges from approximately G3 to G5 (196-831 Hz), while adult females had a range of approximately E3 to F5 (165-698 Hz). Therefore, it can be assumed that similar difficulties in both the production and the perception of singing at high fundamental frequencies will be observed in children's singing as in that of adult sopranos.

Since children have shorter vocal tracts than adult females, their vocal tract resonances are higher by approximately 20 % [152, 154]. However, because children are also generally able to sing at a higher range of pitches, it is to be expected that they will encounter similar effects to adult female singers, as there will be a part of their range when the fundamental frequencies become higher than one or both of the first two resonances. There exists very limited research on young singers, although Barlow et al. [48] investigated closed quotient in adolescent girls.

In previous studies on adult singers [14, 15], evidence of R_1 and R_2 tuning has been observed in the upper part of the soprano range. These tuning techniques were found to be employed extensively by professional singers and also by advanced amateur singers and, to some extent, by non-expert singers (one had trained but had not sung for 7 years, the other three had experience in choirs and two of them had had some singing lessons).

The detection of resonance tuning in non-expert singers in the Garnier study [15], who all employed R_1:f_0 tuning over some part of their range, raises the question of whether resonance tuning is a technique unique to trained adult singers and learned with singing training and experience, or one which is also employed by experienced young singers.

As discussed in sections 2.2.3 and 2.2.1, it is difficult to measure the frequencies of the vocal tract resonances in higher voices, where the harmonics of the voice source are so widely spaced. However, it is possible to extract information about the resonances of the vocal tract using external noise excitation, such as swept-sine or broad band noise. Using an external sound source to excite the vocal tract during phonation allows an accurate picture of the vocal tract transfer function to be generated. It causes little inconvenience to the singer, compared to methods injecting sound into the vocal tract or requiring singers to maintain vocal tract shapes.

The method of vocal tract resonance measurement used in this preliminary experiment is the method proposed by Epps et al. [95] and used by others including Garnier et al. [15], Henrich et al. [14] and Joliveau et al. [58]. It involves exciting the vocal tract at the mouth with a synthesised broad band signal consisting of harmonics spaced approximately 5 Hz apart (referred to in this thesis as "broad band noise"). The use of harmonic components, as opposed to true broad band noise such as white noise, allows a "calibration" procedure to be carried out, where the subject keeps their mouth closed while the signal is played and recorded at the mouth, and the amplitudes of the different frequency components are adjusted individually to ensure that the excitation signal produces a flat frequency response when recorded. The measurement is then taken by exciting the vocal tract with this "calibrated" signal and the transfer function calculated. This can be assumed to be representative of normal singing, since the subject can sing normally while the measurement is being taken, rather than being asked to hold an artificial position. This method is also more robust to small movements of the subject than swept-sine excitation (as discussed in section 2.2.3).

The main focus of this chapter will be the protocol used and the quality of the results obtained by taking measurements of vocal tract resonances using broad band noise excitation. The use of resonance tuning in young singers will also be considered.

4.2 Method

4.2.1 Subjects

The subjects for this experiment were three female choristers from York Minster and one chorister from a well-respected local church choir (aged 13-15). All three took singing lessons and performed regularly; the Minster choristers, 6 days per week; the Church chorister, at least twice per week. Details of the choristers are summarised in Table 4.1. Older choristers were chosen (York Minster choristers are aged around 7-13 years) so that the effects of experience and training were most likely to be observed, based on the current understanding of increased resonance tuning with singing experience in adults [15].

Older female choristers, who may have started to undergo physical changes due to puberty (typical age of puberty onset is 11 years in girls [155]), were chosen so that any resonance strategies employed could be compared with those of their adult counterparts. The subjects chosen all reported vocal ranges up to around A5.

Subject	Age (years)	Choir	Years as chorister	Singing lessons	ABRSM grade exams[1]
1	14	York Minster	6	3.5 years	Grade 5 Singing
2	15	St Olave's Church, York	9	3 years	Grade 5 Singing, Grade 5 Clarinet, Grade 5 Piano
3	13	York Minster	5	1.5 years	Grade 4 Singing, Grade 3 Clarinet, Grade 2 Piano

Table 4.1: Details of the choristers' ages and singing experience.

---

[1] The Associated Board of the Royal Schools of Music (ABRSM) is an examinations board and registered charity based in London, UK. (http://gb.abrsm.org/en/home).

4.2.2 Resonance detection

The method used in this study to measure the vocal tract resonances (initially developed by Epps et al. [95]) involved exciting the vocal tract at the mouth with a synthesised broad band noise signal and recording the response with a lavalier microphone also placed at the subject's mouth, as shown in Figure 4.1.

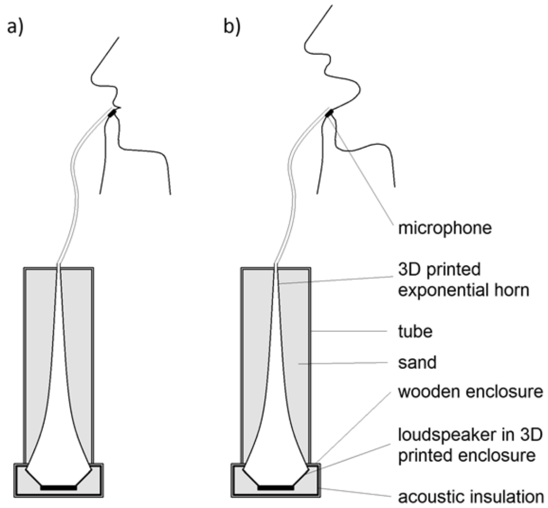


Figure 4.1: The equipment used to simultaneously play and record a signal at the subject's mouth using a 3D-printed impedance-matching horn and a microphone. The impedance-matching horn is mounted in a wooden enclosure filled with sand. The flexible tubing allows the subject to position the acoustic source and microphone on their bottom lip.

The excitation signal consisted of 606 harmonics spaced 5.38 Hz apart[2] (synthesised by adding sine waves), from 250 Hz to 3500 Hz. Their phases were randomised to improve the signal-to-noise ratio (as investigated by Smith et al. [156]). The set-up consisted of a loudspeaker with an impedance-matching horn contained within a sand-filled wooden enclosure. A flexible tube was used to deliver the sound to the subjects' mouth. Although the

---

[2] The frequency resolution of 5.38 Hz was a result of the 44.1 kHz sampling frequency, and 2^13-point FFT, which was found to be sufficient for observing the vocal tract resonances.

equipment was a little heavy, the length of the tube meant that, once positioned, it did not need to be moved and the subject could simply stand next to it. The flexible tube was held by the subject, who lightly pressed it against their bottom lip. A calibration measurement was first carried out, measuring the pressure response at the mouth with the subject's mouth closed (P_{closed}), then taking an 8192-point Fourier transform of the signal, and adjusting the amplitudes of the frequency components individually to flatten the magnitude response of the signal from the microphone at the subject's mouth. This calibrated signal was then used as the excitation signal for the measurements taken while the subject sang the required note (P_{open}) (see Epps et al. [95]).

The ratio P_{open}/P_{closed} is a measure of the ratio of the impedance of the vocal tract to that of the radiation field, as the source approximates an ideal current source [95]. The spectrum of the signal recorded at the subject's mouth therefore shows the harmonics of the voice source superimposed on the approximate transfer function of the vocal tract. The amplitude of the excitation signal (for all subjects) was approximately 96 dB SPL, which introduced sufficient acoustic energy to generate a transfer function accurate enough to allow the researchers to deduce the resonance values, but was low enough to allow the subject to hear themselves, to cause minimal interference.

As the calibration procedure relied on the random phases of the frequency components, it did not always generate a signal with a sufficiently flat frequency response. It was therefore sometimes necessary to repeat the calibration procedure several times to get a sufficiently frequency-independent calibration signal. This however only took a few seconds and, once a suitable calibration was obtained, did not need repeating until the subject stopped holding the tube against their lip, so was not a great inconvenience.

4.2.3 Protocol

Subjects were asked to answer a short questionnaire about their singing experience, read an information sheet explaining the nature of the experiment, and sign a consent form (included in Appendix B). Due to their ages, the subjects in this study were accompanied at all times by a parent/guardian, who did not participate in the experiment.

The subjects were fitted with a head-mounted microphone (DPA4066, DPA microphones) approximately 4 cm from the lips, which was used to record the speaking task required to find the formant values in speech. A second microphone (Earthworks M30, Earthworks microphones) was placed approximately 1 m from the subject. This was used to record the entire experiment for later reference. These signals were recorded simultaneously during the entire experiment, using a Tascam DR680 portable digital recorder. The procedure took place in an anechoic chamber.

The first task for the subject, was to read a short text in their normal speaking voice. They were given a practice attempt to familiarise themselves with the text (without broad band noise excitation). For the second task, the subject was asked to sing individual notes into the wide-band vocal tract measuring device, each on one breath, in an ascending chromatic sequence (12 notes per octave) from C4 to the top of their range, for three vowel sounds (/ɑ/, /u/ and /i/) (see Appendix A for IPA vowels). They were played each note on an electric piano before singing and required to hold each note for approximately 6 seconds. The subject was asked to sing in their normal singing voice, as if they were singing a solo with their choir, at a medium level and keeping their mouth shape constant. They were reminded if necessary during the tasks. Notes were only repeated if the measurement was insufficient or if the subject failed to maintain the note until the end of the measurement.

Data collected consisted of: (1) answers to a questionnaire to determine the subject's level of singing and their training; (2) their acoustic speech and singing recordings, and (3) recordings made by the broad band vocal tract measuring device.

Prior ethical approval was gained from the Physical Sciences Ethics Committee at the University of York (see Appendix J).

4.3 Results

4.3.1 Data Analysis

As in previous works [14, 15], the frequencies of the vocal tract resonances were measured manually from the plot of P_{open}/P_{closed} against frequency by one author, and then checked by another researcher. This was done in MATLAB [68], using a graphical user interface, with a frequency resolution of 5.38 Hz. An example plot of P_{open}/P_{closed} against frequency is shown in Figure 4.2. In some cases, for example when the subject did not remain completely still while singing[3], it was not possible to accurately identify the vocal tract resonances. In these cases the data were omitted from the results. The percentages and numbers of measurements omitted for R_1 and R_2 for each vowel are shown in Table 4.2.

Vowel	/ɑ/	/u/	/i/
R_1	10.8 % (9/83)	9.7 % (7/72)	14.5 % (11/76)
R_2	2.4 % (2/83)	0.0 % (0/72)	27.6 % (11/76)

Table 4.2: The percentage (and number) of measurements omitted for each resonance, for each vowel

Perhaps unsurprisingly, the measurement method was more effective for the open vowels investigated (/ɑ/ and /u/) than for the closed vowel (/i/). This can be deduced from the number of measurements excluded (see Table 4.2); significantly more measurements for the /i/ vowel were excluded than for either of the other two vowels, including over a quarter of the R_2 measurements. This is likely to be caused by the small lip opening of closed vowels not allowing enough acoustic energy to enter the vocal tract to produce a measurement. In spite of this, a sufficient number of results of suitable qual-

---

[3] In some cases this could be identified by observing the subject; however different types of movements produced characteristic errors in the transfer function obtained. If the subject altered the position of the tube and microphone during measurement, this introduced a wave-like error to the transfer function. A change in fundamental frequency caused a "smearing" of the harmonic peaks, and movements of the articulators caused widening of the resonance peaks.

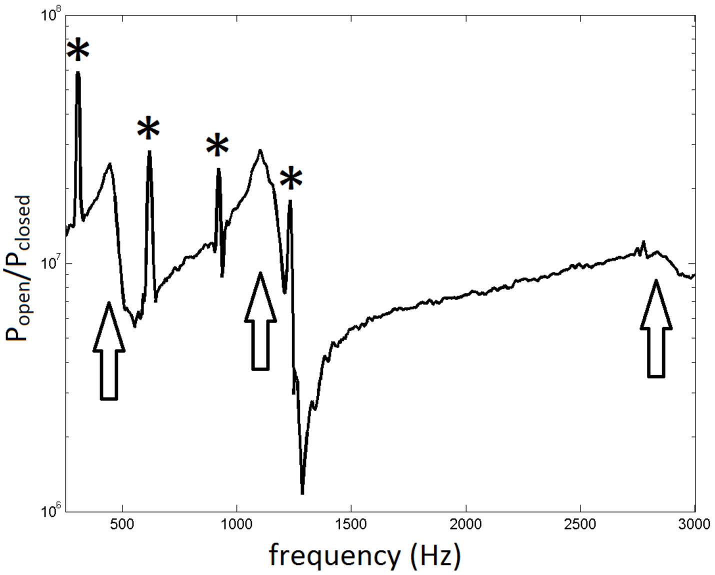


Figure 4.2: A plot of P_{open}/P_{closed} against frequency for an /u/ vowel. The first four harmonics are marked with asterisks, and the resonances are marked with arrows. These were measured manually as described above.

ity were still obtained for the /i/ vowel. The broad band noise measurement method is therefore still suitable for future research on the /i/ vowel, as long as the protocol allows the results to be viewed immediately, so that rejected measurements could be repeated.

4.3.2 Speech formant measurements

As the fundamental frequencies in speech were significantly lower than for most of the sung notes investigated in this experiment, it was possible to calculate the frequencies of the formants in speech (which are assumed to be equivalent to the vocal tract resonances) by analysing a spectrum of the spoken voice for each vowel sound. The recordings of the choristers reading a short text were used for this purpose. The average reading rates of the choristers were 234, 242, and 174 words/minute respectively. Samples of the relevant vowels were extracted from the audio signal by hand: the samples chosen were from the middle of the vowel, where the formant values remained approximately constant across the entire sample, and had an average length of 0.2 seconds. These samples were then analysed using *Praat* [70], by plotting a spectrogram and using the built in formant detection function to extract formant values, and then collated using *Microsoft Excel* [157]. Between two and six values were obtained for each formant measurement, and the average and standard deviation of these were calculated and shown on the left sides of Figures 4.3(a)-(c).

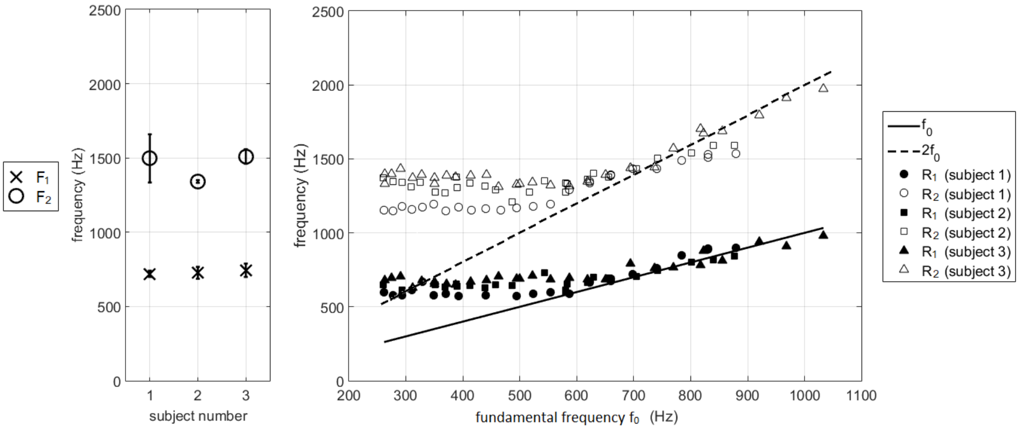
4.3.3 Resonance tuning

The measurements of the vocal tract resonances R_1 and R_2 obtained for all three subjects (as described in section 4.2.2) were plotted against frequency (separately for each vowel), and can be seen on the right of Figures 4.3(a)-(c). The first and second harmonics are represented as solid and dashed lines, respectively. The vocal tract resonances of the first subject are represented by circles, the resonances of the second by squares, and the resonances of the third by triangles.

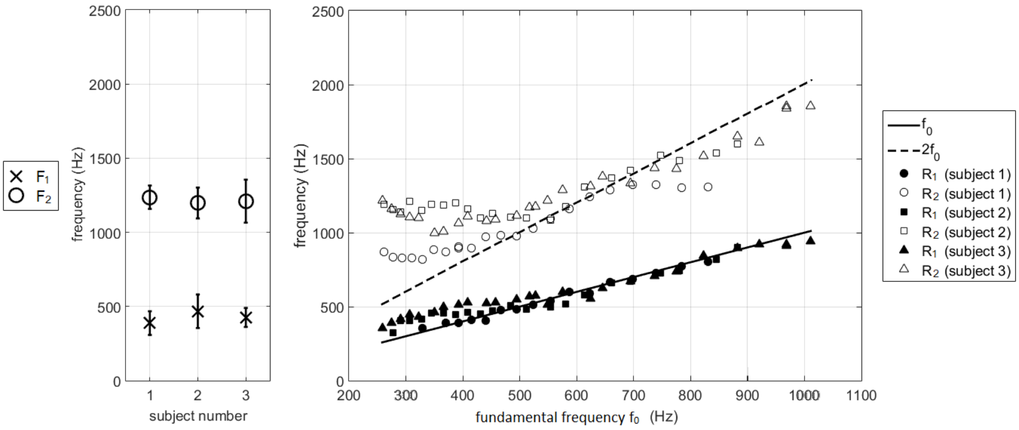
The average formant values in speech were also plotted against frequency, for all three subjects, for each vowel. These can be seen on the left of Figures 4.3(a)-(c).

In order to maintain a perceptually constant pitch interval, the most logical way to define resonance tuning would be using a frequency-dependent interval such as a fraction of a tone. However, it was decided in this work to use the same criteria for determining resonance tuning as in previous works, so that the results would be directly comparable.

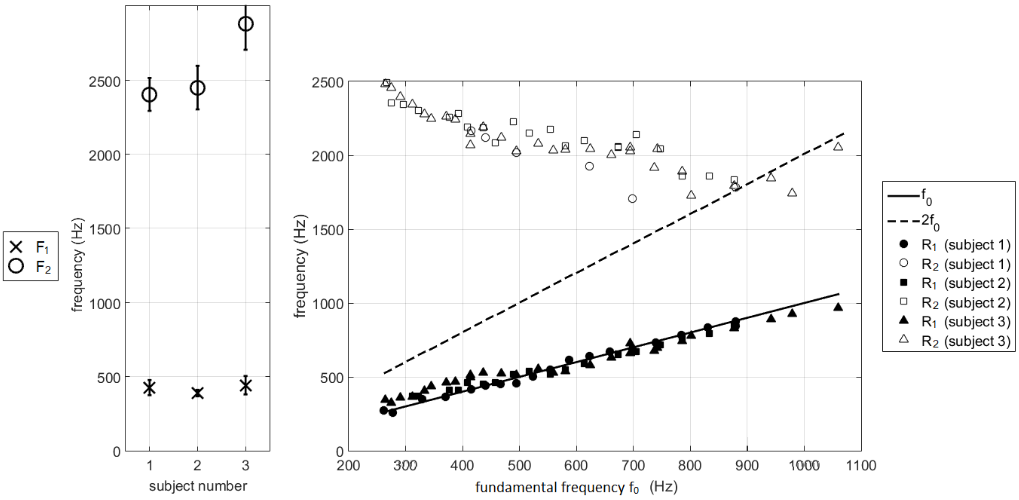
Henrich et al. [14] identified resonance tuning in adult voices by plotting a histogram of the *difference* in frequency between the first resonance and the fundamental frequency (R_1 - f_0). In their study, a central peak approximately 50 Hz wide was found, so they defined resonance tuning as occurring when R_1 was within 25 Hz of f_0.



(a) /ɑ/ vowel



(b) /u/ vowel



(c) /i/ vowel

Figure 4.3: The first (crosses) and second (circles) formants in speech for each subject are shown on the left side, and the resonances against fundamental frequency for all the subjects singing three different vowels (the solid line represents f_0, and the dashed line represents $2f_0$) are shown on the right side.

A similar technique was adopted in this experiment, and Figure 4.4 shows the difference between R_1 and f_0 (R_1 - f_0) for all measurements of R_1 (left). As expected, it shows a strong correlation between R_1 and f_0. It can be seen that there is a much broader central peak in R_1 measurements around f_0 than those in Henrich et al. [14], and that this is approximately 140 Hz wide. In this experiment therefore, resonances were assumed to be "tuned" to a harmonic when they fell within 70 Hz of them. The resonance measurements were inspected using Microsoft Excel, to look for four types of resonance tuning: R_1:f_0, R_1:$2f_0$, R_2:f_0, and R_2:$2f_0$.

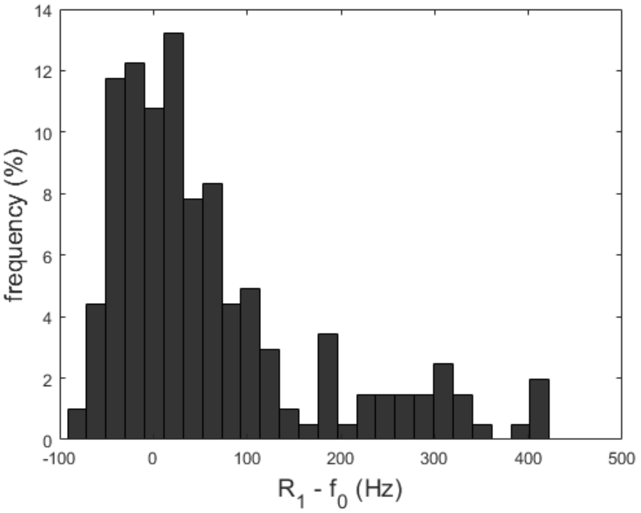
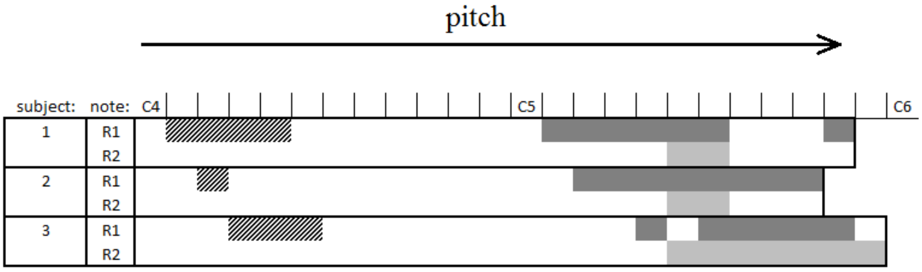
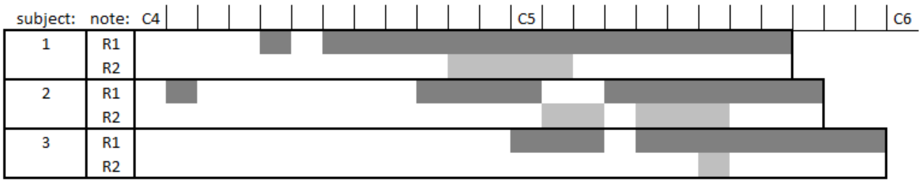


Figure 4.4: Histogram showing the distribution of the difference in frequency between the measured values of R_1 and f_0 (R_1 - f_0).

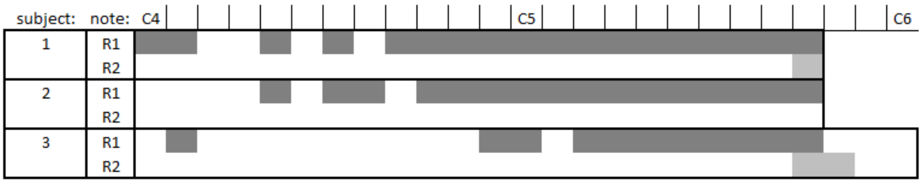
These results are illustrated in Figure 4.5, which show the extent of resonance tuning used by each of the three subjects, across the entire fundamental frequency range sung, for the three vowels investigated. This Figure indicates that all three choristers employed R_1:f_0 tuning over a wide range of fundamental frequencies, and some R_1:$2f_0$ and R_2:$2f_0$ tuning, and that the pattern of resonance tuning is highly dependent on the vowel sung.



(a) /ɑ/ vowel



(b) /u/ vowel



(c) /i/ vowel

Figure 4.5: The resonance tuning strategies employed by each subject, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom). The tuning strategies investigated were $R_1 : f_0$ tuning (dark grey), $R_1 : 2f_0$ tuning (dark stripes), $R_2 : f_0$ tuning (not observed) and $R_2 : 2f_0$ tuning (light grey).

It can be seen from Figure 4.3(a), that for the /ɑ/ vowel, at the lower end of the frequency range investigated, the formant values remained roughly constant at approximately the same values as in speech. Figure 4.5 shows that the subjects tuned R_1:$2f_0$ at the lower end of the range investigated, and then began tuning R_1:f_0 when f_0 was between 4 and 2.5 tones below R_1, and continued this to the upper limit of their ranges (with the exception of one note for subject 3). R_2:$2f_0$ tuning was less consistent, beginning when R_2 was 1-4 tones[4] below $2f_0$, and extending over a range of 3 tones for subjects 1 and 2, but continuing to the top of her range for subject 3. R_1:$2f_0$ tuning

---

[4] A “tone” here is defined as 2 semitones, where there are 12 semitones in an octave.

was only observed for this vowel.

Figure 4.3(b) shows that for the /u/ vowel, at low frequencies, the formant values were again similar to those in speech, and again all the subjects tuned R_1:f_0, this time over the whole range of the fundamental frequencies investigated. In addition to this, R_2:$2f_0$ tuning was employed by all three choristers to some extent; from a range of 4 tones for the first subject to 1.5 tones for the third subject.

For the /i/ vowel, all the subjects tuned R_1:f_0 over almost the entire range of fundamental frequencies investigated (although subject 2 had a gap of around 3 tones around F4); however, only a small amount of R_2:$2f_0$ tuning was seen at the upper limit of the range investigated for subjects 1 and 3, and none for subject 2. For this vowel Figure 4.3(c) shows that the value of F_2 in speech, and R_2 at low frequencies was very high, and although it steadily reduced with increasing frequency, R_2 only approached $2f_0$ at the very highest frequencies investigated.

4.4 Discussion

In general, this experiment showed that broad band noise excitation was a very effective method for measuring vocal tract resonances. The subjects generally found the sensation of the broad band noise excitation a little strange, but became very comfortable using it after a practise session. The practise session was also very helpful to the subjects, as none of them had been in an anechoic chamber before, so this allowed them time to acclimatise, as well as learning to use the equipment. None of the subjects expressed any feelings of discomfort and all enjoyed the experiment.

The plots of P_{open}/P_{closed} against frequency produced for each measurement clearly showed the frequencies of the vocal tract resonances to a suitable degree of accuracy, and the MATLAB GUI used allowed the researcher to observe immediately if a result was unsatisfactory, which allowed it to be repeated without causing significant interruption to the procedure. The number of measurements omitted was small enough to provide information about resonance tuning in choristers across the entire frequency range investigated.

To improve the robustness of these results, it would have been desirable to collect several repeated sets of data. However, the current protocol took around 45 minutes on average and it became apparent that a longer experiment would have been tiring to the participants. An alternative would have been to use more participants, but these were unavailable at the time and unnecessary for a preliminary experiment.

4.4.1 Resonance tuning behaviour in Choristers

In this experiment, the method of measuring vocal tract resonances was tested on young singers - experienced girl choristers. Figure 4.3 shows a distinct pattern in the resonance measurements for all three vowels, and Figure 4.5 shows clear evidence of R_1 tuning (to both f_0 and $2f_0$) in all three of the choristers for all the vowels. Evidence of R_2 tuning was seen in both the /ɑ/ and /u/ vowels, but very little in the /i/ vowel, suggesting that the extent to which the different resonance tuning techniques were used was highly vowel specific.

For the /ɑ/ vowel (seen in Figure 4.3(a)), the pattern of resonance tuning used was similar to resonance tuning in adult singers. All three choristers maintained approximately constant values of both R_1 and R_2 until they became close in frequency to f_0 and $2f_0$, and then tuned them to the relevant harmonic until near the tops of their ranges.

The resonance tuning observed for the /u/ vowel (Figure 4.3(b)) was very similar to that of the /ɑ/ vowel, with R_1:f_0 tuning beginning a little earlier. The range of R_2 tuning varied a great deal between subjects, with subject 1 employing second resonance tuning over a range of 4 tones, but subject 3 over only 1.5 tones.

For the /i/ vowel, it can be seen from Figure 4.3(c) that all three choristers employed R_1:f_0 tuning over a very wide range of the fundamental frequencies investigated, whereas there was little evidence of any R_2:$2f_0$ tuning at all. For all the choristers R_2 fell as f_0 increased. Only one chorister was able to sing at a fundamental frequency above 900 Hz, and some evidence of R_2:$2f_0$ tuning is seen here, however, since the other choristers were not able to reach higher fundamental frequencies, it is not known if they might also have employed R_2:$2f_0$. Because the frequencies of R_2 in speech fell between 2404 and 2880 Hz for this vowel, for f_0 to reach R_2 it would need to be approximately 1400 Hz, which would be unlikely for any chorister. It is unsurprising therefore that this method of resonance tuning was not observed in this experiment.

This difference in tuning techniques between vowels was also seen in the study by Henrich et al. [14], which investigated four vowel sounds: /ɑ/, /ɜ/, /ɔ/, and /u/. As in this study, the sopranos showed a similar range of R_1 tuning, but a greater range of R_2 tuning for the /u/ vowel than for the /ɑ/ vowel.

The frequency range over which resonance tuning was employed on an /ɑ/ vowel in this study was approximately the same as the non-expert singers in the studies by Garnier et al. [15] for subjects 1 and 2, although subject 3 employed the same techniques over a slightly smaller range. However, it is important to realise that the criteria for determining resonance tuning are different in this study. Resonance tuning is said to occur when a resonance is within 70 Hz of a harmonic (the left part of Figure 4.5). This extends the wide range of resonance tuning observed.

It seems therefore reasonable to conclude that the resonance tuning techniques used by the choristers were similar to those observed in adult singers (in the Garnier et al. study), although they did not tune their resonances as *closely* to the relevant harmonic as the adults. There were no great differences in resonance tuning behaviour between the three choristers but as expected, based on speech formant values, differences between the vowels were observed.

4.5 Conclusion

This chapter has presented a preliminary experiment to evaluate the suitability of measuring vocal tract resonances using broad band noise excitation. It has been shown to be highly effective. The measurements of vocal tract resonances obtained in this experiment were generally of good quality, allowing clear observation of resonance tuning to within 5.38 Hz.

The protocol allowed for data to be collected across the entire vocal range of the subjects, was minimally intrusive and not tiring or unpleasant for the singers despite their youth and inexperience. Singing a scale on each vowel in turn is a familiar exercise for a singer and the real-time feedback available to the researcher allowed unsatisfactory results to be repeated with minimal disturbance.

Clear evidence was observed of R_1:f_0, R_1:$2f_0$ and R_2:$2f_0$ resonance tuning by the choristers in this experiment, which is as expected, considering the values of the formants in speech. This was comparable with the resonance tuning techniques used by non-expert adult singers, although the subjects in this study did not tune their resonances as close to the relevant harmonic as the adults.

Further investigation would be beneficial in this area, with more subjects at different stages of training. This would allow investigation into the points at which these strategies are employed and the extent of the effect of training.

This experiment has shown that broad band noise excitation is a reliable method of measuring the vocal tract resonances non-invasively. The protocol is a valid method for analysing resonance tuning behaviour in young singers.

This technique will therefore be used as part of the method for the experiments in the next two chapters, investigating resonance tuning in professional soprano opera singers, in conjunction with magnetic resonance imaging to obtain images of the vocal tract structure.

Chapter 5

Experimental Protocol and Validation

This chapter outlines the protocol and testing of the method used in the main experiment of this PhD research, which used magnetic resonance imaging (MRI) to consider how professional soprano opera singers alter the shape of their vocal tracts to employ resonance tuning techniques. It aims to provide a solid understanding of the effects of MRI measurement conditions on singers, as per the second objective of this PhD research: *to understand the effects of MRI measuring conditions on singers, and verify the usefulness of MRI in singing research.*

In Chapter 4, a method for measuring the vocal tract resonances was tested. It was found that the method provided measurements of vocal tract resonances of sufficient quality, and that the protocol used was suitable for investigating resonance tuning in girl choristers.

The current chapter details the protocol followed for MRI data collection in the main experiment of this work and the data collected. It then details three types of analysis carried out to establish the validity of the measurement method, and finally, draws conclusions about the suitability of this method for investigating resonance tuning. Once this has been established, Chapter

6, will address the first (main) objective of this PhD: *to understand the methods by which resonance tuning is produced by soprano singers*, using the MRI data described and validated in the current chapter.

5.1 Introduction

Magnetic resonance imaging (MRI) is an invaluable tool in singing research, making it possible to visualise the inside of a singer's body and observe the movements of the individual articulators during sound production. Despite advances in imaging technology, however, the conditions experienced by the singer during an MRI scan are far from natural when compared to their normal performance environments. With the increased use of MRI in voice research, it is ever more important to fully understand the effects of the measuring conditions on the singer in order to justify using the results obtained using MRI as an indicator of normal singing behaviour.

Although singers often perform in odd situations, the environmental restrictions necessary for MRI data acquisition (see section 5.2.2) create a distinctly alien environment, even for the most experienced and versatile singer. That they perform as they would on a stage or in the rehearsal studio whilst in an MRI scanner cannot be presumed.

As discussed in section 2.2.3, a number of studies have used MRI to observe the vocal tract during both speech and singing, [98, 100, 101]. However, there is very little research using 3D imaging to specifically investigate singing techniques in soprano voices.

Advances in MRI have led to the development of "open-type" MRI scanners, which allow the subject to sit upright or lie supine. Unfortunately, these are not yet in widespread use. Kitamura et. al [158] investigated the differences in spoken vowels in three male subjects and found differences in the positions of the tongue between upright and supine positions, as well as differences in the shape of the lips and uvula, and larynx height.

The effects of singing training (speaking training was not investigated) may cause subjects to produce sound more consistently between upright and supine conditions. Although there has not yet been any research specifically investigating the effects of MRI conditions on the female singing voice, Traser et al. [159] investigated differences in vocal tract shape between upright and supine MRI in professional tenors and observed small changes to larynx height and jaw protrusion, but no changes in registers.

There has been some research on speech production in a supine position. However, due to the differences in speech and singing production, this is not directly comparable to singing production. Stone et al. [160] compared speech production of non-expert speakers using ultrasound, in both upright and supine positions. They measured the tongue position in both positions and found small differences for most subjects. The largest differences observed were less than 3 mm on average and were seen in the posterior tongue. Importantly, the acoustic differences were found to be negligible between upright and supine positions. Speed [161] also investigated the difference between speaking in upright and supine positions and found that they were consistent at least up to 4 kHz. Engwall et al. [162, 163] studied speech using both MRI and electromagnetic articulography (EMA)/electropalatograpy (EPG), which allow measurement of tongue contact with the roof of the mouth and a two-dimensional mid-sagittal display of the articulatory movements of the tongue. They observed differences in jaw position, lip protrusion and tongue contours and concluded that speech produced during MRI represented “hyperarticulated speech”.

The aim of this chapter is to establish the effects of MRI measurement conditions on the singers involved in order to ensure the reliability[1] of results obtained through MRI. This is achieved by proposing a method for observing singing production using MRI and testing the suitability[2] of this method.

---

[1] *Reliability* in this case is judged as the similarity of sound production between different positions, based on the acoustic output.

[2] A *suitable* protocol in this case is defined as one that allows quality measurements to be obtained while not being distressing or unpleasant for the subject.

Key features of the MRI conditions, such as the supine position and background noise, will be reproduced and their effects tested in a controlled environment.

This will then allow the central research question of this research to be explored: understanding the methods by which resonance tuning is produced by soprano singers by using MRI to investigate how different vocal tract articulators affect the vocal tract resonances with increasing fundamental frequency.

5.2 Method

The purpose of this protocol was to measure vocal tract resonances and allow the collection of MR images of the vocal tract articulators when singing (see Chapter 6), as well as measuring the effects of MRI measurement conditions on singers when singing (for this research) and speaking (to benchmark against previous studies). Therefore in this experiment, six opera singers performed various speaking and singing tasks in three different situations: in an MRI machine, in their normal singing position, and in "simulated MRI" conditions (defined in section 5.2.2). Vocal tract resonance measurements, Long-term Average Spectra (LTAS), and audio recordings were compared for each singer, for speech and singing in different situations.

5.2.1 Subjects

Very highly trained subjects were chosen for this investigation for two reasons: firstly (and most importantly), to shed light on the production of resonance tuning in very advanced singers (addressed in Chapter 6); and secondly, in the hope that their extensive training will mean that they will not be greatly affected by singing in unusual conditions and will be able to sing in an MRI machine in their normal way.

The subjects in this study were six professional female soprano or mezzo-soprano opera singers. Their ages ranged from 32 to 57, with a mean of 43.7 years and a standard deviation of 8.8 years. They all measured level 2.1 or 3.1 on the Bunch-Chapman scale [164] (International or National principal), and reported no health issues. Their details can be found in Table 5.1.

Subject	Age (years)	Voice type	Bunch-Chapman	Nationality
1	47	Soprano	3.1	British
2	57	Mezzo-Soprano	2.1	British
3	32	Soprano	3.1	Irish
4	40	Mezzo-Soprano	2.1	British
5	48	Soprano (light)	2.1[3]	Italian
6	38	Soprano	2.1	Hungarian

Table 5.1: Details of the subjects' ages, voice types (self reported), Bunch-Chapman [164] classifications and nationalities.

5.2.2 Experimental Protocol

Procedure

The experiment was divided into three parts. The first part took place in an MRI machine at York Neuro-Imaging Centre (YNiC), which has a General Electric 3 Tesla HDx Excite MRI scanner. The second part took place in an anechoic chamber with the subject standing up, singing as normally as possible with minimal vibrato, as in a performance or rehearsal. The third part of the experiment also took place in the anechoic chamber, and was designed to simulate the conditions in the MRI machine as closely as possible, to collect data that would allow comparison between normal conditions and MRI conditions (both actual MRI, and simulated conditions). The third part of the experiment involved the subject lying supine on a foam board,

---

[3] Subject 5 was actually an early music singer (medieval, renaissance, and early baroque), but the Bunch-Chapman [164] classification does not distinguish this from Opera.

wearing headphones, through which MRI noise was played at a similar sound pressure level to the MRI machine.

During the MRI scan, the subject was required to lie flat on their back on a hard "bed", they then had a coil surrounding their head and neck fitted, and were inserted into the MRI machine. This was a fairly confined space, a tube of approximately 60 cm diameter. Once the scan began, the patient was required to stay absolutely still, which is difficult to achieve while singing, as any movement (including excessive vibrato) produces blurry images. During scanning, the MRI machine makes a loud noise, which has a pitch of approximately 300 Hz. The subject was therefore required to cope with loud noise, in addition to maintaining the pitch of the required note against the perceived pitch from the hum of the scanner. Finally, the temperature around the MRI machine was quite cold, at approximately 21 $°C$, so they were given a blanket to wear for warmth.

Subjects were also asked to fill in a questionnaire, detailing their singing experience and asking about their techniques for singing high notes. They were asked to fill in part of this before the experiment and then the remaining part afterwards, to see if their experience had affected their opinions on their techniques. The questionnaire, information sheet, and consent forms are included in Appendix C.

Prior ethical approval was gained from the Physical Sciences Ethics Committee at the University of York (details in Appendix J).

Part 1 - MRI conditions

For this part, the subject and experimenter had to relocate to York Neuroimaging Centre (YNiC) to make use of the MRI machine there (a 2-minute walk). The subject was briefed by a member of YNiC staff and final written consent to be scanned was obtained. They were then positioned in the scanner by a member of YNiC staff (as shown in Figure 5.1) and fitted with foam ear-plugs and headphones. One optic microphone (Sennheiser MO2000) was attached to the head mount and another set up as far away as possible within the scanner room. A number of configuration scans were then taken to prepare the scanner for the image capture.

Figure 5.1: Shows a patient (not from this study) being placed into the MRI scanner (reproduced with permission from [96])

Tasks:

1. The subject sang three different vowel sounds (/ɑ/, /u/, and /i/) for 16 seconds, on 8 different pitches (shown in Table 5.2), chosen to be equally spaced (on a logarithmic scale) across the whole frequency range investigated. A separate scan was taken for each note. Before each scan, a reference tone (a recorded piano note) was played over the intercom and the subject was told which vowel to sing. The scan duration was 16 seconds per note, which required the singer to maintain the shape of her vocal tract during this time.
2. The subject was asked to hold a neutral vocal tract shape with their mouth slightly open, and to breathe as normally as possible without phonating or moving for 16 seconds, as this was the scan duration.

After the MRI scan, the subject was encouraged to take a break with food and drink, as required, before proceeding to parts 2 and 3 of the procedure.

Table 5.2 shows the fundamental frequencies investigated for each vowel. The G4 measurement for the /ɑ/ vowel for subject 2 was initially thought to be of poor quality (blurred due to movement) and repeated, but later found to be adequate and included in the study. The C6 measurement for the /ɑ/ vowel for subject 2 was discarded due to poor quality. The C6 measurement for the /ɑ/ and /u/ vowels for subject 6 were initially thought to be of poor quality and repeated, but later found to be adequate and included in the study.

Vowel	/ɑ/	/u/	/i/
C4	3 6	1 2 3 6	3 6
E4	1 2 3 6	1 2 3 6	1 2 3 6
G4	1 2 3 6	1 2 3 6	1 2 3 6
C5	1 2 3 6	1 2 3 6	1 2 3 6
E5	1 2 3 6	1 2 3 6	1 2 3 6
G5	1 2 3 6	1 2 3 6	1 2 3 6
A#5	1 2 3 6	1 2 3 6	1 2 3 6
C6	1 2 3 6	3 6	1 2 3 6

Table 5.2: The fundamental frequencies investigated for each vowel sound. The numbers represent the subjects that sung each vowel and fundamental frequency investigated.

Part 2 - "Normal" conditions

The subjects were asked to stand on a board[4] in their "normal performance stance" in the fully anechoic chamber at the University of York. This will be referred to in this work as "normal" conditions. They were then fitted with a head-mounted microphone (DPA4066, DPA microphones) approximately 4 cm from the lips, which was used to record the speaking and singing tasks. A second microphone (Earthworks M30, Earthworks microphones) was placed approximately 1 m from the subject, and this was used to record the entire experiment for later reference. The subjects were also fitted with a neck band carrying laryngograph electrodes [165] and the signal generated by this

---

[4] to counteract the effects of the mesh floor in the anechoic chamber.

was also recorded. These signals were recorded simultaneously during the entire experiment, using a Tascam DR680 portable digital recorder (sampling frequency 96 kHz). The subject was asked to sing a calibration tone and the sound pressure levels (A-weighted) at both the head-mounted and room microphones were measured and noted.

The subjects were then shown how to sing into the vocal tract resonance measuring device and a calibration recording was taken with the subject's mouth closed.

Then followed three tasks:

1. The subject was asked to read the standard text, "Arthur the rat" [166] (see Appendix D) in their normal speaking voice.
2. The subject sang individual notes, each on one breath, in an ascending chromatic sequence (12 notes per octave) from C4 to the top of their range[5], for three vowel sounds (/ɑ/, /u/ and /i/), while their vocal tract resonances were measured using the broad band noise device. They were played each note on an electric piano before it was sung and required to hold each note for approximately 6 seconds. The subjects were asked to sing in their "normal resonant performing voice" at a medium volume, keeping their mouth shape constant. They were reminded of the protocol during the tasks, if, for example, they moved (these measurements were repeated as necessary).
3. The subject sang one verse of the song "Once in Royal David's city" (see Appendix D).

---

[5] Notes were only repeated if the measurement was unsatisfactory or if the subject failed to sustain the note until the end of the measurement.

Part 3 - "Simulated MRI" conditions

The subject was then required to lie supine in the anechoic chamber, still wearing the microphone headset, to simulate the conditions in the MRI machine (demonstrated in Figure 5.2). To protect them from the floor of the anechoic chamber, the subject lay on a foam-covered board.

This experimental set-up will be referred to in this work as "simulated MRI" conditions. The vocal tract resonance measuring device was adjusted to be in the same position (relative to the singer's mouth) as when standing. Another calibration recording was taken, and the subject was then re-fitted with the laryngograph electrodes and fitted with earplugs and headphones to play recorded MRI noise[6].

Tasks 1 to 3 were repeated from part 2 (speaking and singing tasks), with MRI noise playing over the headphones for tasks 1 and 2 (speaking and singing individual notes).

Figure 5.2: "Simulated" MRI conditions, from the side and from above. Demonstrated using the head and torso simulator K.E.M.A.R. [167]

---

[6] This was recorded from the MRI machine, carrying out the same scanning procedure as in part 1. The subject was asked to indicate when they felt the level was the same as in the MRI machine.

Resonance measurements using broad band noise excitation

The method used to detect vocal tract resonances was tested in a preliminary experiment involving girl choristers (see Chapter 4). It employs the broad band noise excitation (at the lips) method described in section 4.2.2, after Epps et al. [95]. This method was chosen as it is minimally invasive to subjects, allows them to sing naturally and produces reliable resonance measurements. As in Chapter 4, the excitation signal consisted of harmonics spaced 5.38 Hz apart, from 250 Hz to 3500 Hz, and the method included a calibration procedure. The average amplitude of the excitation signal was 87 dB SPL.

As in previous works [14, 15, 151] and in Chapter 4, the frequencies of the vocal tract resonances were measured from the plot of P_{open}/P_{closed} against frequency by one researcher, and then checked by another. The values were then averaged between the two observations (where the results were different by more than approximately 100 Hz, they were discarded). The percentages and numbers of measurements omitted for R_1, R_2 and R_3 for each vowel are shown in Table 5.3.

	Vowel	/ɑ/	/u/	/i/
"normal" conditions	R_1	8.4 % (16/190)	4.9 % (9/183)	14.6 % (26/178)
	R_2	2.1 % (4/190)	7.7 % (14/183)	12.4 % (22/178)
	R_3	11.5 % (22/190)	34.4 % (63/183)	5.6 % (10/178)
"simulated MRI" conditions	R_1	16.6 % (27/163)	7.0 % (12/171)	14.6 % (24/164)
	R_2	0.0 % (0/163)	2.9 % (5/171)	9.8 % (16/164)
	R_3	9.2 % (15.163)	29.8 % (51/171)	3.0 % (5/164)

Table 5.3: The percentages (and number) of measurements omitted for each resonance, for each vowel

There was considerable variation between subjects, with total measurements omitted per subject: subject 1: 5.6 %, subject 2: 14.4 %, subject 3: 9.5 %, subject 4: 10.8 %, subject 5: 18.1 %, subject 6: 5.8 %.

5.2.3 Data

In part 1, the data collected consisted of 3-dimensional MRI images (between 19 and 26 for each singer), and audio recordings of the whole scanning procedure (taken from optic microphones). Since the MRI scan was only started *after* the subject had started singing, a short sample (average length 0.7 seconds) of clean audio was also collected for each vowel and fundamental frequency.

In parts 2 and 3 of the procedure, anechoic recordings of speech and singing were obtained, as well as resonance measurements for 3 vowels on all pitches from the chromatic scale task. As in part 1 of the experiment, the broad band noise excitation was only started *after* the subject had begun to sing, to allow samples of "clean" audio (without broad band noise) to be collected for each vowel and fundamental frequency (average length 1.1 seconds). Figure 5.3 shows how the clean audio samples were selected in Audacity [168], based on the waveform and spectrogram of the signal.

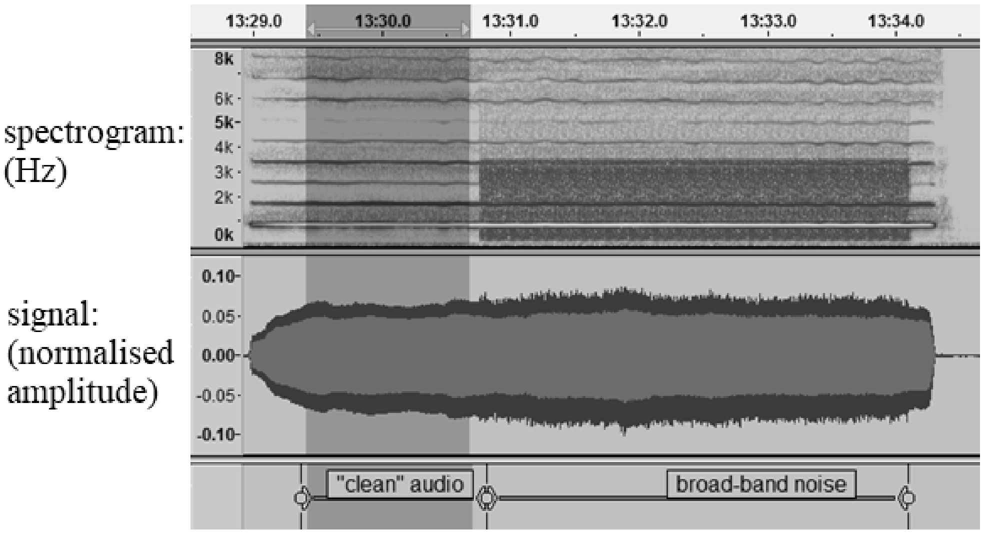


Figure 5.3: Screen shot of spectrogram (top) and waveform (bottom) of singing with broad band noise excitation, indicating the part of the signal chosen as "clean" audio.

In order to remove differences due to the different microphones used in the anechoic chamber (parts 2 and 3) in the MRI machine, the clean audio samples were inverse filtered in MATLAB (code included in Appendix L). This was achieved by convolving the audio samples with the inverse of the relevant microphone response (the DPA response for the anechoic recordings, and the optic microphone response for the MRI recordings). This was deemed necessary as, in some cases, the third resonance approached 3 kHz, where the microphone responses started to deviate from each other.

The microphone responses taken from the manufacturers datasheets are shown in Figure 5.4. Better compensation might have been achieved if the individual microphone responses had been measured, as the individual microphone responses may deviate a little from the published responses. However, in this application, minor deviations from a nearly-flat response are not of concern.

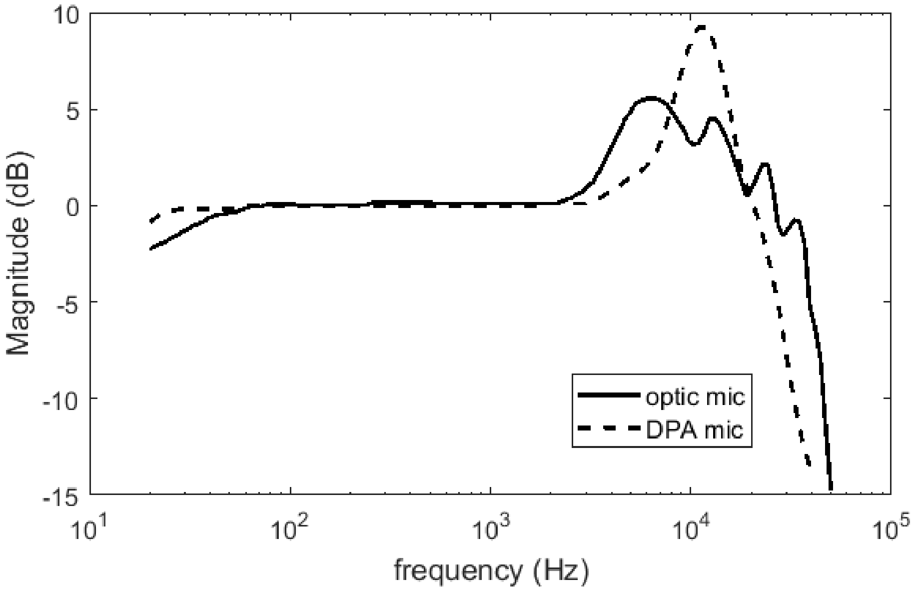


Figure 5.4: Frequency responses of the DPA microphone used in the anechoic chamber (dashed line), and the optic microphone used in the MRI machine (solid line).

The data collected in each part of the procedure is shown in Table 5.4.

The images obtained from MRI (green in Table 5.4) allowed the observation

Task:	Reading: The North Wind and the Sun	Sung scale with resonance measurements (3 vowels)	Singing: One verse of short song	Sung vowels (3 vowels, 8 pitches)
Part 1 (MRI machine)				MRI images Clean audio samples
Part 2 ("normal")	Anechoic recordings of speech	Resonance measurements Clean audio samples	Anechoic recordings of singing	
Part 3 ("simulated MRI")	Anechoic recordings of speech	Resonance measurements Clean audio samples	Anechoic recordings of singing	

Table 5.4: Data obtained from each task, for each part of the procedure. Data of the same colour were compared to each other.

of modifications to the shape of the vocal tract between the different fundamental frequencies and vowels, not only a mid-sagittal slice, as in previous studies [6, 7, 169], but full 3D images of the vocal tract. This will allow effects in other planes to be seen, such as widening of the pharynx and lateral changes to the shape of the tongue, and will be discussed in Chapter 6.

Unfortunately, although all six singers completed the second two parts of the protocol, only four of the singers were able to complete part 1 (MRI scan). Subject 4 found singing in the MRI machine a very unpleasant experience and was unable to sing properly. Although Subject 5 was willing to be scanned, the MRI machine malfunctioned and was out of order for several days, so only a calibration scan was obtained for this subject, with no images of the vocal tract captured during singing.

5.3 Analysis and Results

In this section, the differences between speech and singing produced in MRI conditions, "normal" conditions and "simulated MRI" conditions are investigated. The purpose of conducting this analysis was to better understand the effects of MRI measurement conditions on subjects, to inform singing research using data obtained through MRI.

As described in section 5.2.3, the data obtained consisted of measurements of vocal tract resonances, audio recordings of speech and song, and short audio recordings of sung vowels. Using the three different types of data collected, three different tests were carried out.

The first analysis was carried out to ascertain whether the differences between "normal" and "simulated MRI" conditions (red in Table 5.4), elicited a change in the vocal tract resonances. This was done by investigating the difference in vocal tract resonance measurements of the three sung vowels, obtained in the two different conditions using a Kolmogorov-Smirnov test. Neither the resonance measurements used in the first analysis, nor the magnitudes of the harmonics used in the third analysis followed normal distributions (tested using a one-dimensional K-S test in MATLAB), so non-parametric tests were used.

The second set of data to be considered was the anechoic recordings of speech and singing (orange in Table 5.4), produced in both "normal" and "simulated MRI" conditions. The comparison of this data served two purposes, firstly it allowed the effects of MRI conditions on both speech and singing to be studied and, secondly, it allowed for comparison of the effects of MRI conditions between speech and singing, since much of the existing literature on the effects of MRI conditions has focussed on speech. This was achieved by the comparison of the Long-Term Average Spectra (LTAS) of speech and singing produced in both "normal" and "simulated MRI" conditions.

Apart from the MRI images, the only data collected from part 1, were the short samples of clean audio (blue in Table 5.4). Since the same data were collected from all three parts of the experiment (anechoic and MRI), this was used to further test the similarity of voice production in the different situations, and therefore the suitability of MRI for research into resonance tuning involving MRI. The third analysis of the data therefore involved comparing audio samples between different conditions, by extracting information about the harmonics. The use of the data obtained is summarised in Table 5.5.

Resonance measurements	Analysis 1: Comparison of anechoic resonance measurements (KS testing)
Anechoic recordings of speech and singing	Analysis 2: Comparison of standing and supine acoustic recordings (LTAS)
Clean audio samples	Analysis 3: Comparison of anechoic/MRI audio recordings (Spearman correlation)
MRI measurements	Chapter 6 analysis

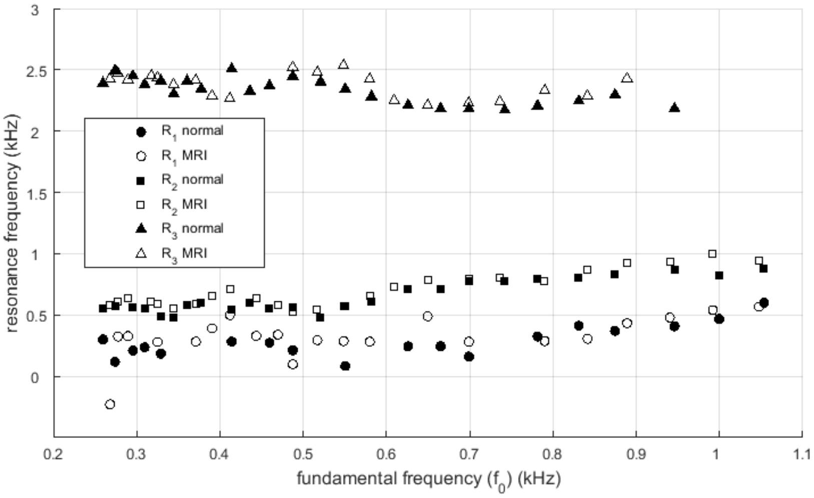
Table 5.5: Data obtained from each task, for each part of the procedure. Data of the same colour were compared to each other.

5.3.1 Analysis 1: Comparison of resonance measurements between "normal" and "simulated MRI" conditions (KS testing)

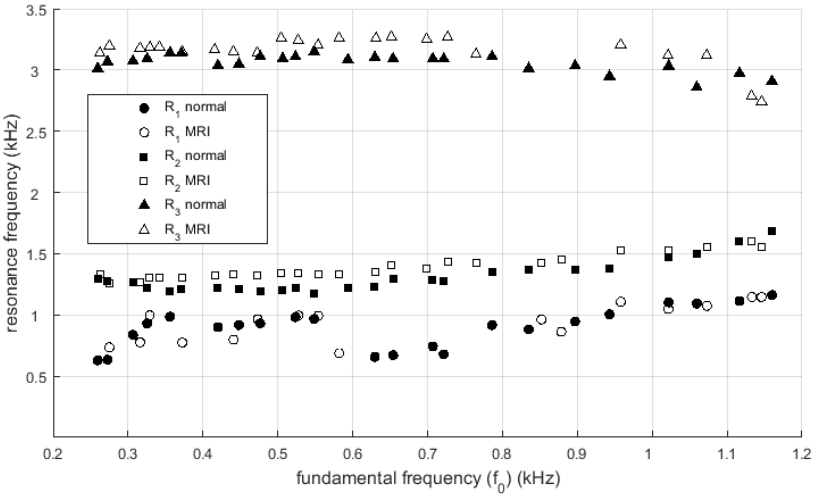
The aim of this analysis was to investigate the differences in singing production between "normal" conditions and "simulated MRI" conditions. This was achieved by comparing resonance measurements of sung vowels produced in parts 2 and 3 of the experiment.

The resonance measurements (obtained using broad band noise excitation, as discussed in section 5.2.2) were first collated in Excel [157], and then imported into MATLAB to evaluate the similarity between the "normal" and "simulated MRI" resonance measurements (code included in Appendix L). This was done using a two-dimensional Kolmogorov-Smirnov (K-S) test, a non-parametric test, where the null hypothesis is that both data sets were drawn from the same continuous distribution (after Peacock [170]).

The resonance measurements for all singers in both "normal" and "simulated MRI" conditions, are included in Appendix E (MATLAB code and raw data in Appendix L). Examples to highlight the differences between resonance measurements in the different conditions are shown in Figure 5.5. These are plots of the first three resonances against fundamental frequency, for an /ɑ/ vowel, for subjects 2 and 3. The "normal" and "simulated MRI" conditions were not found to be significantly different for subject 2 (top), however for subject 3 (bottom) the second and third resonances were significantly different (according to the K-S test, at the 5 % significance level) in the different



(a) subject 2, /ɑ/ vowel.



(b) subject 3, /ɑ/ vowel.

Figure 5.5: The first three resonances (R_1, R_2 and R_3) against fundamental frequency for an /ɑ/ vowel for subjects 2 (top) and 3 (bottom).

conditions.

Table 5.6 gives the results of the 2-dimensional Kolmogorov-Smirnov (K-S) test between resonance measurements in “normal” anechoic conditions and “simulated MRI” conditions, for each vowel and resonance, for all subjects.

It can be seen that in most cases, the results accept the null hypothesis

Subject 1				Subject 2				Subject 3			
Vowel	R_1	R_2	R_3	Vowel	R_1	R_2	R_3	Vowel	R_1	R_2	R_3
/ɑ/	0	1	0	/ɑ/	0	0	0	/ɑ/	0	1	1
/u/	0	0	0	/u/	0	0	0	/u/	0	0	0
/i/	0	0	1	/i/	0	0	0	/i/	0	0	0
Subject 4				Subject 5				Subject 6			
Vowel	R_1	R_2	R_3	Vowel	R_1	R_2	R_3	Vowel	R_1	R_2	R_3
/ɑ/	1	1	0	/ɑ/	0	0	0	/ɑ/	0	1	0
/u/	0	0	0	/u/	0	1	0	/u/	0	0	0
/i/	0	0	0	/i/	0	0	0	/i/	0	1	0

Table 5.6: The H-values for Kolmogorov-Smirnov test results for comparison of samples sung normally and in "simulated MRI" conditions (in the anechoic chamber). An H-value of 1 indicates that the results were significantly different (at $p \leq 0.05$).

Vowel	R_1	R_2	R_3
/ɑ/	1	4	1
/u/	0	1	0
/i/	0	1	1

Table 5.7: The H-values for Kolmogorov-Smirnov test results from Table 5.6, summed over all six subjects.

(that the samples were drawn from the same population), meaning that they are not significantly different. Table 5.7 summarises the results presented in Table 5.6, across all subjects. It can be seen that the /ɑ/ vowel more often showed a significant difference between the two conditions than the other vowels, which never showed a significant difference in first resonances, and only showed a significant difference in second and third resonances for one subject (apart from /u/ R_3). This may suggest that subjects produced the /ɑ/ vowel less consistently (between the different conditions) than the other two vowels investigated. This may be due to the fact that other vowels tend to "neutralise" towards the /ɑ/ vowel at high frequencies. This will be discussed further in section 5.4.

5.3.2 Analysis 2: Comparison of acoustic recordings between “normal” and “simulated MRI” conditions (LTAS)

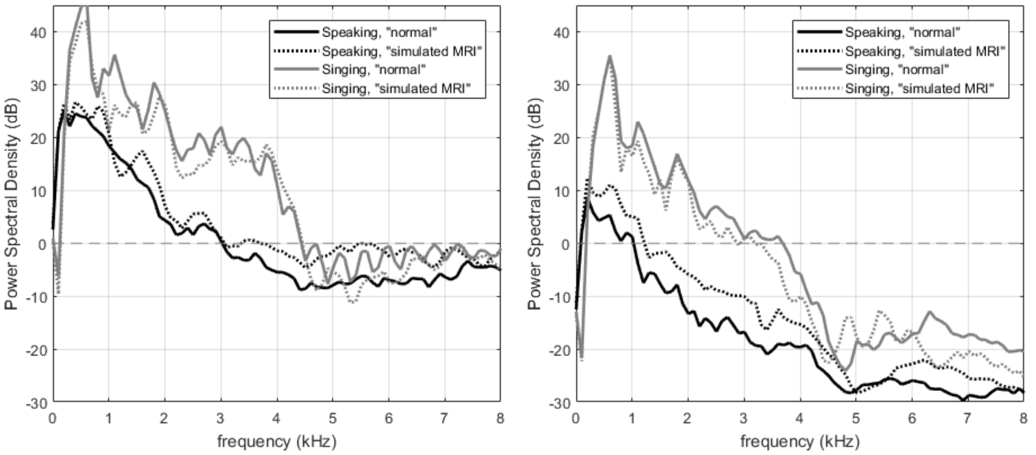
The purpose of the second analysis was two-fold: firstly to observe the effects of “simulated MRI” conditions on spectra of both speech and singing; and secondly, to compare the effects of “simulated MRI” conditions *between* speech and singing.

Using Audacity [168] the recordings of the spoken reading and song verse for each singer were cut from the audio recordings of the second and third parts of the experiment. In order to view the distribution of energy over the whole frequency range, the LTAS of each speech or song recording for each subject was then generated using Praat [70], in 100 Hz intervals. These values were then exported as a text file (.txt) and imported into MATLAB [68].

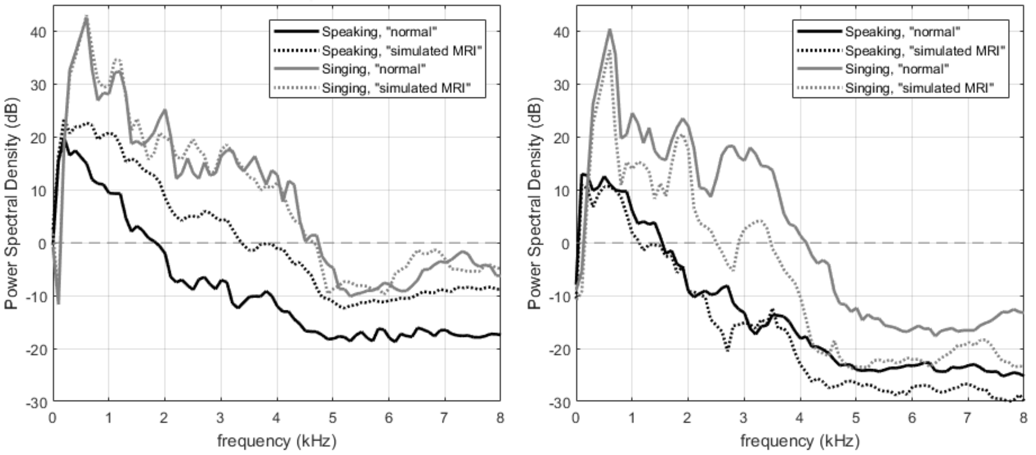
Figure 5.6 shows the LTAS results for each singer, plotted on the same axes for each subject. The results were not normalised, to allow differences in amplitude to be observed. The LTAS of speech are in black and of singing are in grey. Sound produced in “normal” (anechoic) conditions is represented by a solid line and those in “simulated MRI” (anechoic) conditions by a dashed line. The difference in LTAS between the two conditions was also calculated, for both speech and singing (as indicated in equation 5.1):

$$difference = {''normal''}\ LTAS - {''simulated\ MRI''}\ LTAS \qquad (5.1)$$

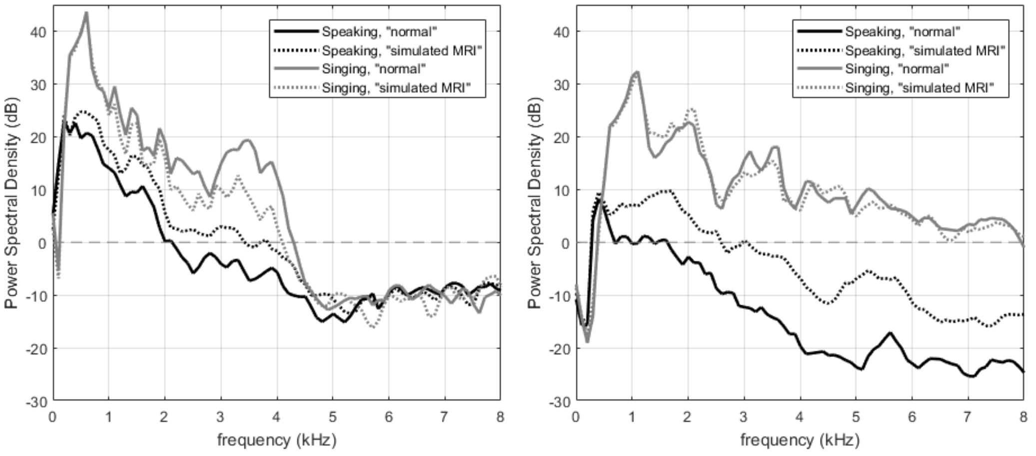
This difference was plotted against frequency in Figure 5.7.



(a) subject 1 (b) subject 2

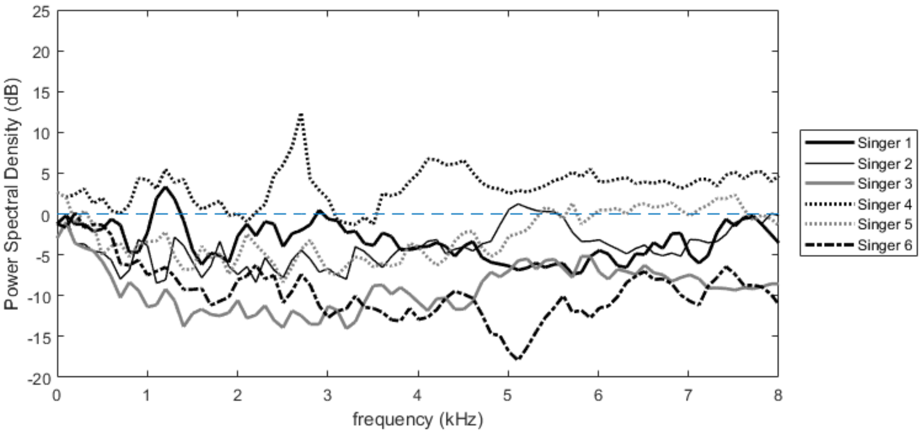


(c) subject 3 (d) subject 4

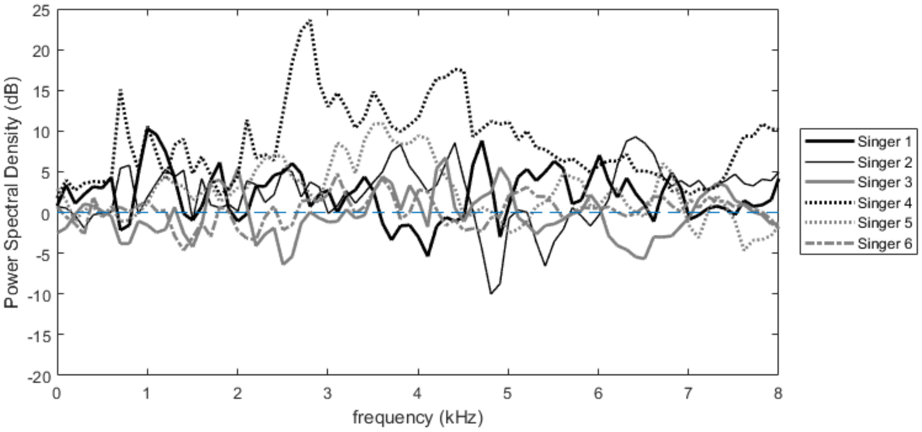


(e) subject 5 (f) subject 6

Figure 5.6: The Long-Term Average Spectra for speech (black) and for singing (grey), in “normal” (solid line) and “simulated MRI” (dotted line) conditions, for all 6 subjects.



(a) Speech



(b) Singing

Figure 5.7: The difference in Long-Term Average Spectra for speech (top) and singing (bottom), between “normal” and “simulated MRI” conditions, for all 6 subjects.

It can be seen from Figure 5.6 that the LTAS for singing consistently has more energy in the 1-4 kHz range than speaking, for both conditions investigated. Five of the six subjects produced sound of a higher amplitude in the “normal” conditions than “simulated MRI” conditions for singing (grey) (up to 15 dB difference), but for speech (black) this was true for only four of the six subjects.

The differences between the “normal” and “simulated MRI” conditions were mostly very small; for speech, the difference between conditions was more constant with frequency than for singing, consisting of a shift in amplitude of approximately 5-10 dB (normal conditions had a higher amplitude than "simulated MRI" for all but one subject). The small acoustic differences agree with previous studies. However the Lombard effect [171] suggests that subjects should *increase* the volume of voice production in the presence of other noise, rather than decrease it.

For singing, the difference between the conditions was smaller than for speech (up to approximately 8 dB), suggesting that the singers in this study were less affected by MRI conditions when singing than when speaking.

As expected [172], the LTAS showed differences between speaking and singing for all singers (irrespective of conditions), suggesting that the production of speaking and singing is very different. Conclusions drawn about the effects of MRI conditions on speech (from previous studies) therefore cannot necessarily be assumed to be true for singing. Both the differences in LTAS between conditions and between speaking and singing varied considerably between subjects, suggesting that this is quite an individual matter.

5.3.3 Analysis 3: Comparison of Anechoic/MRI audio recordings (Spearman correlation)

The third set of analyses involved the short samples of clean audio recordings of sung vowels, gathered during both the anechoic and MRI recordings (all three parts of the experiment). This is the only data collected from all three parts, so plays an important part in testing the similarity of voice production in the different situations.

Using Audacity [168], individual "clean" audio samples (that did not include any background MRI or broad band noise) for each note and vowel were cut out from the audio recordings from all three parts of the experiment. Samples were chosen by inspection of the waveform and spectrogram of the recordings. Samples were selected to be as long as possible, while maintaining a near-constant amplitude and frequency spectrum. Care was taken to avoid the onset period and to select an integer number of periods of vibrato, when this was present.

A MATLAB program (see Appendix L) was used to detect the frequencies and magnitudes of the first 20 harmonics below 10 kHz in each sample. An example plot of the spectrum of a sample, including the harmonics detected, is shown in Figure 5.8. Similarity of the samples was then deduced by carrying out a Spearman test for correlation on the harmonics in pairs of samples of the same fundamental frequency (code in Appendix L). The Spearman correlation is similar to the Pearson correlation (which is used for parametric data), but represents a statistical measure of the strength of a monotonic relationship between paired data [173]. It can take a value between -1 and +1, with a value of close to ±1 indicating the strongest correlation between the two variables. This test was chosen as it is a common test of correlation between bivariate, non-parametric data. The data to be compared by this test were the amplitudes of harmonics in two different samples of audio, which, if identical, would correlate perfectly.

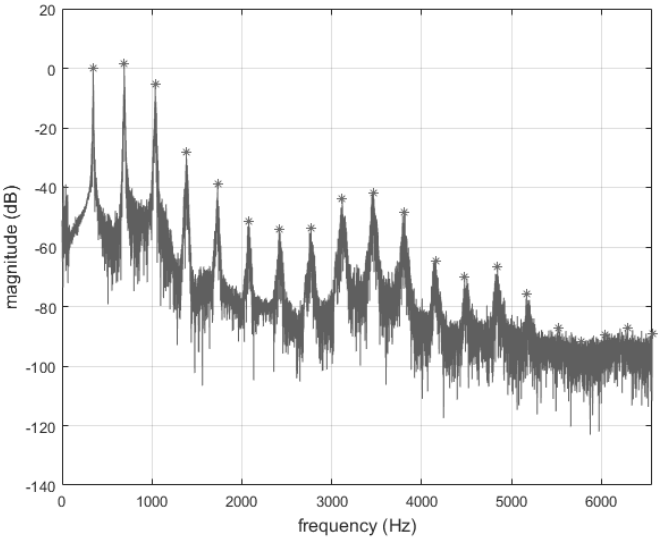


Figure 5.8: The spectrum of an anechoic clean audio sample and the harmonics detected (asterisks), plotted in dB.

One example plot of the magnitudes of corresponding harmonics of two samples of the same fundamental frequency, in "normal" conditions and "simulated MRI" conditions (parts 2 and 3) is shown in Figure 5.9. As expected, considering that the two samples represent the same vowel, sung by the same subject, at the same frequency, the correlation between the magnitudes of harmonics in the two samples is very strong, and the line of best fit has a gradient close to 1.

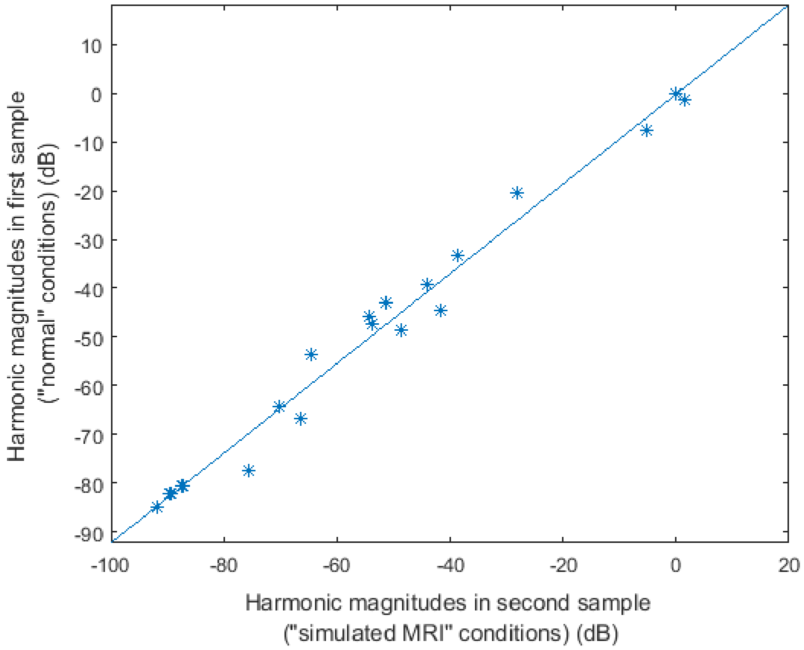


Figure 5.9: A plot of relative magnitudes of harmonics in two samples, for the "normal" and "simulated MRI" conditions. (Subject 6, $f_0 = 343$ Hz, /ɑ/ vowel)

The mean of results of the Spearman correlation comparing the amplitudes of relevant harmonics across the same vowels and pitches in different conditions, were taken across each subject, and these results are summarised in Tables 5.8 (comparison between "normal" and "simulated MRI" conditions), and 5.9 (comparison between vowels sung in the anechoic chamber and in MRI conditions).

Subject	min	max	**mean**	stdev
1	0.583	0.967	**0.881**	0.061
2	0.676	0.983	**0.866**	0.072
3	0.615	0.992	**0.868**	0.077
4	0.661	0.980	**0.859**	0.072
5	0.541	0.983	**0.889**	0.068
6	0.697	0.983	**0.877**	0.062

Table 5.8: The Spearman correlation results (minimum, maximum, mean and standard deviation) for comparison of harmonics in vowels sung in "normal" and "simulated MRI" conditions in the anechoic chamber.

Table 5.8 summarises the comparison between vowel sounds produced in parts 2 and 3 of the experiment ("normal" and "simulated MRI" conditions).

From these results, it can be seen that the mean Spearman correlation was between 0.859 (subject 4) and 0.889 (subject 5). A correlation value exceeding 0.8 indicates a *very strong* positive correlation (according to Evans' [174] classification[7]) showing that the sung vowel sounds were in general very similar between the two conditions investigated.

The left half of Table 5.9 summarises the comparison between vowel sounds produced in parts 1 and 2 of the experiment (the real MRI recordings and "normal" conditions), and the right half shows the comparison between vowel sounds produced in parts 2 and 3 of the experiment ("simulated MRI" conditions and the actual MRI recordings). Only 4 singers completed all three parts of the experiment, and hence, these results can only be calculated for them.

For the comparison between MRI and "normal" conditions, the means lay between 0.793 and 0.876. For the comparison between MRI and "simulated MRI", the means lay between 0.832 and 0.888. Again, these values indicate a *strong/very strong* positive correlation, showing that the sung vowel sounds were in general very similar between the three conditions investi-

---

[7] According to Evans [174], R values are classified as follows: 0.00 - 0.19 : "very weak", 0.20 - 0.39 : "weak", 0.40 - 0.59 : "moderate", 0.60 - 0.79 : "strong", 0.80 - 1.0 : "very strong".

comparison:	MRI / “normal”				MRI / “simulated MRI”			
Subject	min	max	**mean**	stdev	min	max	**mean**	stdev
1	0.600	0.967	**0.852**	0.080	0.297	0.943	**0.832**	0.122
2	0.314	1.000	**0.793**	0.194	0.657	1.000	**0.862**	0.107
3	0.371	1.000	**0.876**	0.147	0.371	1.000	**0.888**	0.128
6	0.674	0.976	**0.847**	0.098	0.459	0.983	**0.848**	0.139

Table 5.9: The Spearman correlation results (minimum, maximum, mean and standard deviation) for comparison of harmonics in vowels sung in the anechoic chamber (“normal” and “simulated MRI”) to those from the MRI machine.

gated. For three of the four subjects, the MRI conditions and “simulated MRI” conditions produced higher correlations than between MRI conditions and “normal” conditions, which is as expected.

The results of the audio comparison using Spearman correlation showed that the correlation between the harmonics in all three situations was either strong (0.60-0.79) or very strong (>0.8), indicating that the samples were very similar. As expected, the actual MRI and “simulated MRI” audio were more similar to each other than the MRI and “normal” conditions.

5.3.4 Singers’ Experience

As discussed in section 5.2.2, the subjects were asked to complete a questionnaire which was split into two parts, one to be answered before the experiment began, and one after. The purpose of this was to observe any change in a subject’s perception of their own singing and singing production, and to allow them to reflect on any differences in their singing technique between the different conditions investigated (MRI conditions, and “normal” and “simulated MRI” conditions in the anechoic chamber). The questions allowed for open-text answers.

The full questionnaire responses can be found in Appendix K. The comments relevant to this chapter are summarised below. A popular method of analysing questionnaire data is qualitative data analysis. However, the purpose of this is often to reduce the complexity of large datasets, so was not deemed necessary in this study due to the small sample size. Only the subjects that completed the MRI part of the experiment successfully (subjects 1, 2, 3, and 6) have been included.

The subjects were asked if they had noticed any differences in the way they sang in the different positions (standing up and lying down), and what they were. Comments mostly focussed on the subjects' spatial perception, and were similar to the comments made prior to the experiment, with subject 1 commenting on the lack of a feeling of "connection" and the different head and neck alignment. Subject 2 said that she found the MRI machine "very restrictive", and remarked that standing on the netting in the anechoic chamber in parts 1 and 2 of the procedure also affected her feeling of "support" during singing. Subject 3 commented that she found it harder to sing high notes in the MRI machine, and said that she felt "stretched" and wanted to move her legs for support. Subject 6 said that singing lying down was challenging but she thought that it went surprisingly well.

Subjects generally found the MRI machine a very strange environment to sing in, with subject 1 describing it as "surreal". She commented on the lack of projection or sense of space, said she found it awkward and had tension in her shoulders. Subject 2 also found it "odd" and said she was affected by the lack of freedom. Subject 3 commented that she "hated not being able to hear properly" and the lack of feedback. Subject 6 said she found it difficult at first (finding it claustrophobic).

All subjects agreed that they felt they sang differently in the MRI machine when compared to more normal conditions. Three mentioned "support". Subject 1 said she felt constricted, deprived of connection to the outside world and her "emotional centre". Subject 2 commented on the lack of pelvic/leg support. Subject 3 said she "had to try much harder". Subject 6 commented that she would normally use more vibrato and would position breathing support lower (she claimed to have used "chest" support in the

MRI machine).

5.4 Discussion and Conclusions

In this chapter, a three-part protocol was tested, which allowed for the collection of MRI images, vocal tract resonance measurements, recordings of speech and singing and audio samples of sung vowels. The subjects performed speaking and singing tasks in three different situations (MRI conditions, "normal" and "simulated MRI"). This achieved the aim of collecting MRI images and anechoic resonance measurements of the vocal tract, as well as audio data. The overlap of data between the three different situations allowed for three statistical analyses to be carried out, for the purpose of evaluating the validity of MRI in investigating the singing voice.

The first statistical analysis used a 2-dimensional Kolmogorov-Smirnov test to evaluate the difference in resonance measurements collected in "normal" and "simulated MRI" conditions. It showed that in most cases there were no significant differences at the 5 % level. The significant differences that did occur were not consistently in the same vowels and resonances for different subjects, suggesting that these results were individual. The resonances of the /ɑ/ vowel were found to be different between "normal" and "simulated MRI" conditions more often than the /u/ or /i/ vowels (6/18 tests were significantly different for the /ɑ/ vowel, compared to 1/18 and 2/18 for the /u/ and /i/ vowels, respectively). This suggests that the /ɑ/ vowel was produced least consistently between conditions.

The second statistical analysis investigated the difference between speech and singing produced in "normal" and "simulated MRI" conditions, using Long-Term Average Spectra (LTAS) to examine the distribution of energy with frequency, in both speech and song. The difference between "normal" and "simulated MRI" conditions for speech was almost frequency-independent, but consisted of a shift in amplitude of approximately 5-10 dB (higher amplitude in normal conditions for all but one subject), which may be explained by the lack of auditory feedback in the anechoic chamber. For singing, the difference between “normal” and “simulated MRI” conditions was considerably smaller than for speech (up to approximately 8 dB), suggesting that singers were less affected by MRI conditions when singing than when speaking, which could suggest that this is an effect of training, as singers train very hard to be able to produce vowels in a certain way, making them very consistent in reproducing them in different conditions. These results appeared to vary considerably between subjects, suggesting, like the first analysis, that this is quite an individual matter. Cross-subject analysis has not been carried out in this section and differences in articulator movements between subjects will be examined in Chapter 6. As expected [172], the LTAS results showed differences between speech and singing. However, these were similar across the conditions investigated (“normal” and “simulated MRI”), with the singing containing more energy in the 1-4 kHz range.

The third statistical analysis investigated the differences between the harmonics of sung vowels produced in actual MRI conditions and “simulated MRI” conditions. This was done using a Spearman correlation to evaluate the correlation between the magnitudes of corresponding harmonics in pairs of samples (on the same vowel and fundamental frequency). The results indicated a strong or very strong correlation between the magnitudes of the harmonics, which demonstrates that the harmonic content of sound produced in the “simulated MRI” conditions was not significantly different to that produced in actual MRI conditions. This is further supported by the fact that the correlation between the MRI conditions and “normal” conditions, although still strong, was not as high as between actual MRI and “simulated MRI” conditions.

The results of the questionnaire indicated that singers felt the MRI machine was a very difficult environment to sing in, due to the feeling of constriction and lack of “support”, even though this was not reflected in the acoustic parameters investigated. The comments made generally referred to the physical sensation of being in the MRI machine, rather than the auditory effects. Several commented that they lacked a feeling of being “grounded” in the MRI machine, and found it restrictive and claustrophobic. Several also made comments about specific articulators, such as feeling a difference in the positions of their tongue and jaw. Articulator measurements were of course only observed in a supine position, but there were no consistent patterns in the acoustic spectrum to suggest that this was the case. It is possible that there may be more subtle differences in the sounds produced than has been discovered in this experiment. A larger data set, with repeated measurements and more subjects, would be useful for investigating this.

It is worth considering at this point that the singers in this study were self-selecting to an extent; not only did one singer not complete the MRI, but when arranging the study, several singers were approached who declined to take part, which could have been because they would find it too stressful. In addition to this, although the major aspects of the MRI conditions were replicated in the anechoic chamber (such as the supine position, ear occlusion and noise), more subtle aspects such as the cold temperature and reverberation of the MRI scanner were not. Further results using audio recorded in MRI machines would be of use in studying the effects of MRI conditions more extensively.

The results of these three analyses indicate that, in spite of the subjects’ perception, the singing produced in the different conditions investigated was very similar in terms of the three measures investigated: vocal tract resonances, overall frequency content (LTAS) and harmonic content. These results are in agreement with previous studies such as Traser et al. [159], who observed small changes to the vocal tract in tenors, but no changes in registers. This also supports the previous studies on speech, such as Speed [161], who found that speakers were consistent at least up to 4 kHz, and Stone et al. [160], who found small changes in the vocal tract between upright and supine positions, but negligible acoustic differences.

Results obtained using MRI should still be treated with caution, however, it should be noted that there does exist some variability between subjects, and it is not possible to predict from these results how singers of different ages, voice types and experiences would react when in MRI conditions. For this reason it would be beneficial if further voice research involving MRI were to include some kind of verification procedure as standard to investigate the effects of MRI conditions on singers, similar to the analysis carried out in this work.

Further study investigating the auditory effects of the MRI environment would be a valuable addition to all research studies involving this method of imaging the human body. It is also worth noting that although an anechoic chamber is a much more natural performance environment than an MRI machine, it is still an unusual performance environment for singers. The methods used in this study were sufficient for contextualising and validating the results obtained in this work, as they show that the effects of MRI conditions on singing are small enough that results obtained by this method are still valuable. However, they do not give a complete picture of the perceptual and auditory effects of MRI conditions on subjects.

Overall, the results presented in this chapter have shown that the effects of MRI conditions are generally not significant for this group of singers, although it is not known whether these results can be extrapolated to other groups. These results therefore provide a solid foundation for analysis of the professional operatic soprano singing voice using MRI. Since MRI provides an insight into the inner workings of the vocal tract which is unobtainable by any other method, the unpleasant effects of the conditions for the singer represent an acceptable compromise. Based on the findings of this chapter, therefore, the following chapter will report the results of the full experiment, focussing on the analysis of images obtained using magnetic resonance imaging and how they can improve our understanding of the production of resonance tuning.

Chapter 6

Production of Resonance Tuning

In Chapter 5, the methods of the main experiment at the centre of this thesis were outlined, involving resonance measurements, audio recordings and MRI image capture. It was established that MRI conditions did not have a significant effect on singing production in most cases.

This chapter will present the results of the experiment that address the main objective of this PhD research: *to understand the methods by which resonance tuning is produced by soprano singers.*

The chapter begins by discussing the type and extent of resonance tuning observed in this experiment (in parts 2 and 3 of the protocol, "normal" and "simulated MRI" conditions) and how it compares to previous studies. This will be followed by a description of how the MRI images obtained were processed to generate area functions (using a novel algorithm) and two-dimensional measurements of the vocal tract articulators. Finally, it discusses the statistical analysis carried out to explore the relationships between the articulators and the vocal tract resonances, and the meaning and implications of these results.

Parts of this work have been previously published in a journal paper, "Determining the relevant criteria for three-dimensional vocal tract characterization", in Journal of Voice [8].

6.1 Introduction

In Chapter 5, the protocol used for the main experiment was described, which involved collecting measurements of vocal tract resonances, audio recordings, and MRI data. Various tests were carried out to evaluate the appropriateness of comparing data obtained in anechoic "simulated MRI" conditions to actual MRI measurements. The different types of data gathered from each part of the experiment (MRI, "normal" and "simulated MRI" conditions) are summarised in Table 6.1, which illustrates the comparisons that can be made between them.

Task:	Reading: The North Wind and the Sun	Sung scale with resonance measurements (3 vowels)	Singing: One verse of short song	Sung vowels (3 vowels, 8 pitches)
Part 1 (MRI machine)				MRI images Clean audio samples
Part 2 ("normal")	Anechoic recordings of speech	Resonance measurements Clean audio samples	Anechoic recordings of singing	
Part 3 ("simulated MRI")	Anechoic recordings of speech	Resonance measurements Clean audio samples	Anechoic recordings of singing	

Table 6.1: Data obtained from each task, for each part of the procedure. Data of the same colour were compared to each other. (repeat of Table 5.4)

The current chapter focusses on measurements of the vocal tract resonances obtained using broad band noise excitation and the analysis of images of the vocal tract obtained using MRI, in order to investigate the resonance tuning behaviour of soprano singers and the links between specific articulators and resonances.

The first stage of the analysis will examine the resonance tuning measurements, to establish the type and extent of resonance tuning strategies used by singers and identify whether this aligns with expectations (based on both acoustic theory and previous studies).

In addition to this, the differences in resonance tuning between vowels will be investigated. It is expected that different resonance tuning strategies will be used for different vowels, due to the differences in formant values between vowels. The first resonance is significantly different for different vowels, for example the closed-front /i/ vowel has its first formant at approximately 310 Hz, while the open-front/closed-back /a/ vowel has its first formant at approximately 850 Hz [117]. Acoustic theory [13] suggests that resonance tuning is employed once the resonance frequency equals the fundamental frequency, to produce sound more efficiently.

Differences in resonance tuning between subjects will also be examined. Although the subjects in this experiment have similar backgrounds, it is possible for singers to produce a very similar sound by different methods (known as *articulatory compensation* [64]), which might not be apparent from the acoustic spectrum, but could be identified from MRI images.

In the next stage of the analysis, the resonance tuning techniques used by the singers will be assessed in light of the MRI data. Area functions will be obtained by generating a 3D model of the vocal tract, and then “slicing” this at regular intervals to produce an area function. The iterative bisection algorithm, adapted from that used by Story [102], was used to generate a 2D area function from each mid-sagittal slice. Area functions for each subject and fundamental frequency will be examined for relationships with fundamental frequency, and this will be compared to the resonance tuning results.

The final stage of the analysis, which represents the fundamental purpose of this research, will aim to determine which articulators affected the vocal tract resonances the most strongly. To achieve this, the resonance tuning results and the MRI data will be combined, using a feature selection process to identify which variables (articulator measurements) produce the best fitting regression models for the data. Additionally, to determine whether the movements of articulators used to alter the vocal tract resonances vary between vowels and between singers, patterns of resonance tuning and area functions will be compared between singers and statistical analysis carried out on data for individual singers.

6.2 Data Processing

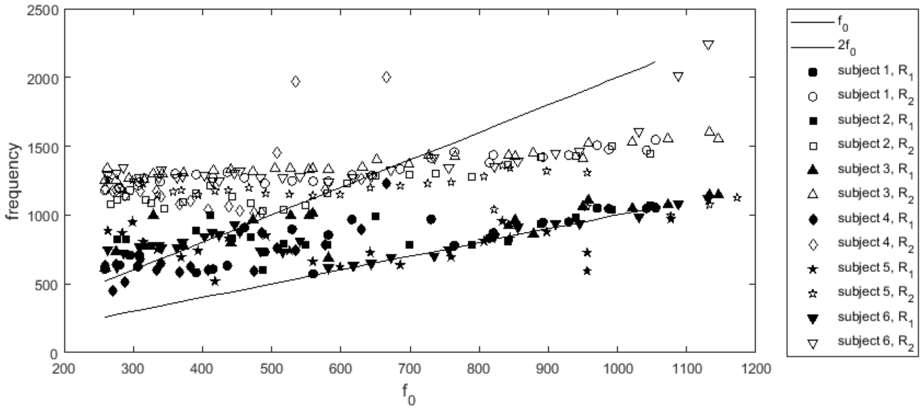
The experimental protocol for this experiment was discussed in detail in Chapter 5, (section 5.2.2). In summary, it consisted of three parts: the first part taking place in a "normal" standing position in an anechoic chamber, the second part in "simulated MRI" conditions in an anechoic chamber, and the third part in an MRI machine. In the first two parts the subjects carried out speaking and singing tasks, producing acoustic recordings and vocal tract resonance measurements (obtained using broad band noise excitation). In the third part of the experiment, acoustic recordings and MR images of the vocal tract were obtained as singers sang a range of notes in the MRI machine.

As discussed in Chapter 5, only four of the six subjects completed the MRI part of the experiment, so analysis involving data obtained through MRI only includes these subjects.

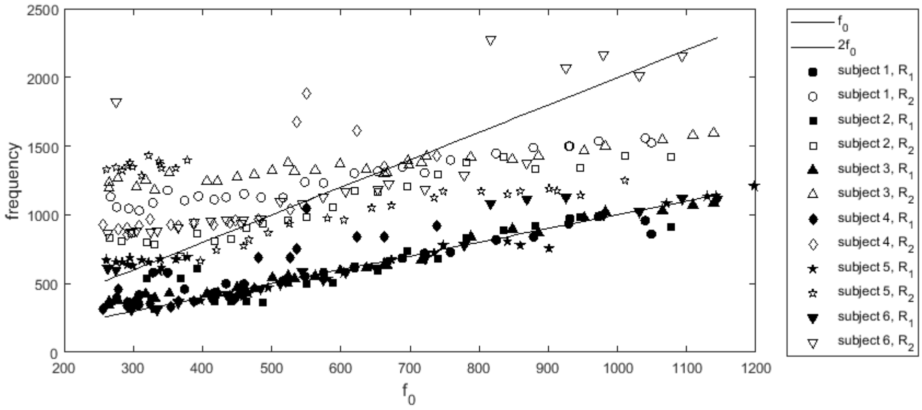
The following section will detail the processing and analysis of the vocal tract resonance measurements, and the MRI images obtained.

6.2.1 Resonance Measurements

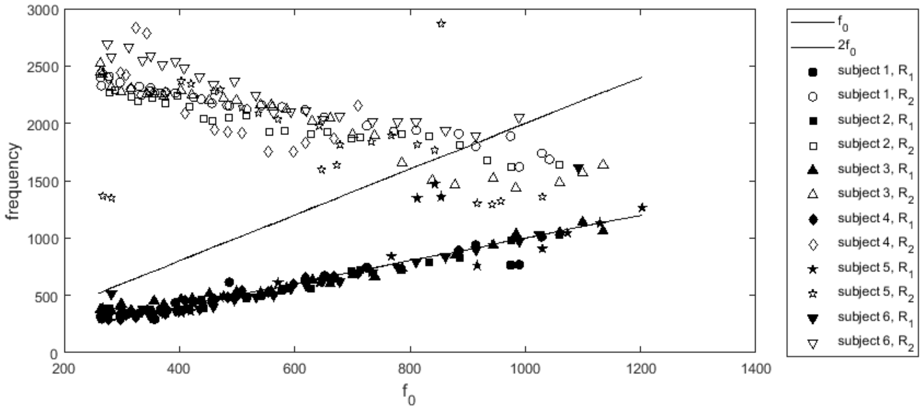
As described in section 5.2.2, the resonances of the vocal tract were measured using broad band noise excitation of the vocal tract. Each singer sang notes of approximately 4-5 seconds' duration from C4 (262 Hz) to the top of her range, on three different vowels, /ɑ/, /u/ and /i/. This yielded suitable measurements of vocal tract resonances over a wide frequency range.



(a) /ɑ/ vowel



(b) /u/ vowel



(c) /i/ vowel

Figure 6.1: The first two resonances plotted against fundamental frequency, for all subjects, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom). The first resonances are represented by filled-in shapes, and the second resonances are represented by empty shapes.

The first and second resonances for all subjects were plotted against the fundamental frequency sung for each vowel and these are shown in Figure 6.1. Plots of the first three resonances for each singer individually are included in Appendix E.

As in previous works [14, 151], resonance tuning was identified by plotting a histogram of the frequency difference between the first resonance and the fundamental frequency (R_1 - f_0), which is shown in Figure 6.2. In this study, a central peak approximately 120 Hz wide was found, so resonance tuning was defined as occurring when the resonance was within 60 Hz of the relevant harmonic.

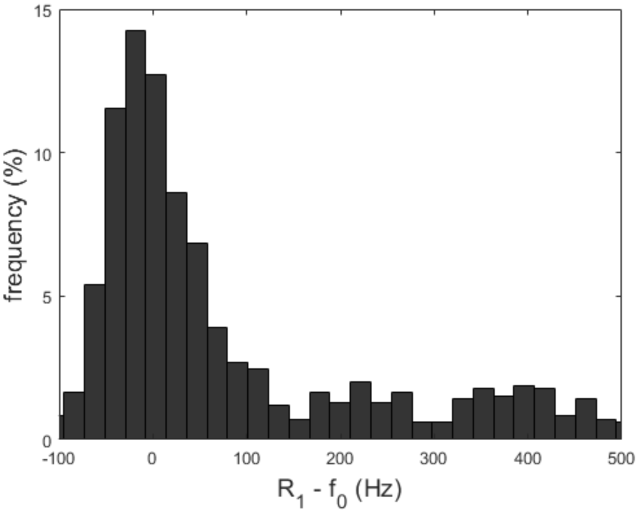


Figure 6.2: Histogram showing the distribution of the difference in frequency between the measured values of R_1 and f_0 (R_1 - f_0).

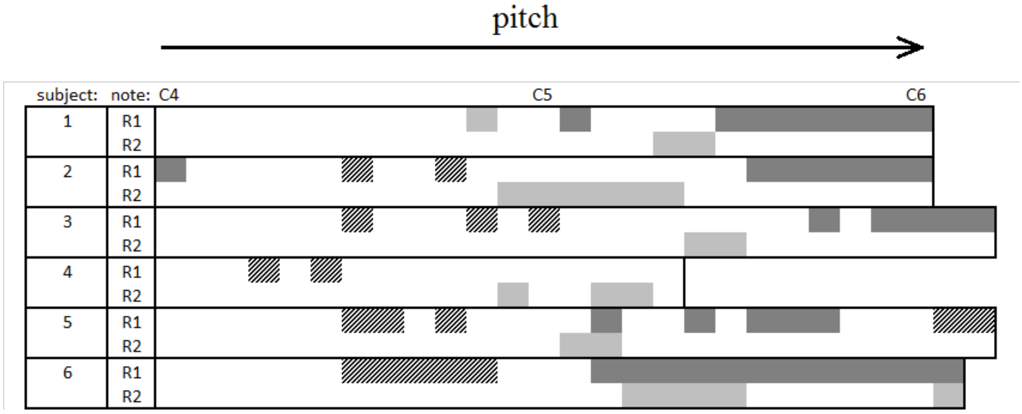
The resonance tuning strategies used by each subject are summarised in Figure 6.3 (a figure including both "normal" and "simulated MRI" measurements is included in Appendix F). The four tuning strategies investigated were $R_1 : f_0$ tuning (dark grey), $R_1 : 2f_0$ tuning (dark stripes), $R_2 : f_0$ tuning (not observed) and $R_2 : 2f_0$ tuning (light grey). The highest fundamental frequency reached varied between singers, as can be seen from the upper limit of Figure 6.3.

The results for the /ɑ/ vowel showed that most subjects employed $R_1 : f_0$ over the upper half of the frequency range investigated. Some $R_2 : 2f_0$ tuning was seen for all subjects, mostly just above the middle of the frequency range. All subjects employed $R_1 : 2f_0$ tuning sporadically over the lower part of their range, and subject 5 also employed $R_2 : f_0$ tuning near the very top of her range.

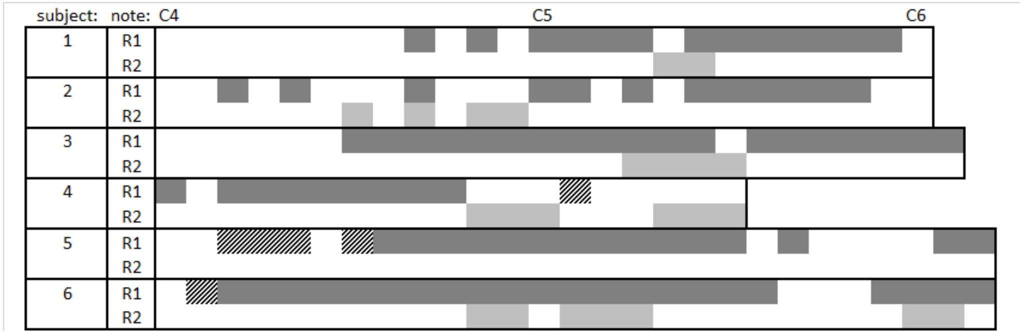
For the /u/ vowel, all subjects employed $R_1 : f_0$ tuning extensively across most of the range investigated. Subjects 4, 5, and 6 exhibited some $R_1 : 2f_0$ tuning at the lower end of the range investigated. $R_2 : 2f_0$ tuning was also widely employed across the middle of the frequency range, but no $R_2 : f_0$ was observed for this vowel.

Finally, for the /i/ vowel, again all subjects employed $R_1 : f_0$ tuning extensively across the entire frequency range. Subject 6 exhibited a very small amount of $R_1 : 2f_0$ tuning at the lower end of the range investigated. Two of the subjects also used $R_2 : 2f_0$ tuning for one note near the top of their range.

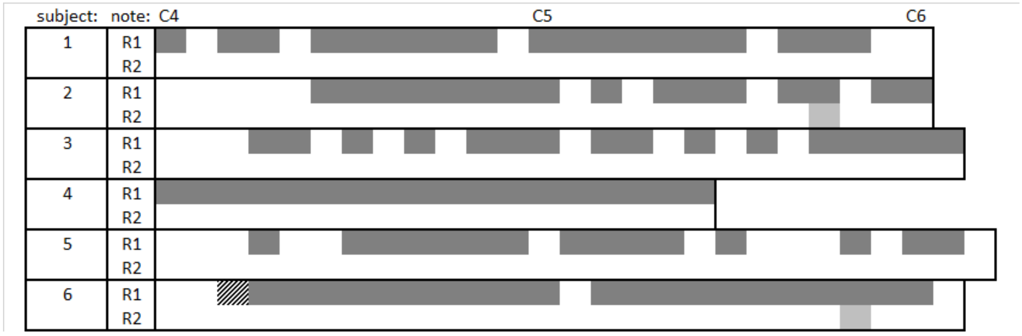
The subjects showed the least similarity (with each other) for the /ɑ/ vowel. This vowel also had the least resonance tuning overall of all the examples studied and, as seen in Chapter 5, showed the least consistency between different conditions. The resonance tuning patterns for the /u/ and /i/ vowels were very similar, between both singers and vowels, although the /u/ vowel had more R_2 tuning than the /i/ vowel. These two vowels were very consistent between subjects, and did not show a large variation with frequency, with singers typically employing the same resonance tuning techniques over wide frequency ranges.



(a) /ɑ/ vowel



(b) /u/ vowel



(c) /i/ vowel

Figure 6.3: The resonance tuning strategies employed by each subject, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom). The tuning strategies observed were $R_1 : f_0$ tuning (dark grey), $R_1 : 2f_0$ tuning (dark stripes) and $R_2 : 2f_0$ tuning (light grey). ($R_2 : f_0$ tuning was not observed.)

In previous studies on resonance tuning, Garnier et al. [15] found that in 12 soprano singers (described as 4 non-experts, 4 advanced and 4 professional singers), both advanced and professional singers employed extensive R_1:f_0 tuning, between C5 and C6, for the /ɑ/ vowel. A few of the singers also exhibited R_2:$2f_0$ tuning, beginning around C6. In another study, Henrich et al. [14] found that for sopranos, for the /ɑ/ vowel, R_1:f_0 tuning was employed between C5 and C6, and R_2:$2f_0$ tuning was also observed for 3-5 notes near the frequency of R_2. Extensive R_1:f_0 tuning was also observed for the /u/ vowel, with similar R_1:f_0 tuning to the /ɑ/ vowel (2-5 notes for the /ɑ/ vowel, and 2-4 notes for the /u/ vowel). The /i/ vowel was not investigated in either of these studies. The singers in these previous works were described as ranging from "nationally recognized to amateur singers" [14].

In this current study, much less resonance tuning was observed in general for the /ɑ/ vowel: R_1:f_0 tuning only occurred sporadically over the upper part of the subjects' ranges (across 12 semitones for subject 6, but not at all for subject 4). Very little R_2:$2f_0$ tuning was observed (0 - 5 semitones) and this was mostly around the middle part of the frequency range investigated (C5). The resonance tuning pattern observed for the /u/ vowel was similar to previous studies, with extensive R_1:f_0 tuning (over one-and-a-half octaves for subjects 3 and 6) and a little R_2:$2f_0$ tuning (0 - 7 semitones).

6.2.2 Two-Dimensional MRI measurements

To obtain measurements of the vocal tract articulators, to compare to the resonance measurements already collected, the MRI images were imported into ITK-snap [175] for analysis. In ITK-snap, the "annotation" tool was used to directly measure the dimensions of the vocal tract in the mid-sagittal plane. After Echternach et al. [6], the parameters measured were: lip opening a, jaw opening b (between fixed points on the lower and upper jaws), height of tongue dorsum c (height of highest point above fixed point on jaw), jaw protrusion d (from the wall of the pharynx), oropharynx width e, and uvula elevation f (relative to hard palate). In addition to these, the oropharynx breadth g (perpendicular to e), larynx height h (relative to the collarbone), lip spreading i, and vocal tract length j (the length of the mid line of the vocal tract calculated with the area function - see section 6.2.3), were measured. The larynx height was measured by taking the distance of the larynx to a fixed point (the collarbone) for all sung notes and the "neutral" position, then subtracting the distance for the neutral position. The mid-sagittal measurements are shown in Figure 6.4.

These measurements were collated using Microsoft Excel [157] and later imported into MATLAB for analysis (see section 6.3). The two-dimensional measurements are included in Appendix L.

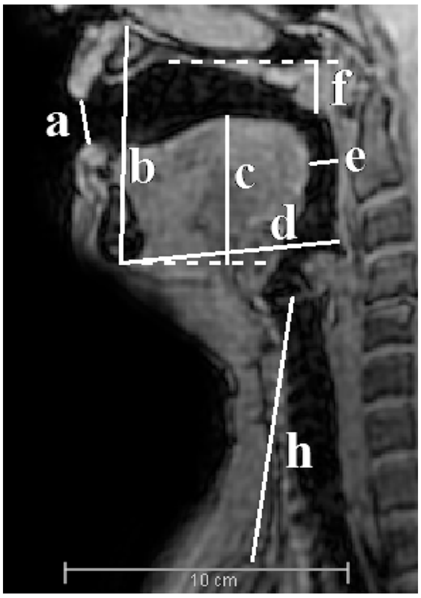


Figure 6.4: 2D MRI measurements: lip opening *a*, jaw opening *b*, height of tongue dorsum *c*, jaw protrusion *d*, oropharynx width *e*, uvula elevation *f* and larynx height *h*. (oropharynx breadth *g*, lip spreading *i*, and vocal tract length *j* not shown). Figure after [6].

Since the vocal tract articulators are all physically connected, a high degree of correlation was expected between the variables. To investigate the type and degree of the relationships between variables, the linear correlation between each pair of variables was calculated and plotted [173]. Figure 6.5 shows the correlation between variables for all singers, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom). Since only four of the six singers completed the MRI part of the experiment, only results for these singers could be analysed in this way.

From Figure 6.5 it can be seen that, as expected, there is a high degree of correlation between variables and the pattern of correlation varies between vowels.

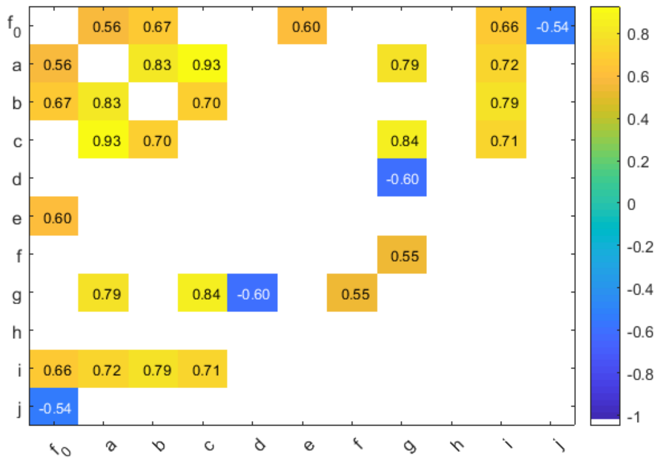
Several correlations were found to be consistent across all three vowels; for all three vowels, the lip opening a and jaw opening b showed positive correlation with fundamental frequency, which is expected based on acoustic theory [19]. Both the lip opening a and lip spreading i also correlated positively with jaw opening b, which is expected, as both these measures describe the degree of openness of the singer's mouth.

The number of statistically significant correlations only varied a little between vowels, with 15/55 found for the /ɑ/, and 23/55 and 21/55 for the /u/ and /i/ vowels, respectively.

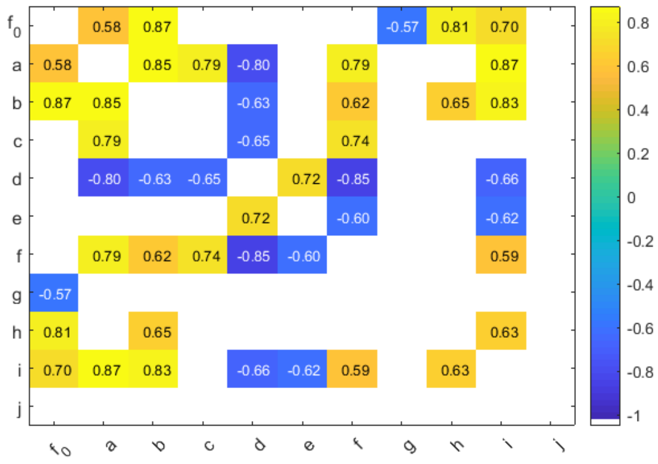
For the /ɑ/ vowel, the variables found to correlate with others most often (5 times) were the fundamental frequency and lip opening a. The jaw opening b, tongue height c, oropharynx breadth g and lip spreading i also appeared often (4 times).

The results for the /u/ vowel show that the lip spreading i was found to correlate most often (7 times), and the lip opening a, jaw opening b, jaw protrusion d, uvula elevation f and lip spreading i also frequently showed a significant correlation (6 times each).

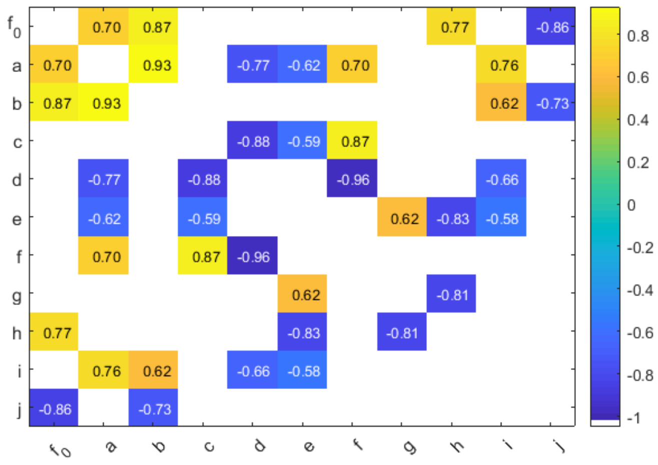
Finally, for the /i/ vowel, the variable showing the most correlation was the lip opening a (6 times), followed by the oropharynx width e (5 times).



(a) /ɑ/ vowel



(b) /u/ vowel



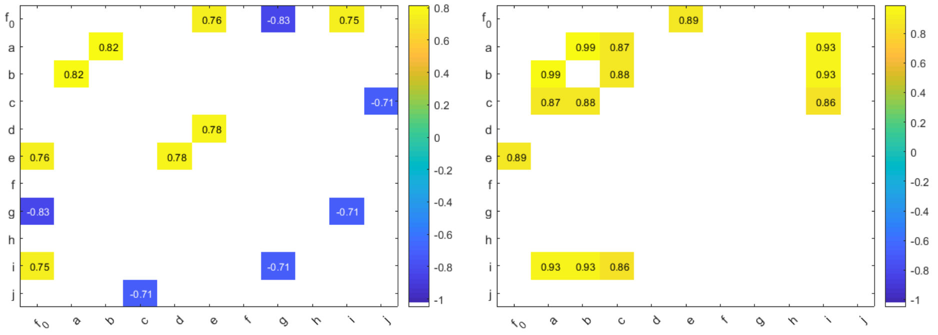
(c) /i/ vowel

Figure 6.5: The linear correlation between all variables[1] for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom), for all subjects. Non-significant results are not shown.

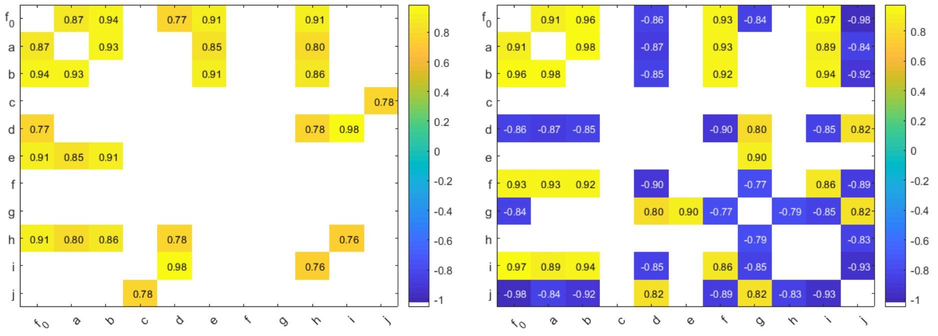
To investigate the differences between singers, the same correlation was calculated again, but split by singer as well as vowel. These results are shown in Figure 6.6. Unfortunately the number of articulator measurements obtained was not sufficient to calculate the correlations for subjects 3 and 6 (as the sets of measurements were not complete for all fundamental frequencies - see data in Appendix L), but the correlations between variables for subjects 1 and 2 show distinct differences, both in the variables that show correlations between each other and in the type of correlations (positive or negative). This may imply that the different variables either serve different functions (in terms of resonance tuning), depending on the vowel sung, or do not relate to the resonances at all. In the latter case, the singer may have developed a habitual movement with a negligible bearing on the acoustic output (although based acoustic theory, it seems unlikely that articulator movement could have no acoustic effects).

---

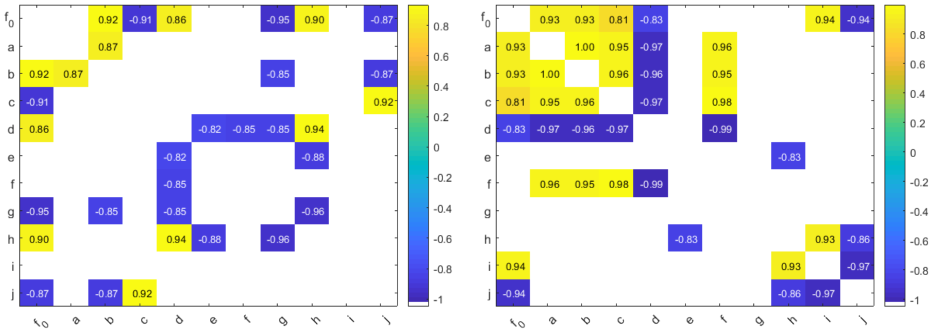
[1] As described in section 6.2.2, these variables are: lip opening *a*, jaw opening *b*, height of tongue dorsum *c*, jaw protrusion *d*, oropharynx width *e*, uvula elevation *f*, oropharynx breadth *g*, larynx height *h*, lip spreading *i*, and vocal tract length *j*.



(a) Subject 1, /ɑ/ (b) /Subject 2, /ɑ/



(c) Subject 1, /u/ (d) /Subject 2, /u/



(e) Subject 1, /i/ (f) /Subject 2, /i/

Figure 6.6: The correlation between all variables[1] for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom), for subjects 1 and 2. Non-significant results are not shown.

6.2.3 Generation of 3D Area Functions

Once the two-dimensional articulator measurements had been generated in ITK-snap [175], the airway was segmented to produce a 3D vocal tract volume. This was done using the built-in feature of ITK-snap, which is designed for the analysis of anatomical structures and performs a segmentation based on the image contrast. As discussed in section 2.2.3, due to the lack of hydrogen in teeth, these were indistinguishable from air. In this experiment, the position of the teeth was estimated by inspecting all the MRI images for each subject, as in some of these the subjects had their tongue or lips pressed against their teeth, allowing the position of the teeth to be deduced in other scans and the teeth to be removed from the segmentation. The MRI segmentation procedure is prone to leakage into the surrounding tissues, so the segmentation volume was allowed to expand until the whole volume of the vocal tract had been filled and the volume had continued out of the mouth to form a rough “radiation dome”.

The segmentations were therefore inspected and the parts representing teeth (incorrectly included in the segmentation) were removed by hand, as well as any leakages into the surrounding tissue. Finally, the radiation dome was defined as ending at the narrowest point between the lips and removed. The segmentation was then imported into ParaView [176] and exported as a list of 3D points on the surface of the vocal tract, as well as connectivity data for the points to be loaded into MATLAB [68] for analysis. All segmentations are included in Appendix M.

The positive x direction was defined as transverse (left-right), the positive y direction as anterior-posterior (front-back), and the positive z direction as superior-inferior (up-down). All measurements were taken in mm.

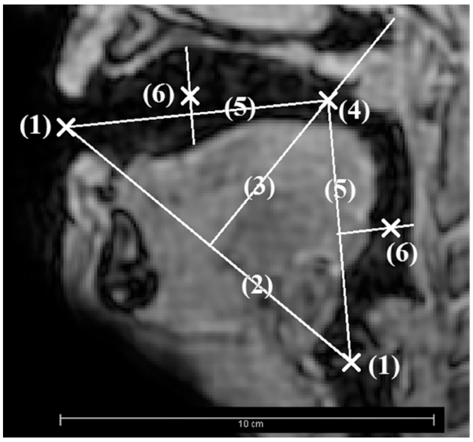


Figure 6.7: Illustration of the algorithm to determine slicing of vocal tract.

Then, following an algorithm originally developed to analyse upper airway geometry and volume with regard to sleep disorders [177] and adapted to generate a 2D area function from a mid-sagittal slice [102], the area function was calculated using an iterative bisection algorithm.

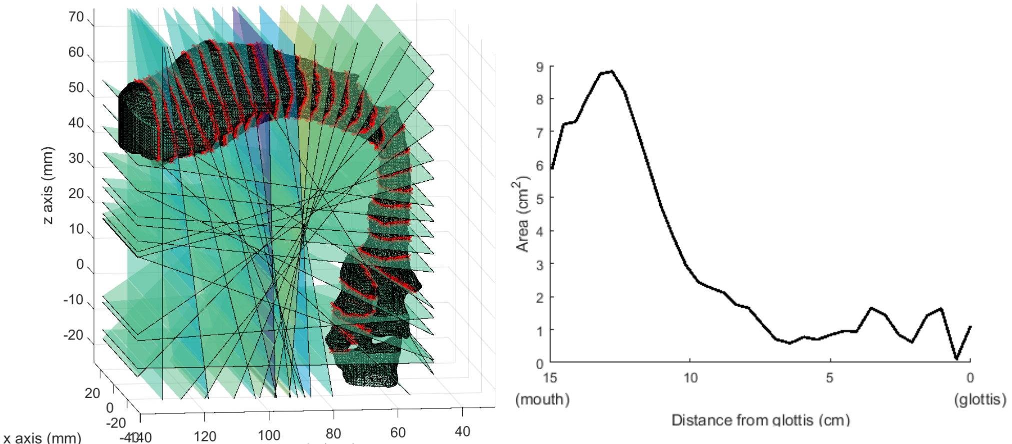
Unlike existing methods of calculating 3D area functions, such as those employed by Baer et al. [119], which assume that the two ends of the vocal tract are straight and the middle section has constant radius of curvature, this algorithm adapts better to the shape of the vocal tract.

The start (glottis) and end (mouth) of the vocal tract were manually defined by the researcher (as the narrowest point of the glottis, and the point between the lips at the narrowest point, respectively) and labelled as (1) on Figure 6.7.

Firstly, the line joining the start and end of the vocal tract was calculated (2) and then a plane was defined at the midpoint of this line, normal to it (3). The intersection of this plane with the vocal tract was found and its area and centre were then calculated (4). The coordinates of the centre of this area was stored as a point, the full set of which defined the mid line of the vocal tract. This process was repeated between the start of the vocal tract and the midpoint, and between the midpoint and the end. This "sliced" the vocal tract into quarters (5) and the areas and midpoints of these intersections (6) were again calculated.

This slicing was repeated (slicing into eighths, sixteenths etc.) to produce a vocal tract cut into 2^n parts. The areas of the start and end points were also included, with the first "slicing" plane defined as horizontal (xy), and the last as vertical (xz). This yielded an area function of $2^n + 1$ slices, and in this study, n was chosen to be 5, giving 33 slices in total. This value was found to provide a sufficient level of detail for analysis, while not taking an excessively long time to calculate. Increasing the number of slices above this was found to introduce errors in the area function produced, due to slices overlapping. An example of the 3D vocal tract mesh, with the planes used to slice it, is shown in Figure 6.8(a) and the area function generated by this is shown in Figure 6.8(b). The MATLAB code used to produce the area functions is included in Appendix L.

A number of restrictions were implemented in this procedure to make the process more robust. Firstly the x component of the centre of each area slice was restricted to the midpoint of the previous and following x components, to prevent transverse shifts. In addition to this, the "slicing" plane was forced to face forwards, (the x component of the normal was made zero), to reduce the likelihood of areas overlapping with the previous or following ones.

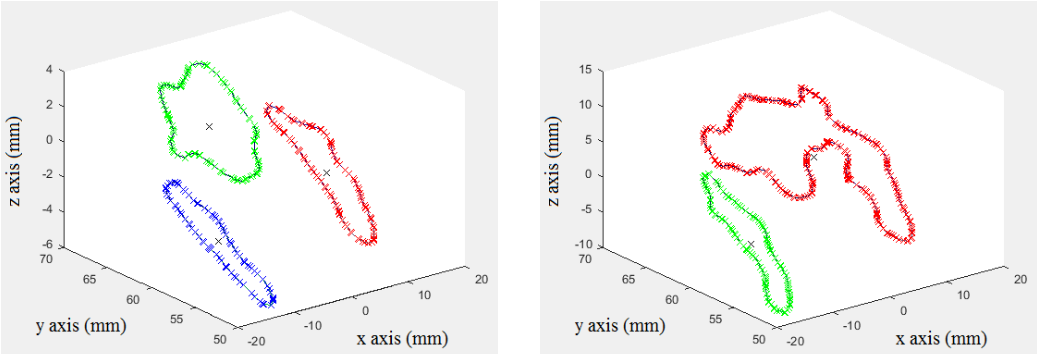


(a) 3D mesh of vocal tract, sliced by 33 planes.

(b) Area function produced

Figure 6.8: An example of the planes used to generate an area function (left), and the resulting area function generated (right).

Some difficulty was encountered in analysis due to the piriform fossae and other side branches, as in some cases the intersection of the vocal tract with the "slicing plane" produced more than one area. If there was more than one separate area identified, the most central one was chosen and its area calculated. Due to the slight asymmetry of the piriform fossae, however, this meant that occasionally one (or part of one) of them was included in the area (as it was not quite separate from the main area of the vocal tract), while the other one was discarded. This led to some error in the measurements of cross-sectional area in the region around 1-2 cm from the glottis. An example of this is shown in Figure 6.9. For the same subject, vowel and pitch as Figure 6.8, the 4th plane from the glottis slices through three separate areas (Figure 6.9(a)), however the 6th plane (Figure 6.9(b)) only identifies 2 areas.



(a) Slice through the 4th plane from the glottis.

(b) Slice through the 6th plane from the glottis.

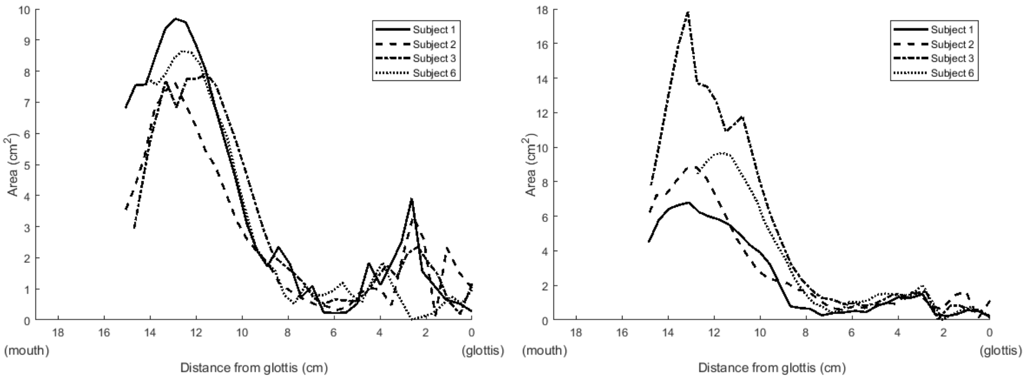
Figure 6.9: An example of two slices through the vocal tract used to generate the area function; the 4th and 6th slices show how the cross-sectional areas of side branches may be inadvertently incorporated into the central vocal tract area.

Although the resonances of the vocal tract could be calculated directly from the area functions generated from MRI images, effects such as the radiation impedance at the subject's mouth, or the wall compliance within the vocal tract are not precisely known and so could not be taken into account. Since the resonance measurements made in this experiment (using broad band noise excitation) measure the resonances directly, they can be assumed to be already taking these effects into account.

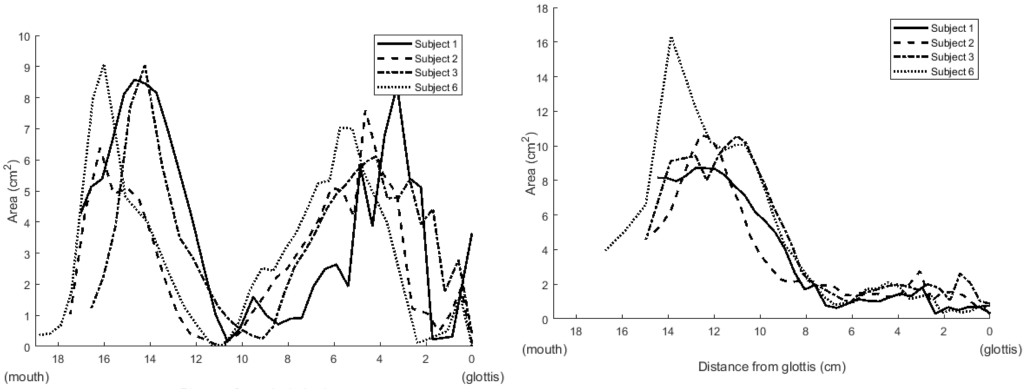
Using the algorithm described earlier in this section, the area functions were calculated for each MRI scan. The area functions generated for each subject individually are shown in Appendix G, Figures 13 (subject 1), 14 (subject 2), 15 (subject 3) and 16 (subject 6).

The results are summarised in Figure 6.10 which shows the area functions for the lowest note sung by all singers (E4 - 330 Hz), and the highest note sung by all singers (G5 - 784 Hz). As discussed in section 5.2.2, not all singers reached the same range of frequencies, so the highest note common to all singers was chosen here.

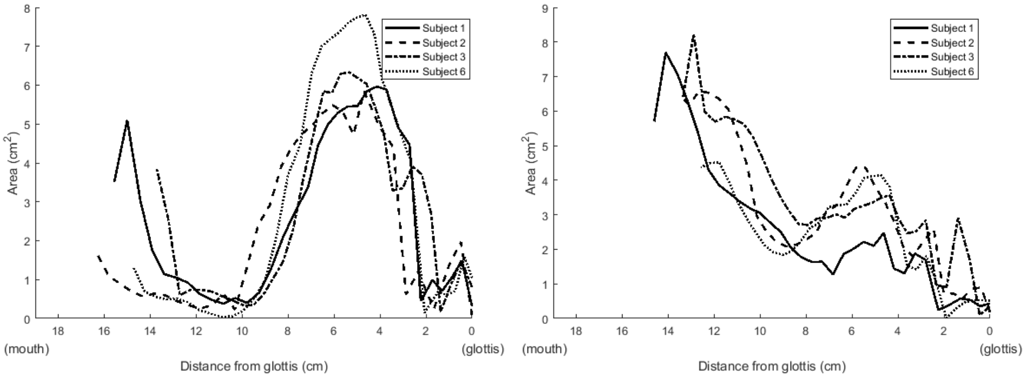
The area functions for the /ɑ/ vowel (Figure 6.10(a) and (b)) are charac-



(a) E4, /ɑ/ vowel (b) G5, /ɑ/ vowel



(c) E4, /u/ vowel (d) G5, /u/ vowel



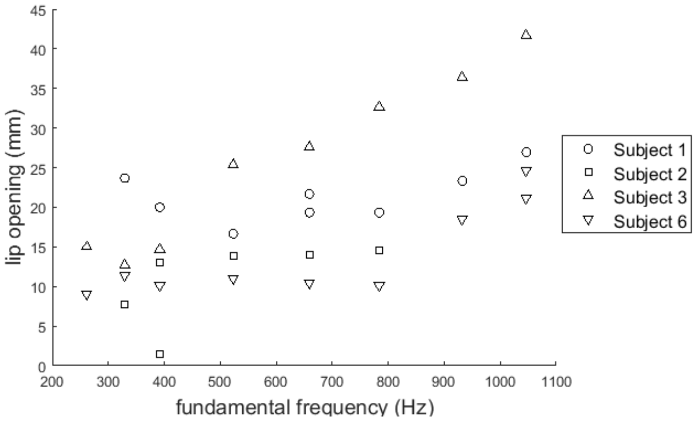
(e) E4, /i/ vowel (f) G5, /i/ vowel

Figure 6.10: A plot of area functions for all subjects, (a-b) for the /ɑ/ vowel, (c-d) for the /u/ vowel and (d-e) for the /i/ vowel, (a, c, e) on the lowest note sung, E4, and (b, d, f) highest note sung, G5.

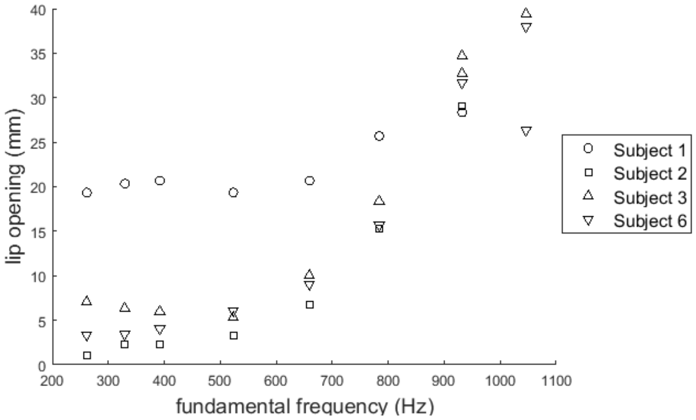
terised by an approximately bell-shaped vocal tract: narrowing to approximately 1 cm^2 around 6 - 7 cm from the glottis (right side of Figure 6.10) (around the back of the tongue), then opening out around 13 cm, before narrowing again near the mouth (left side of Figure 6.10). Although the extent of the mouth opening varies for different fundamental frequencies (between 3 and 7 cm^2), there does not appear to be any relationship between fundamental frequency and mouth opening, and no consistent trends with fundamental frequency can be observed across all subjects.

The area functions for the /u/ vowel (Figure 6.10(c) and (d)) are very consistent between singers; at the lower fundamental frequencies a large area is observed of about 8 cm^2 around the pharynx (approx. 5 cm from the glottis), which then decreases to a very small cross-sectional area around 12 cm from the glottis and then opens up a little, before a final restriction at the mouth. For the higher fundamental frequencies, the shape is very similar to the /ɑ/ vowel, with a narrowing around 6 cm, then a large opening up to approximately 14 cm^2, before a slightly smaller mouth area. At certain points along the vocal tract, a relationship between the cross-sectional area and the fundamental frequency can be observed. For example, around the pharynx (5 cm from the glottis), the lowest fundamental frequency has the highest area and the highest fundamental frequency has the lowest area. The opposite effect is seen at 13 cm from the glottis, where the highest fundamental frequency has the lowest area and vice versa. A noticeable shortening of the vocal tract is also seen with increasing fundamental frequency, possibly due to the corners of the mouth being pulled back, changing its effective length.

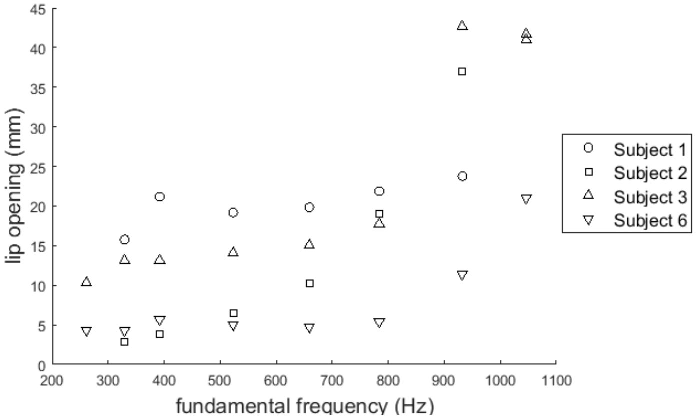
The same patterns between the cross-sectional area and fundamental frequency are also seen for the /i/ vowel (Figure 6.10(e) and (f)): at the mouth the lowest fundamental frequencies has the lowest cross-sectional areas, 4-6 cm from the glottis (pharynx) the lowest fundamental frequencies have the highest areas (approximately 6 cm^2) and a shortening of the vocal tract is observed with increasing fundamental frequency. For both this vowel and the /u/ vowel, the area function at high frequencies is very similar to that of the /ɑ/ vowel.



(a) /ɑ/ vowel



(b) /u/ vowel



(c) /i/ vowel

Figure 6.11: The lip opening for each singer, plotted against fundamental frequency. Figure (a) shows the /ɑ/ vowel, (b) shows the /u/ vowel and (c) shows the /i/ vowel. Not all fundamental frequencies and vowels were captured for all singers (see Table 5.2).

Figure 6.11 shows the lip opening plotted against fundamental frequency for all three vowels. It can be seen that while there is a positive correlation between lip opening and fundamental frequency for the /u/ and /i/ vowels (Figure 6.10(b) and (c) respectively), this is less pronounced for the /ɑ/ vowel (Figure 6.10(a)).

The area functions are similar between subjects, with the same general changes in area function observed for all subjects.

6.3 Statistical Analysis

To address the research question identified in Chapter 1, *investigating how singers achieve resonance tuning*, the statistical analysis carried out in this section aims to link the datasets obtained. These consist of vocal tract resonance measurements (obtained by broad band noise measurement) and parameters describing the vocal tract shape (articulator measurements obtained from MRI images).

As it was not known in advance what the relationship between the predictor variables (articulator measurements) and the response variables (resonances) would be, it was necessary to initially consider all the available variables. A model of the lowest complexity possible could then be created by excluding all but the necessary variables. The number of predictor variables was also very large (11) for traditional methods such as regression analysis, so a method to reduce this number was required.

6.3.1 Feature Selection

The algorithm used to explore the links between articulator measurements and vocal tract resonances is similar to feature selection [178], which aims to reduce the dimensionality of a model by including only a subset of the predictor variables that provide the best prediction of the data. The method used in this work was an “all possible subsets” method [179], which involved generating regression models for all possible subsets of variables, then evaluating which of these models represented the best fit of the data. In this case, since there were 11 predictor variables, there were $2^{11} - 1 = 2047$ possible subsets of variables (not including the empty set). This method is very useful for dealing with data with a large number of predictor variables [178]. Its main disadvantage is generally its high computational cost. However, in this analysis computation speed proved to be adequate.

The model generation was achieved using an iterative algorithm, which involved the following steps:

1. A small amount of random noise (equivalent to 1 mm variation) was added to the data (as explained below);
2. a list of all possible subsets of the variables was generated;
3. for each subset of variables (with a constant term added to represent the intercept [180]), a regression model was generated;
4. the mean squared error (MSE) for each model was calculated and the model with the smallest MSE was selected and stored;
5. steps 1-4 were repeated 1000 times (with random noise re-generated for each iteration);
6. a bar chart was plotted, showing the number of times that each variable appeared in the model over the 1000 iterations, and the variables appearing most often were selected for a final regression model;
7. a final regression model was calculated using the variables selected in step 6 and a constant term.

This analysis was carried out in MATLAB [68], and the code is included in Appendix L. As described in point 6, a bar chart showing the most selected variables was generated for each vowel and singer. An example is shown in Figure 6.12, for the second resonance of the /u/ vowel.

Adding normally-distributed random noise to the data at each iteration of the algorithm has an effect similar to *regularisation.* Adding small variations to the data improves the robustness of the model and prevents *overfitting* [181], where the model fits the training data too well and is therefore unnecessarily complex and ultimately inaccurate.

The quality of fit of the final regression models can be evaluated by examining firstly the R^2 values, which should be as close as possible to 1, secondly, the p-values, which should be below the chosen significance level of 0.05, and finally the residual errors, which should be random, rather than patterned.

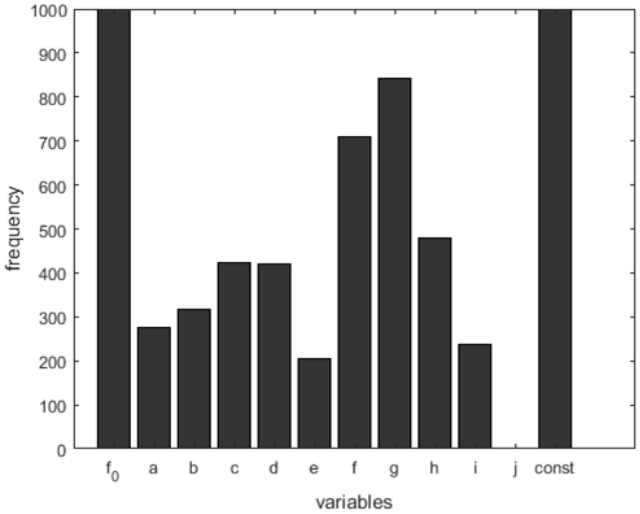


Figure 6.12: Bar chart showing the number of times that each of the 12 variables was selected by the feature selection algorithm (fundamental frequency, 10 articulator variables (a-j[2]), and a constant term), for the second resonance of the /u/ vowel.

---

[2] As described in section 6.2.2, these variables are: lip opening *a*, jaw opening *b*, height of tongue dorsum *c*, jaw protrusion *d*, oropharynx width *e*, uvula elevation *f*, oropharynx breadth *g*, larynx height *h*, lip spreading *i*, and vocal tract length *j*.

6.3.2 Evaluation of Models

The main method of evaluating the quality of fit of a regression model is by examining the residual error of the model. This has been done for all nine vowels and resonances in Figure 6.13. From this plot it can be seen that the residuals are random and do not appear to follow any pattern or have an envelope shape. A test for randomness was applied using the MATLAB function *runstest* [182]. The test was applied to the residuals for each of the nine models above and returned the null hypothesis (showing that the data was random) for eight of the nine models. Applying the test to the model for the /u/ vowel did not give a random result. However, performing the regression individually for the different singers *did* produce random residuals. This is encouraging in the context of this thesis, as it suggests that the significant models generated fit the data well and that there are no missing variables or non-linear effects that have not been accounted for. The magnitude of the residual errors suggests that the data contains a large amount of variation, which is to be expected, as the data consists of measurements of real movements.

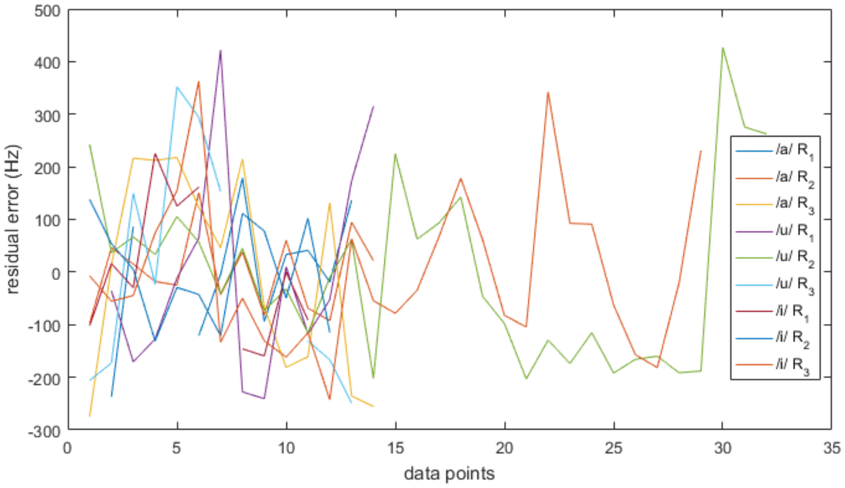


Figure 6.13: Residual errors of the nine regression models generated (three vowels x three resonances), showing random errors in the models.

Although regression analysis provides a useful tool for evaluating the influence of various articulators on the vocal tract resonances, this method should be treated with caution, as regression analysis is not ideal in cases where there exists a large degree of multicollinearity between variables. When the vocal tract system is considered as a whole, it is clear that since the articulators are all physically linked together, their motion is similarly tied together and the correlation between variables is high. The correlation between variables for each vowel was shown in Figure 6.5. From this it was observed that for all vowels there existed a large amount of correlation between variables, as expected.

One method commonly used to reduce the effects of multicollinearity is principal component analysis [183], which reduces collinear variables to new, orthogonal variables. However, although this would improve the regression models able to be generated, the new variables generated by principal component analysis would not be particularly meaningful, compared with the original real measurements. It is important therefore not to lose sight of the purpose of this research, which is not to create a perfect statistical model of vocal tract resonance production, but rather to use a more holistic approach to gain an insight into how the system functions, and regression analysis provides a useful method of determining the relative influence of different variables.

6.3.3 Regression Models

For each resonance and vowel, a regression model was generated for all singers using all the “all possible subsets” method described above to select the most appropriate variables to fit the data. The nine resulting regression models (3 vowels x 3 resonances) are given in Table 6.2.

Of the nine models generated by the feature selection algorithm (three vowels x three resonances), only three of these produced a good fit of the data (R^2 value of 0.7 or above to 1 decimal place): the models for the second resonance

Model	Regression equation	R^2	F	p value
/ɑ/ - R_1	-7.42*j + 1996.75	0.11	0.98	0.35
/ɑ/ - R_2	10.90*b -19.07*d -9.79*j + 3347.23	0.68	7.07	0.01
/ɑ/ - R_3	-18.93*j + 5870.94	0.21	3.25	0.10
/u/ - R_1	-1.91*j + 941.38	0.03	0.35	0.57
/u/ - R_2	0.91*f_0 + 18.23*f + 1.24*g + 285.18	0.74	27.22	0.00
/u/ - R_3	0.44*j + 2908.36	0.00	0.01	0.92
/i/ - R_1	-17.86*j + 3225.55	0.53	8.90	0.02
/i/ - R_2	18.27*j -660.58	0.76	31.79	0.00
/i/ - R_3	0.12*f_0 -6.85*d + 7.50*g + 3001.60	0.08	0.72	0.55

Table 6.2: The regression models generated for each vowel and resonance investigated. Models with a good fit (high R^2 value) are shaded grey.

for all three vowels. In addition to this, the model for the /i/ vowel, first resonance, had a quite good fit (R^2 value of 0.53).

Of the variables included (identified in section 6.2.2), the variables that appeared in the nine regression models were:
f_0 - *fundamental frequency*
b - *jaw opening*
d - *jaw protrusion*
f - *uvula elevation*
g - *oropharynx breadth*
j - *vocal tract length*

This suggests that these are the variables that are the most important when considering resonance production. Both the jaw opening and jaw protrusion (variables b and d) feature in the model for the second resonance of the /ɑ/ vowel. Based on the tube acoustics of the vocal tract, previous studies [184] have suggested that the tongue position is very important for the second resonance, however the tongue height was not selected by the feature selection algorithm for any of the vowels or resonances investigated. The range of variables included in the regression models for R_2 suggest that this resonance is actually affected by several articulators jaw opening (b, jaw protrusion d, uvula elevation *f*, oropharynx breadth g and vocal tract length j).

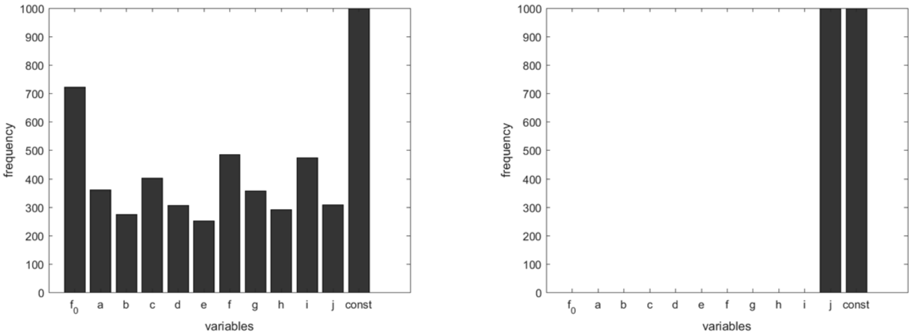
The only vowel where the feature selection algorithm produced a model with quite good fit for R_1 was for the /i/ vowel (R^2 value of 0.53), and only included the vocal tract length. In addition to this, the coefficient pertaining to the vocal tract length was almost always negative, suggesting that shortening the vocal tract raises the resonances. This is not surprising based on acoustic theory - shortening the length of a tube raises the frequencies of all its resonances. However, based on previous understanding of the effects of jaw opening on the first resonance [67], it is surprising that the jaw variables are not included.

The vowels/resonances for which the feature selection algorithm failed to produce a good fit (R^2 value below 0.5) mostly included just the vocal tract length. This implies that this is one of the most important factors in determining the vocal tract resonance. Interestingly, it was chosen in preference to the fundamental frequency, implying that, although there was some correlation between the vocal tract length and fundamental frequency, the vocal tract length is a better predictor of the vocal tract resonances.

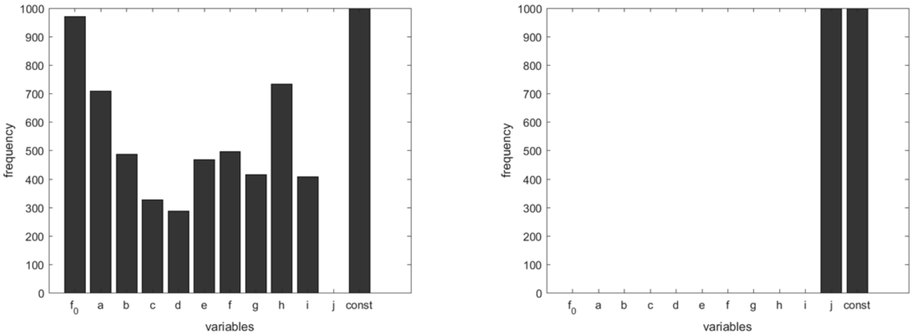
Unfortunately, a large proportion of the measurements for the third resonance were not of suitable quality (see Chapter 5, Table 5.3), as it was difficult to accurately measure the resonances at high frequencies, so it is not surprising that this method failed to generate a suitable regression model for this resonance.

Feature selection split by singer

In order to explore the differences between singers, the feature selection process was repeated for individual subjects. The bar charts showing the frequency with which each variable was chosen by the feature selection algorithm for the first resonance of the /i/ vowel is shown in Figure 6.14. The full results for all vowels, resonances and subjects, are shown in Appendix H.



Subject 1 Subject 2



Subject 3 Subject 6

Figure 6.14: The variables chosen by the feature selection algorithm, for R_1 for the /i/ vowel, split by subject (The 12 variables are fundamental frequency, the 10 articulator variables (a-j), and a constant term ("const")).

The reduced datasets do not allow statistically meaningful regression models to be generated, but it can be seen that for some of the resonances and vowels that did not produce a good model for all subjects, the feature selection process selected different variables for the individual singers.

From Figure 6.14, it can be seen that when the data was split by subject, the variables chosen by the feature selection algorithm were not the same. For subject 1, the fundamental frequency (and constant term) appeared the most often. For subjects 2 and 6, only the vocal tract length was chosen. For subject 3, however, the variables appearing most often were the fundamental frequency, lip opening a, and larynx height h.

This suggests that the reason why no good general model was generated for some resonances and vowels, was because the subjects were using different methods to alter their resonances. A much larger dataset would be required to fully investigate this with any statistical robustness, including many more singers and multiple repeats. This is discussed in further detail in section 6.4.1.

6.3.4 Singers' Experience (continued from Chapter 5)

As discussed in section 5.3.4, the subjects completed a questionnaire to investigate the effects of the MRI conditions on their singing and on their perception of their own singing and singing production. This aimed to probe the singers' understanding of the specific techniques they used to sing in the upper part of their ranges and their understanding of how they achieved this (articulator movements, posture, visualisations etc.).

All the subjects described the changes they made to the vocal tract for singing high notes very similarly, mentioning lifting the palate, dropping the jaw, and being aware of their breath support. When asked if these changes differed between vowels and if any were more difficult to sing, subject 1 said that the /ɑ/ was "tricky" and the shape of the /u/ vowel was hard to maintain. Subjects 2 and 3 found the /i/ vowel difficult and subject 6 found the /u/ vowel hardest to sing at high pitches.

After the experiment, the subjects were again asked to describe the changes they made to the vocal tract for singing high notes and whether these changes differed between vowels. Subjects 1, 3 and 6 all said that they found the /u/ vowel hardest, which represented a change in perception for subject 3, who had said that /i/ was hardest when asked before the experiment commenced.

6.4 Discussion

Previous studies on resonance tuning in sopranos [14, 15] generally found that for the /ɑ/ vowel, singers employed $R_1{:}f_0$ tuning extensively, in conjunction with $R_2{:}2f_0$ tuning, both beginning at around C5. For the /u/ vowel, similar patterns were observed, with either similar or little (up to a couple of tones) more $R_1{:}f_0$ tuning than for the /ɑ/ vowel (the /i/ vowel was not investigated). In this current study, the type of resonance tuning observed was broadly as expected, based on acoustic theory, which predicts that sopranos should increase sound production by tuning resonances to nearby harmonics whenever possible. However, for the /ɑ/ vowel, the subjects in this experiment used surprisingly little $R_1{:}f_0$ tuning. Five of the six singers in this experiment did however employ $R_1{:}2f_0$ around G5. One explanation for this difference in behaviour may be the higher level of experience of the singers in this study. The singers in the previous works were described as ranging from, "nationally recognized to amateur singers" [14], whereas the subjects in this study were either national or international principals.

The changes in area functions with fundamental frequency were generally as expected for the /u/ and /i/ vowels, showing increased mouth opening, leading to a shorter vocal tract length, more open jaw and a reduced pharynx space. These results were reflected in the subjects' comments on their techiques (section 6.3.4) and are also consistent with the observations of Echternach et al. [6], who also observed widening of the lips and greater jaw opening, as well as elevation of the tongue dorsum in soprano singers at high fundamental frequencies.

The area functions generated for the /ɑ/ vowel, however, showed a lack of clear changes in area function with fundamental frequency, very different from the /u/ and /i/ vowels, which both showed very clear patterns of resonance tuning and clear changes in area function. This is unexpected, as a common idea in singing teaching [3] is that singers should attempt to "neutralise" their vocal tract shape to sing at very high fundamental frequencies, losing the distinction between the vowels. The /ɑ/ vowel is generally considered an easy vowel to produce and other vowels tend towards it at high frequencies.

The comments from the questionnaire suggest that the subjects were aware of this idea. For example, when asked if the changes they made to the shape of the vocal tract for high notes varied between vowels, subject 3 commented that the /i/ vowel tended towards an /ɑ/ sound at high pitches. The area functions generated in this experiment support this idea to an extent, as the area functions for different vowels indeed became very similar at high fundamental frequencies. They are not completely identical, however, as shown in Figure 6.15, which shows the highest and lowest fundamental frequency area functions averaged across all 4 subjects (Figure 6.10 shows the area functions for each singer individually, which are generally consistent between subjects). It can be seen that although at high frequency the /ɑ/ and /u/ vowels were very similar, the /i/ vowel had a slightly smaller mouth area (left side of Figure 6.15) and slightly larger pharynx area (4-6 cm from the glottis at the right side of Figure 6.15).

Inspection of the resonance tuning results and area functions showed that there were strong differences in both of these between the three vowels investigated in this experiment. However, there do not seem to be very great differences between subjects. Although the resonance tuning strategies and area functions are not exactly the same, they appear to follow very similar patterns with changes in fundamental frequency. Even the resonance tuning results for subject 4 were consistent with the other singers, in spite of failing to complete the MRI part of the experiment and showing different changes between “normal” and “simulated MRI” conditions from the other subjects (as discussed in section 5.3.2). The results were also consistent with results from previous studies [14, 15, 185] (allowing for differences due to different levels of experience).

The results of the feature selection process and regression analysis seem to suggest that the most important factors affecting resonances are the funda-

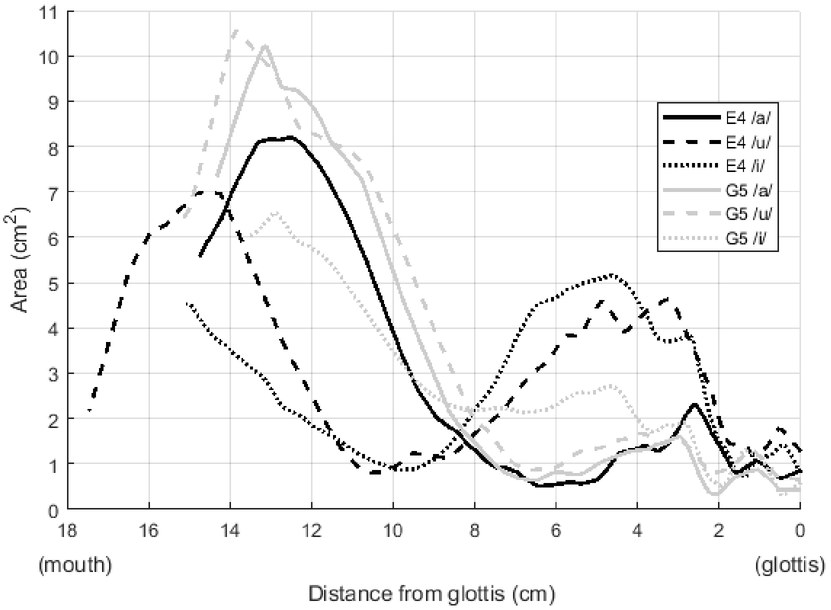


Figure 6.15: Area functions for each vowel, for the lowest note (E4 - black) and highest note (G5 - grey) sung by all subjects, averaged across all 4 subjects. The area functions for individual singers are shown in Figure 6.10.

mental frequency, jaw opening b, jaw protrusion d, uvula elevation f, oropharynx breadth g and vocal tract length j.

Both the jaw opening and jaw protrusion (variables b and d) feature in the model for the second resonance of the /ɑ/ vowel. The importance of these variables is to be expected, as tube acoustics suggest that the most effective way to raise the first resonance of a tube is to widen the opening [67]. Previous studies have suggested that the tongue position is very important for the second resonance [184], although the tongue height was not selected by the feature selection algorithm for any of the vowels or resonances investigated.

Not surprisingly, the analysis suggested that the vocal tract length was a strong influence on all the vocal tract resonances. Not only was this variable chosen by the feature selection algorithm, but it also appeared in many of the non-significant models (more often than fundamental frequency), suggesting that the vocal tract length is a better predictor of the vocal tract resonances than fundamental frequency. It is to be expected that the vocal tract length is inversely related to the vocal tract resonance and that the resonance frequencies are not necessarily related to the fundamental frequency, since this is controlled by the vocal folds. The regression model coefficient pertaining to the vocal tract length was almost always negative, suggesting that shortening the vocal tract raises the resonances. Again this is in agreement with acoustic theory: shortening the length of a tube raises the frequencies of all its resonances.

The importance of the lip, tongue and jaw correlates well with previous studies, such as Sundberg and Lindblom [67, 184]. However, the results of this work are not as conclusive as those of Sundberg and Lindblom. One reason for this may be the large sample size in this study, compared to typical values for voice research.

It is, however, a little surprising that the regression model calculated for the second resonance of the /u/ vowel included the uvula elevation f, as this has not previously been considered as an articulator. However, the uvula is part of the soft palate, so the "uvula elevation" more accurately describes the motion of the soft palate, which is known to be important in singing, not least because it separates the vocal tract from the nasal cavity [186]. It is surprising that this variable was only selected for this vowel, however, and not /ɑ/ or /i/.

The significant advantage of three-dimensional magnetic resonance imaging (compared to two-dimensional) is, of course, that it provides the opportunity to study the transverse properties of the vocal tract, as well as in the mid-sagittal plane. The oropharynx breadth g (perpendicular to the oropharynx width) was included in the regression model for the second resonance for the /u/ vowel (and in the non-significant model found for the third resonance for the /i/ vowel). This demonstrates the benefits of studying the entire vocal tract, rather than just a mid-sagittal slice.

It is surprising that only three of the nine models generated produced a good fit of the data: specifically the models for the second resonance for all three vowels. From the resonance measurements in Figure 6.1, it could be seen that extensive R_1 tuning was observed for both the /u/ and /i/ vowels for all subjects. Only one model, with a “quite good” fit (R^2 value of 0.53), was found for the first resonance, for the /i/ vowel.

There are several possible explanations for the lack of a model for R_1: either there are additional variables that should have been considered (perhaps variables describing the effects of the cheeks in some way should have been included), or the effects of the articulators influencing the first resonance are non-linear, and so cannot be approximated by linear regression models. This could possibly be true if two or more articulators were working together in a complementary fashion, with each articulator responsible for R_1 over a narrow frequency range. The overall effects would then suggest no significant relationship between the articulators and R_1.

Alternatively, it may be that the subjects used slightly different methods to alter R_1, so no overall pattern was found for all singers together. Both the individual correlation matrices generated in section 6.2.2, and the feature selection split by singer (section 6.3.3) would suggest that this is quite likely, as the correlations between subjects showed considerable variation.

It should be remembered that the lack of a good regression model does not necessarily mean that there are no meaningful relationships in the data. If the four singers employed different methods of altering resonances, then when the data was combined together, this would produce seemingly random relationships between the variables and the resonances. Therefore, although the resonance tuning results and area functions suggest that singers are making similar overall changes to the shape of their vocal tracts with fundamental frequency and producing similar effects on the voice (in terms of resonances), they may be producing these effects in different ways.

This may represent a much more subtle example of articulatory compensation, which refers to the possibility of two different vocal tract shapes producing almost exactly the same acoustic output [187]. This idea is not surprising when it is considered that singers typically learn by making a similar *sound* to their teachers, rather than copying some other aspect of the sound production process. This idea is supported by the acoustics of tubes, as it is possible to produce the same resonances using tubes of different shapes. For example, a vocal coach may ask a singer to "create space" in the throat, which is not only a subjective term, but assuming that the aim is to increase the pharynx volume, something that could be achieved in subtly different ways. One singer could rely on lifting the soft palate, while another might widen the pharynx, and a third could use a combination of these techniques, all producing a very similar effect on the voice. This is hardly surprising when one remembers that, in addition to a unique learning experience, all singers also have a unique physiology - the size, range of motion, and interconnection of different articulators is an individual matter (some physiological differences were evident from inspection of MRI scans).

6.4.1 Limitations

This study was carefully designed to quantify the differences between results of the same or similar tasks across anechoic and MRI conditions, to allow the results relating to resonance tuning to be contextualised for practice, and this was discussed in Chapter 5. Technology permitting, it would have been better to obtain measurements of vocal tract resonances simultaneously with MR image collection, which would remove the need for analysis of similarity between conditions. It is hoped that advances in MRI technology will make this possible in the future.

The current study suffered from the limitations of the MRI scanner that was available, such as the loud noise, and the subject being required to lie supine. Upright MRI is available currently [159], although not yet widespread.

Although feature selection using regression analysis is helpful in identifying which variables (articulators) have the most influence on the vocal tract resonances, the limitations introduced by the large amount of potential multicollinearity between variables restricts its usefulness somewhat, meaning that the significance of this should not be overestimated.

It is also worth noting that producing robust statistical models to fit sets of data generally requires considerably larger data sets than the one in this study. Although this data set is unusually large for a study of the singing voice involving MRI, it still lacks statistical power, and a much larger dataset consisting of many more singers and multiple repeats of measurements (at least 3 repeats, to allow detection of outliers) would be desirable if any very meaningful statistical models were to be obtained. Even if the population of opera singers was of the order of magnitude of a few thousand singers, this would mean that to obtain results with a confidence interval of 5 % would require a sample size of several hundred singers.

It would have been beneficial to carry out repeat measurements of all the data in this experiment, but the current protocol already took approximately four hours to complete and was quite tiring for the singers involved, so this was not feasible as part of this current study.

6.5 Conclusions

This experiment has highlighted the variety of resonance tuning methods used by singers for different vowels and the corresponding vocal tract shapes used to produce these sounds. The uniqueness of this research lies in the richness of the dataset - never before have three-dimensional MR images of the vocal tract been captured for a group of such highly experienced singers, in conjunction with measurements of their vocal tract resonances.

The results of the statistical analysis, bringing together information about the vocal tract resonances and the changes in vocal tract articulators, shows that although some existing ideas about singing production are supported, such as the importance of jaw position and vocal tract length, the effects are not as straightforward as previously thought. The overwhelming message to be gained is that, as with its perception, resonance production is very complex and varies depending on the vowel and fundamental frequency sung, and according to the individual singer.

Differences between singers have been observed in every kind of data examined in this work, showing that singing production is a highly individual matter. This suggests that early work on resonance tuning involving small numbers of singers may benefit from being expanded; larger numbers of very high quality singers would allow expert techniques to be better understood and a wider range of singers would capture a broader variety of resonance tuning behaviour. This has been achieved to some extent with the rich dataset obtained in this work, but it would be beneficial to revisit established theories in addition to this.

Although the resonance tuning methods used by different singers are very similar, the alterations to the vocal tract that produce these effects are not necessarily the same for all singers. That is, although singers achieve a similar result (in terms of vocal tract resonances), this may be arrived at in different ways.

These findings highlight the importance of considering the differences in resonance production between different vowels. Strong differences were seen between vowels for resonance tuning behaviour, area functions and resonance production, which supports the results of the perceptual study in Chapter 3. This work has also emphasised the highly individual nature of singing, illustrating the differences in voice production between singers. It would be highly beneficial to singers if future research in this area investigated the differences in resonance production in singers of different genres, with different levels of experience and across a range of vowels.

Chapter 7

Conclusions

> The heavenly bodies are nothing but a continuous song for several voices (perceived by the intellect, not by the ear); a music which... sets landmarks in the immeasurable flow of time. It is therefore, no longer surprising that man, in imitation of his creator, has at last discovered the art of figured song, which was unknown to the ancients. Man wanted to reproduce the continuity of cosmic time... to obtain a sample test of the delight of the Divine Creator in His works, and to partake of his joy by making music in the imitation of God.
>
> Johannes Kepler, Harmonices Mundi (1618)
> Book V, Ch. 7

This thesis has focussed on the technique of *resonance tuning* in soprano opera singers. Firstly, a subjective test was conducted to explore the perception of common methods of resonance tuning. Following this, the usage and extent of resonance tuning by both girl choristers and adult singers was considered. Finally an investigation into the production of resonance tuning by the movement of the various vocal tract articulators was carried out.

This chapter reviews the experiments carried out in this thesis, their findings and how these relate to the objectives set in Chapter 1. It sets out the conclusions that can be drawn from this research, as well as the implications and potential future work arising from it.

7.1 Summary

Chapter 2, the literature review, introduced the subject matter of this thesis by presenting an overview of voice production and introducing the concept of resonance tuning in soprano singing. Previous studies on both male and female voices were considered and the specific challenges facing soprano singers were discussed. Studies considering resonance tuning in soprano voices were also examined and showed that sopranos have been found to tune their first and/or second resonances to the first or second harmonic. Since this work involved measuring the vocal tract using magnetic resonance imaging, different imaging methods were considered and previous studies utilising MRI to study the voice were discussed.

Chapter 3 described the first experiment carried out for this thesis; a perceptual test to investigate listeners' impressions of synthesised vowel samples, with different resonance tuning strategies, on different vowels and fundamental frequencies. The purpose of this study was to understand the perceptual effects of increasing the acoustic efficiency of the voice using resonance tuning. The perception of resonance tuning was shown to be both very complex and highly dependent on the vowel investigated.

Chapter 4 described a preliminary experiment conducted for the main work of this thesis, which tested the use of broad band noise excitation for measuring vocal tract resonances in girl choristers. It was shown that broad band noise excitation is a suitable method for measuring the frequencies of vocal tract resonances in high voices. This experiment not only provided insights into the use of resonance tuning by girl choristers, but also verified that the software and hardware tested could be used in the main experiment of this PhD research. The software consisted of a program generating random-phase harmonic components spaced 5.38 Hz apart and included a calibration procedure which adjusted the amplitudes of the harmonic components to produce a flat frequency response when the subject's mouth was closed. The hardware consisted of a loudspeaker isolated in a wooden container, insulated with sand, which incorporated an impedance-matching horn and a flexible tube to deliver the sound to the subjects lips.

Chapter 5 described the method used in the experiment forming the main study of this thesis. This was used to arrive at results relating resonances to articulator measurements. The experiment involved both resonance measurements using broad band noise excitation (as evaluated in Chapter 4) and image capture during singing using MRI. This chapter also discussed the statistical analyses used to evaluate the effects of MRI conditions on singers in order to assess the suitability of MRI for singing voice research, in anticipation of Chapter 6. The results showed that, although the subjects felt that they had sung differently in the MRI machine, the singing produced in the different conditions was not significantly different in terms of the parameters investigated; the vocal tract resonances, overall frequency content (LTAS) and harmonic content.

Finally, Chapter 6 discussed the main work of this thesis; processing of the MRI data obtained during the experiment carried out in Chapter 5. It analysed both this data and the resonance measurements to investigate the effects of the different articulators on vocal tract resonances. A feature selection algorithm was first used to determine which of the predictor variables (fundamental frequency and 10 different articulator measurements) showed a relationship with the vocal tract resonances and then a regression model was generated using these variables. The resulting models suggested that the production of resonance tuning is more complex than previously thought; although the use of resonance tuning was very similar between singers, this may have been achieved by different articulatory means. Unsurprisingly, considering the physiology of the vocal tract, the results suggested that the variables used (articulator measurements) had a high degree of multicollinearity, so although this analysis gives an indication of the important variables for resonance production, the models are not completely reliable.

7.2 Research Objectives

The main objective of this PhD research, as identified in Chapter 1, was:

1. To understand the methods by which resonance tuning is produced by soprano singers.

To achieve this aim, two sub-objectives were also identified:

2. To better understand the purpose of resonance tuning by studying its perception by listeners.
3. To understand the effects of MRI measuring conditions on singers and verify the usefulness of MRI in singing research.

These objectives were addressed using three experiments: a perceptual test using synthesised voice sounds representing different methods of resonance tuning; an experiment investigating resonance tuning behaviour in girl choristers; and an experiment using MRI to investigate resonance production in adult soprano singers. There are four major contributions to the field from this work, that will be discussed in detail below. In brief, these are:

Perception of Resonance Tuning - understanding of listener perception of different resonance tuning methods, using a perceptual study of different resonance tuning methods, as discussed in Chapter 3.

Resonance tuning - understanding of resonance tuning behaviour in both girl choristers and professional opera singers, two groups not previously studied in great detail (from the preliminary experiment in Chapter 4), and national/international principal opera singers (from the experiment carried out in Chapter 6).

Effects of MRI conditions - as discussed in Chapter 5, an investigation into the effects of MRI scanning conditions on singers, to understand how these affect the production of singing and speech.

Production of Resonance Tuning - an investigation into how different articulators affect the vocal tract resonances, using MR images of the vocal tract articulators, as discussed in Chapter 6.

These four contributions are considered in greater depth below.

7.2.1 Perception of Resonance Tuning

Since singing is a performance art, the first contribution of this thesis is to provide insight into the *production* of resonances, in the form of an investigation into the *perception* of different resonance tuning strategies. This was achieved by means of a perceptual test, where subjects listened to synthesised singing samples with different resonance tuning. Samples were compared based on their *preference* and *naturalness*, and subjects were then asked to *identify* the vowel sounds.

The findings of this perceptual test show that, as with its production, the perception of resonance tuning is highly complex, with no clear tuning patterns seen across all vowels. The perception of different resonance tuning strategies is also highly vowel-dependent, with differences in preference, naturalness and vowel identification noted in the results across the different vowels. The differences in perception between vowels were not surprising considering the differences in speech formant values between them. For example, although the /i/ and /u/ vowels have similar first formants in speech (310 and 370 Hz respectively), they have very different second formant values

(2790 and 950 Hz respectively), causing second formant tuning to strongly alter the sound of the /i/ vowel.

The dependence of perception on fundamental frequency also varied between vowels, with the /i/ vowel showing almost no dependence on f_0 for either preference or naturalness. The results also supported the theory that at high fundamental frequencies, all vowels tend to be "neutralised" towards an /ɑ/ vowel [3], as vowels were most commonly mis-identified as the /ɑ/ vowel.

The results of this experiment therefore suggest that for singers, resonance tuning may be a matter of compromise, as employing a certain resonance tuning strategy might improve one perceptual attribute whilst worsening another. It also suggests that it might be perceptually beneficial for singers to employ resonance tuning techniques over a wider range of fundamental frequencies than has previously been observed. This is not observed in practice, however, and the reason for this is unknown; perhaps there exist physiological restrictions on the vocal tract, or perceptual or acoustic attributes of resonance tuning not yet understood.

7.2.2 Resonance tuning

The second major contribution of this thesis is the new understanding of resonance tuning in girl choristers, who have similar vocal ranges to the adult sopranos studied, and therefore the same motivation for resonance tuning (in terms of vocal efficiency). A preliminary experiment was conducted as part of this PhD research (Chapter 4) to test the suitability of using broad band noise excitation to measure the vocal tract resonances during singing. During this experiment, resonance measurements for girl choristers were obtained, which also yielded information on resonance tuning behaviour in girl choristers. The second experiment (Chapter 6) explored resonance tuning in very highly trained singers who are national or international opera principals. Although previous studies such as [59, 99] have studied "professional" sopranos, the level of these is unclear.

Although prior work has observed resonance tuning in adult singers [14, 15], this thesis has also found evidence of resonance tuning in girl choristers, (extensive R_1:f_0 tuning, and a little (1-7 semitones) R_2:$2f_0$ tuning). This shows that even these experienced young choral singers are capable of employing resonance tuning techniques. It is not known exactly what causes singers to begin using resonance tuning techniques. It may be due to the physiological changes occurring in the vocal tract as their bodies develop, or perhaps an effect of training. The latter is supported by the findings of Garnier et al. [59], who investigated resonance tuning in professional, advanced, and non-expert singers. They suggest that more experienced singers employ resonance tuning methods over a wider range than non-expert singers. However, this theory was not entirely supported by the resonance tuning observed in opera singers in the second (MRI) experiment of this study. Generally, they did not seem to tune their resonances as close to the relevant harmonics, within 60 Hz in this study, as opposed to the 25 Hz in [15].

This more flexible approach to resonance tuning may be due to the very high level of the singers in this study. Once a singer has truly mastered the rules of their craft they may be free to use their own judgement to interpret whether to employ the techniques that they have developed. However, this would require further investigation.

As discussed in section 2.3.2, previous studies on the soprano singing voice have generally included only limited numbers of subjects and limited vowels (most commonly the /ɑ/ vowel). There has not been any previous study considering the /i/ vowel, which is acoustically quite extreme (compared to other vowels) due to its high second formant and representing a very closed vocal tract shape. Both the girl choristers and professional opera singers showed very clear resonance tuning patterns for this vowel (as seen in Figure 4.5): R_1:f_0 tuning across nearly their entire range, but no R_2 tuning. Examination of the resonance measurements shows that the frequency of the second resonance decreased with fundamental frequency until it reached the second harmonic near the top of the singer's range.

7.2.3 MRI Conditions

The second major contribution of this thesis has been the investigation into the effects of MRI conditions on singers. Since MRI techniques are becoming increasingly common in research into the vocal tract, it is becoming more important to understand how the measurement process influences the results obtained and to be aware of any measurement errors introduced during the process. The effects of MRI conditions were investigated in Chapter 5, which compared speech and singing produced in "normal" and "simulated MRI" conditions. The effects of MRI conditions (or just the supine position) on speech production has been examined in previous studies [158, 160, 161]. Some of these detected small differences in articulator positions, but no significant differences in the speech produced. Tenor singing [159] in supine positions has also been investigated and, again, found small differences in articulators, but not in the sound produced.

This research has shown that for the national/international principal singers in this study, the vocal tract resonances in singing produced under "simulated MRI conditions" (in an anechoic chamber) were not significantly different from singing produced under "normal" conditions.

The differences in the long-term average spectra generated for speech were almost entirely accounted for by a fall in amplitude of 5-10 dB for "simulated MRI" conditions compared to "normal" conditions. For singing, the LTAS showed that subjects produced more energy in the 1-4 kHz range than for speech, and the amplitude differences between the two conditions investigated were smaller.

This implies that singers are more consistent across conditions for singing than for speech, perhaps due to the effects of training. However, although the "simulated MRI" conditions replicated both the noise and the supine position of an MRI scan, the experience would not have been exactly the same for the subjects, and this would merit further investigation, with more closely replicated conditions (such as matching the temperature and reverberation in the MRI machine room, as discussed in section 5.4).

The differences between the conditions investigated were not completely consistent between singers, which was reflected in their experience of being in an MRI machine. Some found it highly stressful, but others not, which is normal compared to the general population. This highlights the importance of validating the effects of MRI for individual singers and it is hoped that these results, in addition to validating the reliability of the current study on resonance tuning production, may assist in the interpretation of data from previous studies and inform future research on vocal tract resonances utilising MRI techniques. It would be beneficial for future studies to include a verification procedure as standard and to possibly provide training to familiarise singers with the MRI environment.

7.2.4 Resonance Production

The main contribution of this thesis is the investigation into how the different articulators affect the vocal tract resonances. This was investigated in Chapter 4 using MRI to obtain area functions of the vocal tract and measurements of the vocal tract articulators during singing.

Inspection of a plot of the area functions showed that for the /u/ and /i/ vowels there were clear changes with fundamental frequency, which were not observed for the /ɑ/ vowel. Changes observed with frequency for the /u/ and /i/ vowels included shortening the vocal tract, decreasing the pharynx area, and increasing the mouth opening.

The regression models generated support some of the existing theories about the production of vocal tract resonances, such as the importance of *jaw position* and *vocal tract length*, which appeared in the significant regression models, but not as often as perhaps might be expected [63].

This suggests that the production of resonance tuning is much more complex than previously thought [63] and varies depending on the vowel and fundamental frequency sung, and according to the individual singer's preferred methods. Previous works investigating the production of resonance tuning have generally not focussed on the *difference between* vowels [99]. This work has shown the importance of this complexity, by illustrating differences in resonance, area functions and regression models between vowels.

When the feature selection procedure failed to generate a significant model, the process was repeated, but split by singer, and the results of this analysis suggested that singers were altering their resonances by different articulatory methods. These differed between the different vowels investigated, though it should be noted that these data sets were not large enough to produce statistically significant results. Therefore, although the resonance tuning methods used by different singers were very similar, the alterations to the vocal tract that produced these effects were not necessarily the same for all singers, suggesting that singers achieve similar resonance tuning strategies, but through different articulatory methods.

7.3 Hypothesis

As originally stated in section 1.1, the hypothesis informing this thesis is as follows:
Noise excitation to measure vocal tract resonances, and magnetic resonance imaging to observe the articulators can be used to understand the production of resonance tuning methods employed by professional soprano singers, which vary across different vowels and pitches as well as between singers.

Broad band noise excitation was employed in both the preliminary experiment involving girl choristers and the main experiment involving professional soprano opera singers (Chapters 4 and 6) to successfully observe resonance tuning behaviour in these two groups. MRI was used (Chapter 6) to obtain area functions of the vocal tract and measurements of the articulators. The relationships between these measurements was investigated using feature selection and regression analysis to generate statistically significant models for some of the vowels and resonances investigated (4 of 9). This seemed to suggest that the most important factors affecting resonances are the fundamental frequency, jaw opening and protrusion, uvula elevation, oropharynx breadth and vocal tract length.

Substantial variation was observed in both resonance tuning and articulators between different vowels and singers, with some trends in articulator measurements observed across pitches. The results suggest that both the production and perception of resonance tuning are much more complex than previously thought. Although the resonance tuning results and area functions suggest that singers are making similar changes with fundamental frequency, they may be producing these effects in different ways.

The results of the experiments carried out as part of this thesis clearly address the points of the stated hypothesis, while also identifying and leaving scope for further work on the effects of different articulators on the vocal tract resonances. Although the results of this work do not definitively establish how resonance tuning is produced, they provide another step along the road to understanding.

7.3.1 Further work and Impact

Despite the significant contributions to the field identified in section 7.2, there are still many questions left unanswered and there is considerable scope for further work.

Perception

Although this work has begun to consider the perception of resonance tuning within the context of its production, it would be beneficial to investigate the complex relationships between different perceptual attributes utilising *recorded* singing as well as synthetic sounds, as there is a limit to the naturalness of synthesised sounds.

The experiment in this thesis (Chapter 3) involved samples representing single vowel sounds. However, in reality vowels are almost always heard in words or phrases rather than in isolation, so future developments of this work also need to consider the importance of *context* on perception, for instance within a word or musical phrase.

It would also be useful to consider how the physical relationships between vocal tract articulators are linked to the perception of different strategies of resonance tuning. For example, does a resonance tuning strategy sound less natural to a listener if it is physically impossible for a singer to produce?

Use and Development of Resonance Tuning

In order to more fully understand the motivation for employing resonance tuning techniques it would be necessary to investigate the use of resonance tuning in younger, less experienced singers. These include choristers just starting to sing (to compare to the older choristers in this thesis) and music college students (to compare to the adult opera singers). Expanding the study to include both girl and boy choristers would also allow researchers to explore whether boys tune their first two resonances in similar ways to girls and adult female singers, and thus establish whether first and second resonance tuning is exclusively a female behaviour, or is a product of singing at high fundamental frequencies. Alternatively, the study could be expanded to include a control group with no, or very little training. This could then be compared to the Garnier study [15], which included singers of three different levels of experience.

In this work, it was found that the singing produced by the professional opera singers was not significantly affected by “simulated MRI” conditions. It is not known whether this consistency in unusual conditions is a feature of their extensive training or something common to all singers, so further study on the effects of MRI conditions involving non-expert subjects and subjects with less training would shed light on this problem. This would help to inform the development of MRI protocols appropriate for singers of differing levels of experience.

Finally, this work has focussed on the use and production of resonance tuning in choristers (classical church music) and professional opera singers, but it would be interesting to investigate how resonance tuning is used by singers of other music genres and other groups of singers, for example altos. Very little work has been done on “extreme” methods of resonance tuning, such as the whistle register and overtone singing, so a better understanding of these would be valuable to aid with teaching these methods.

Production of Resonances

Although the analysis carried out in this thesis gives some indication as to which vocal tract articulators affect the vocal tract resonances, it did not provide a full explanation of how resonances are produced. Inspection of the area functions generated suggested that singers made similar changes to area functions, but it is possible that these are achieved by different articulatory methods.

A much larger dataset would be required to fully investigate this with any statistical robustness, including many more singers and multiple repeats. This is discussed in section 6.4.1. In addition to this, the measurements of vocal tract resonances would ideally be obtained simultaneously with MRI image collection, removing the need for multiple experimental stages and validation procedures. Current MRI technology does not allow for this. However, it is hoped that advances in MRI technology will make this possible in the future. Advances in technology that reduce or remove some of the other limitations of MRI would also be welcome, for example the significant noise levels and the subject being required to lie supine. Upright MRI is available currently, however this is not yet widespread. Since this work has shown that singing produced under MRI conditions is not significantly different from singing produced in normal conditions, supine MRI is sufficient for studying singing production.

To fully understand the production of resonance tuning in normal singing (as opposed to isolated sustained vowels), it would be useful to study resonance tuning in more natural situations, such as singing in a musical context. A more holistic approach to analysis might also be beneficial, considering the vocal tract as a whole. Both this study and previous studies have considered the general changes to area functions for this purpose [6]. However, a reliable mathematical technique to describe overall changes to vocal tract shape has not yet been established. Morphoacoustic methods have previously been applied to the pinna (outer ear) to understand how small changes in the shape of the pinna affect the head-related transfer function [188]. It is possible, therefore, that applying similar methods to the vocal tract might be helpful in understanding how changes in its shape affect the resonances produced.

Implications

The final aspect of this work to be considered is the impact of these findings. Those most directly affected by this work will likely be singers and singing teachers. The information obtained about resonance tuning in girl choristers may allow practical guidelines to be developed to assist teachers specialising in this age group, who often lack a thorough understanding of vocal acoustics. If girl choristers are developing similar techniques to adult singers, this will influence singing teaching. Teachers may want to pay particular attention to vowel timbre, to ensure that students produce a resonant sound without compromising on identifiability of vowels.

The observations obtained of resonance tuning behaviour of highly professional opera singers will provide a little insight into what makes these singers so exceptional. As well as endeavouring to employ the same resonance tuning techniques, opera singers in training should take note of the adaptive approach taken to resonance tuning techniques.

Finally, the information obtained about the influence of area functions and specific articulators on the vocal tract resonances will give an insight into the importance of different parts of the vocal tract structure. This work and further work following similar protocols will hopefully be able inform the production of synthesised speech and efforts to alter the shape of the vocal tract, for example in corrective surgery involving the jaw, palate or teeth.

7.4 Closing Remarks

This thesis has provided valuable insight into the female singing voice, and the type and extent of resonance tuning used by Western Classical soprano singers. The use of MRI has allowed the effects of different articulators to be investigated, and represents a significant step forward in singing voice research, which is truly an intersection of art and science.

It is hoped that this research not only informs singing students and teachers of the available and prevalent methods of resonance tuning, but also highlights the importance of a singer finding their own particular voice and tailoring their approach to their individual needs to develop as a singer. To researchers interested in the singing voice, it is hoped that this work highlights the need to appreciate both the complex nature of singing and the individuality of singers.

Glossary

R_n - Resonances, resonant frequencies of the vocal tract

F_n - Formants, peaks in the acoustic spectrum of the voice

f_0 - Fundamental frequency of oscillation

MRI - Magnetic Resonance Imaging

SFC - Singer's Formant Cluster

Soprano - High female voice classification. Range approx C4 - C6.

Alto - Low female voice classification. Range approx F3 - D5.

Tenor - High male voice classification. Range approx C3 - G4.

Bass - Low male voice classification. Range approx E2 - E4.

Chorister - Young singer that performs regularly in Churches/Cathedrals

Appendices

The following materials are included as Appendices:

(A) Resonance tuning perception supplementary materials (Chapter 3)

(B) Choristers supplementary materials (Section 3.2.3)

(C) Sopranos supplementary materials (Section 4.2.2)

(D) Standard text and song (Section 4.2.2)

(E) Sopranos plots of resonances for all singers (Section 4.3.1 and 5.2.1)

(F) Sopranos plots of resonance tuning (Section 5.2.1)

(G) Sopranos plots of area functions (Section 5.2.3)

(H) Feature selection plots for all singers (Section 5.3.3)

(I) LF model details (Section 6.2.1)

The accompanying CD includes the following files:

(J) ETHICAL APPROVAL

- Choristers Ethical approval (Section 3.2.3)
- YNiC Ethical approval (Sopranos) (Section 4.2.2)

(K) SOPRANOS QUESTIONNAIRES

- Sopranos questionnaire answers (Section 4.3.4)

(L) MATLAB CODE FOR MRI EXPERIMENT

- Mic inverse filtering code (Section 4.2.3)
- Plotting resonances and K-S test code (Section 4.3.1)
- Area Function code (Section 5.2.3)
- Feature selection and Regression analysis code (Section 5.3.1 and 5.3.3) (includes 2D MRI measurements)
- Harmonic detection and Spearman correlation code (Section 4.3.3)

(M) DATA FOR MRI EXPERIMENT

- Audio samples (Section 4.3.3)
- Vocal tract segmentations (Section 5.2.3)

(N) PERCEPTION

- Vowel synthesis code (Section 6.2.1)
- Perceptual test results and Analysis (Section 6.2.4)
- Anova code (Section 6.3.4)

A Resonance tuning perception supplementary materials

List of IPA symbols

The following Table provides a list of IPA symbols for vowels sounds, and examples of words containing these vowels [189].

IPA symbol	Example sound
/ɪ/	The “i” in “kit”
/e/	The “e” in “bet”
/ae/	The “a” in “cat”
/ɒ/	The “o” in “lot”
/ʌ/	The “u” in “strut”
/ʊ/	The “oo” in “Foot”
/i/	The “ee” in “fleece”
/u/	The “oo” in “goose”
/ɑ/	The “a” in “father”
/ɔ/	The “ough” in “thought”
/ɜ/	The “ur” in “nurse”
/ə/	The “uh” in ‘afraid”

Table 1: List of IPA symbols for vowels, and example words.

Perceptual test examples

Figures 1, 2 and 3 below show the test GUI used to present comparisons to participants, using the qualtrics survey software:

Which do you prefer?

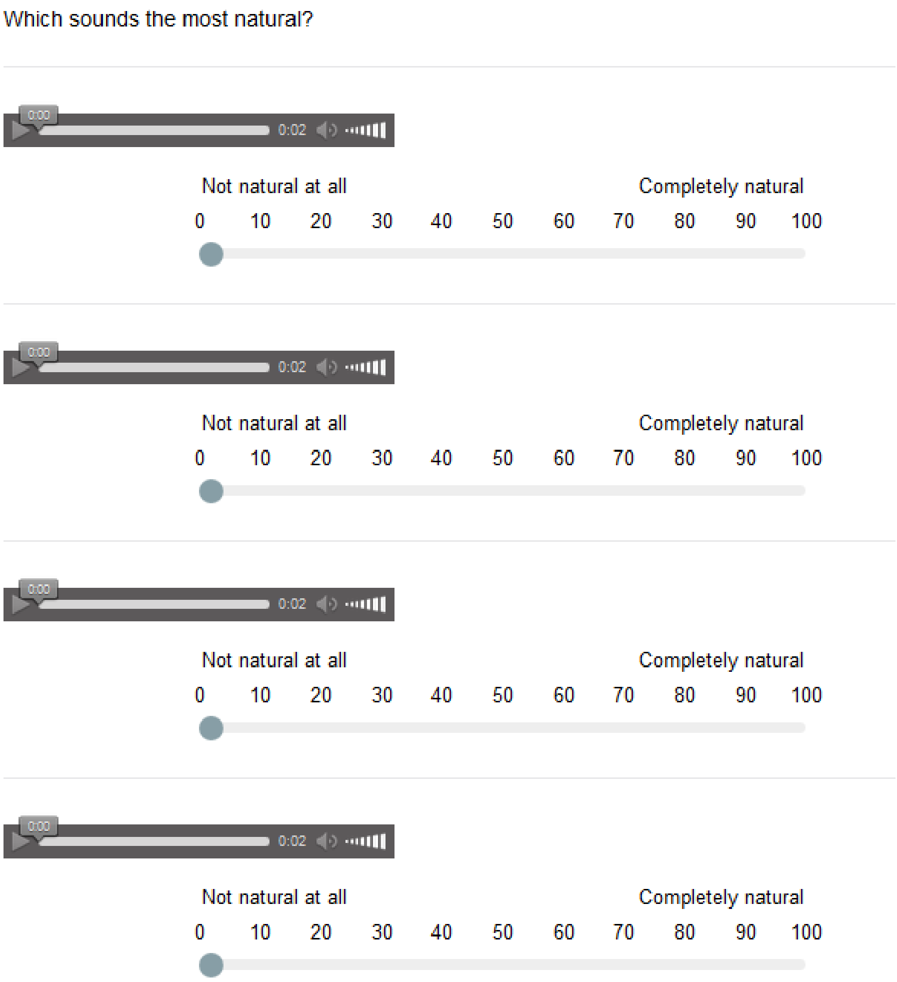
Don't prefer at all Prefer completely
0 10 20 30 40 50 60 70 80 90 100

Don't prefer at all Prefer completely
0 10 20 30 40 50 60 70 80 90 100

Don't prefer at all Prefer completely
0 10 20 30 40 50 60 70 80 90 100

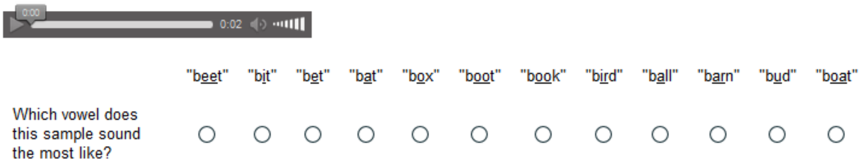
Don't prefer at all Prefer completely
0 10 20 30 40 50 60 70 80 90 100

Figure 1: An example from the set of questions on *preference*.



Which sounds the most natural?

Figure 2: An example from the set of questions on *naturalness*.



Which vowel does this sample sound the most like?

Figure 3: An example from the set of questions on *vowel identification*.

B Choristers supplementary materials

This appendix includes the supplementary materials from the first experiment investigating resonance tuning in girl chorister. This consists of the information sheet, questionnaire and consent forms for both the subjects and their parents.

Information sheet - A Pilot Study: Investigating formant tuning in girl choristers through wide-band vocal tract excitation.

Procedure of the Experiment

In this experiment you will be taken into the anechoic chamber and asked to wear a laryngograph throughout the experiment – this will involve electrodes on an elastic strap being fastened on to your neck by the investigator. This will feel slightly unnatural but will not be painful and should not impede normal breathing, speaking and singing. You will also be fitted with a head microphone, this is small and light and you will still be able to speak and sing normally with it on.

You will then be asked to read a story out loud (seen beforehand) once "normally", into a microphone, and once into a device designed to measure the resonances of the vocal tract. This is a tube-like device and will be mounted on a stand. You will need to stand close to it so that your top lip rests on the end of the tube. Again this will feel slightly unnatural but will not be painful and you will be able to speak and sing normally.

The next part of the experiment will be singing; you will be asked to sing an ascending chromatic scale from the bottom of your range to the top, holding each note for about three seconds. This will be repeated on three vowel sounds, "ah", "ee", and "oo".

The scales will then be repeated singing into the tube-like device.

Please sing in your best performance voice, and try to keep the shape of your mouth and throat constant for the whole note, and to stay at the same pitch and loudness. The note will be given on a piano before each note.

The last task will be to sing the first verse of "Once in Royal David's City". This will be done once normally (into the microphone), and then once into the tube-like device.

If you have any concerns at all (at any point in the procedure), please raise them with one of the investigators (Rebecca Vos or Helena Daffern). If you want to, you can leave at any time during the experiment, without having to give a reason.

Data Protection

Any data collected will be identified only by a number for the purposes of data collection, and the list linking names and numbers will be securely stored in a separate location.

The data will be stored securely after the experiment, and not shared with any person outside the Audio lab without the permission of the participant.

Questionnaire

Please answer the questions below as accurately as possible. If you do not feel that any of the options accurately represent your answer, write a comment in the space below that question.

1. Please select your gender: Male ☐ Female ☐

2. Please enter your age: ☐

3. How many years have you been a minster chorister? ☐

5. Do you have a hearing impairment? (for example loss of hearing, tinnitus etc.) Yes ☐ No ☐

6. Are you currently having singing lessons? Yes ☐ No ☐

If yes, for how long have you been having them? (in years and months) ☐

7. Have you passed any music exams? (if you have several in the same instrument, only write down the highest grade)

____
____
____
____
____

8. Have you ever performed (singing) in public?

☐ no

☐ Only as part of a choir

☐ A few times (less than 10)

☐ Many times (10 or more)

Any other comments?

____
____
____
____
____

Consent form for Subject

Investigator's Names: Miss Rebecca Vos

Title of Study: A Pilot Study: Investigating formant tuning in girl choristers through wide-band vocal tract excitation.

Brief Description of Study:

You will be asked to answer some questions about yourself, and how much and how often you sing.

You will be asked to read an information sheet with your parent explaining what will happen in the experiment.

You will have electrodes strapped on to your neck. Don't worry, this doesn't hurt at all and you will be able to breathe, talk, and sing normally.

You will be asked to do 4 tasks; (1) read a short story out loud, (2) say three vowels, (3) sing scales on three vowel sounds, and (4) sing the first verse of "Once in royal David's City".

These will be done once normally (into a microphone), and once into a tube-like device, which you will be shown how to use.

Please circle "yes" or each point if you understand and are happy to continue:

My data will be stored anonymously (without my name on), and not shared with anyone outside the Audio lab without my permission. **YES/NO**

Even though my data won't have my name on it, and the investigators won't ever tell anyone who it is, somebody who knows my voice well may be able to tell that it is me. If this happens they will never be told if they are right or wrong. **YES/NO**

Even though I agree to take part in the experiment, I am allowed to leave at any time I want to (I don't have to give a reason). **YES/NO**

If I want, I can ask for recordings to be deleted at any point during or after the experiment. **YES/NO**

My anonymous data (without my name on it) can be used for research and teaching purposes, and published at conferences, in journals or online, either as it is, or in other formats. **YES/NO**

By signing this form, I confirm that I have filled out the Participant Questionnaire, read the Information sheet carefully, and am happy to continue with the experiment. **YES/NO**

Print name: ________________________________

Date: ________________________________

Sign name:

Consent form for Parents

Investigator's Names: Miss Rebecca Vos

Title of Study: A Pilot Study: Investigating formant tuning in girl choristers through wide-band vocal tract excitation.

Brief Description of Study:

You and your daughter will be asked to answer a short questionnaire about her level of singing training and experience of singing.

You and your daughter will be asked to read an information sheet explaining the nature of the experiment and what she will be asked to do.

Your daughter will be recorded with a microphone and wearing a laryngograph (consisting of electrodes on an elastic neckband). She will be asked to do 4 tasks; (1) read a short story out loud, (2) say three vowels, (3) sing scales on three vowel sounds, and (4) sing the first verse of "Once in royal David's City".

She will be asked to repeat each of the tasks whilst singing/speaking into a tube-like device to measure her vocal tract resonances.

Please circle either yes or no:

I understand that my daughter's data will be securely stored in an anonymous form, and not shared with any others outside the Audio lab without my permission. **YES/NO**

Due to the nature of this data, it is possible that even though it is anonymised, somebody who knows my daughter's voice may be able to identify her. This will never be confirmed or denied by the investigators. **YES/NO**

I am happy for my daughter's anonymised data to be used for research and teaching purposes within the Audio lab, and published at conferences, in journals or online, either as it is, or in secondary formats. **YES/NO**

By signing this form, I confirm that I have filled out the Participant Questionnaire and read the Information sheet carefully with my daughter, and we are happy to continue with the experiment. **YES/NO**

Even though I consent for my daughter to take part in the experiment, I understand that I may withdraw my daughter at any time, without having to give a reason. If I wish to, I may ask for my daughter's data to be destroyed at any point during or after the experiment. **YES/NO**

Print name: Date:

________________________________ ________________________________

Sign name:

____

C Sopranos supplementary materials

This appendix includes the supplementary materials from the main experiment investigating the production of resonance tuning in adult soprano singers. This consists of the information sheet, questionnaire and consent forms, as well as the safety questionnaire and consent form required for magnetic resonance imaging at The York Neuro-Imaging Centre (YNiC).

MRI Capture of the Female Vocal Tract - Information for Adult Participants

You are being invited to take part in a research study. Before you decide whether to take part 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. Talk to others and particularly members of the project team about the study if you wish.

The first part of this document explains the purpose of this study and what will happen to you if you take part. The second part gives you more detailed information about the conduct of the study.

Please ask us if there is anything that is not clear, or if you would like more information regarding the study. Take time to decide whether or not you wish to take part.

Part 1

Project Background:

Vowel formants are broad resonances that occur in the spectrum of speech and singing, amplifying nearby harmonics. They are responsible for the perception of vowels. For female singers, singing vowel sounds at high frequencies can mean that the fundamental frequency exceeds the normal first vowel formant (in speech), “wasting” the amplification of this resonance.

It is now generally accepted that, classical singers even at a non-professional level, will “tune” their vocal tracts, to raise their first vowel formant, so that it occurs at close to the fundamental frequency.

Recent research in the audio lab has focussed on the perception of formant tuning, using synthetic singing samples with different formant tuning methods applied. The results suggested that formant tuning could be used over a wider range than has formerly been observed in real singers and this study aims to suggest possible reasons for this.

This study also aims to investigate the physiological and acoustic factors at play when a female opera singer sings in the different parts of her range in a full, resonant performance voice.

Acoustic data will be captured during the MRI scans, using optic microphones, and subjects will be asked to hold particular vocal tract shapes (different vowels over 5 pitches) for the duration of a scan.

More details of the procedure will be given in part 2.

What will happen to me if I take part?

If you choose to take part you will be asked to attend YNiC (the York Neuro-Imaging Centre) on the Science Park at the University of York, with members of the project team. You will also be asked to complete some simple audio recordings in the Anechoic Chamber (a sound isolated room with absorbing walls that produce no echoes) of the Audio Lab (also on the Science Park) before the scanning procedure.

The audio recording procedure takes approximately 80 minutes, and the scanning procedure takes no more than 90 minutes. In total we will require approximately 3 ½ hours of your time.

What will happen during each session?

Part 1:
In the first part of the procedure we will ask you to wear a microphone headset and a laryngograph neckband, and stand in the anechoic chamber in your normal singing stance. You will then be asked to read a passage of text in your normal speaking voice. The laryngograph neckband is fitted with two electrodes that allow us to monitor your vocal fold movement. This should not impede speech or singing or feel too unusual.

You will then be asked to sing an ascending chromatic scale from the bottom of your range to the top in your full, “resonant”, operatic voice without excessive vibrato, holding each note for 3 seconds. A piano note will be played for reference before each note. This will be done on three vowels, first “ah”, then “oo”, then “ee”. You will be asked to sing into a device to measure the resonances of your vocal tract (a kind of tube incorporating a speaker and a microphone which you will be shown how to use).

The last task for this part will be to sing a short song of your own choosing, at least a verse or 30 seconds long.

Part 2:
In the second part of the procedure we will ask you to lie on a foam-covered board on the floor in the anechoic chamber, still wearing the microphone headset, and laryngograph electrodes. We will also ask you to wear earplugs and headphones to simulate the MRI conditions.

You will be asked to repeat the speaking and singing tasks from part 1, and while you produce the sounds, a recording of an MRI scanner will be played over the headphones.

You will be then asked to sing 4 pitches for each vowel, again in your full, “resonant”, operatic voice without excessive vibrato, holding each one for 16 seconds. A piano note will be played for reference before each note.

Part 3
For part 3 we will relocate to YNiC where you will be briefed by a member of staff and your final consent to be scanned will be sought. An information sheet is provided with this document that gives more information about what it is like to be scanned.

A member of YNiC staff will position you in the scanner. You will be asked to wear a set of foam ear-plugs and headphones during scanning (both to protect you from the noise of the scanner and to allow us to communicate with you). During the entire scanning process you will be in continuous contact with the operator via the intercom.

A number of configuration scans will first be taken. You are not required to do anything for this time. When these are complete you will be notified over the intercom.

You will first be asked to hold a neutral vocal tract shape, and breathe as normally as possible without moving for 16s.

You will be then asked to sing 4 pitches for each vowel (as in parts 1 and 2), again in your full, “resonant”, operatic voice without excessive vibrato, holding each one for 16 seconds. Before each scan a reference piano note will be played over the intercom, and as with the audio recording this represents the pitch at which you should sing. You will also be told which vowel to sing.

What are the benefits of taking part in this research?

This study is focussed on exploring the ways in which formant tuning is executed by elite opera singers and the difference in the use of formant tuning over a range of pitches. The data gathered in this study will further understanding of the range of pitches over which formant tuning methods are used, and the physiological adjustments of the vocal tract necessary to achieve them.

What are the other possible disadvantages and risks of taking part?

The research project has been planned, and will be conducted in a manner that minimises the risk of harm to its participants. While undergoing scanning you may experience claustrophobia and/or feelings of isolation. It is also possible you will experience dis-equilibrium and/or mild nausea on entering or leaving the strong magnetic field the scanner produces. If you become uncomfortable you may leave the scanner at any time.

** If you are unhappy with any aspect of the process then you may withdraw from the study at any time.

It is of vital importance that participants do not carry any ferromagnetic materials with them when being scanned (e.g. keys, jewellery, coins, belt buckles). You will be asked to leave all similar items outside the scanning room. Please be aware that it is not possible to scan some groups of people (e.g. pregnant women, those with metal implants including dental braces, pacemakers, cochlear or brainstem implants, or those who have any other surgical implants containing metal).

Very high sound levels are produced during scanning. For your protection you will be advised to wear ear plugs and headphones.

What if there is a problem?

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information on this is given in Part 2.

Will my taking part in the study be kept confidential?

Yes. All the information about your participation in this study will be kept confidential. The details are included in Part 2.

Contact Details of the Project Team

Rebecca Vos (Project Principle Investigator) – rrv501@york.ac.uk
Dr Helena Daffern– helena.daffern@york.ac.uk
Prof David Howard – dmh@ohm.york.ac.uk

Part 2

What will happen if I don't want to carry on with the study?

You are free to withdraw from this study at any time, and without providing an explanation. Any data collected from you up to this point may still be used in the study unless you make a request to the contrary. In the latter case all data collected from you will be destroyed.

What if there is a problem?

There are two procedures for complaints. One addresses the case of your mistreatment by the project team, and the other addresses something serious happening during or following your participation in this study.

Complaints:

If you have a concern about any aspect of this study, you should ask to speak with the researchers who will do their best to answer your questions. Their contact details are available in Part 1. If you remain unhappy and wish to complain formally, you can do this through the complaints procedure of the University of York.

Harm:

The York Neuroimaging Centre takes pride and care in ensuring that no harm, or risk of harm, occurs to participants in research. In the event that something does go wrong and you are harmed during the research study and this is due to someone's negligence, then you may have grounds for a legal action for compensation against The University of York.

Will my taking part in this study be kept confidential?

Any information which you give us, and all of the measurements that we collect from you, will be confidential. No names will be used when the research is written up. We shall keep your data for 10 years and will then destroy it securely. We shall comply with the terms of the Data Protection Act 1988. We shall store the information and the measurements in anonymous computer files and in locked filing cabinets. We shall store names and addresses separately from other data. We shall use your data in this study and we may combine your data with data that we gather in future studies. Only three people in our research team will know the contact details of the participants. They are Miss Rebecca Vos, Dr Helena Daffern, and Professor David Howard. In addition, staff of the York Neuro-imaging Centre have privileged access to the computer systems and can link the names of participants with their data. Those people are under a professional obligation not to abuse this privilege. With the approval of the Research Ethics Committee of the York Neuroimaging Centre, other researchers may be allowed access to the data which you will provide for use in research and teaching. Those researchers will be allowed access to your data in anonymous form only.

We are not qualified to interpret brain images clinically. If we suspect that an image of your brain reveals a possible problem, we shall inform your GP (family doctor) who may then contact you and advise you. If you do not want us to do this, then you should not agree to take part in the study.

What will happen to the results of the research study?

Your anonymised data will be used for scientific research purposes, and published at conferences, in journals or online, either as raw data or in secondary formats.
You should be aware that although names will never be connected to data, and the investigators will never confirm or deny the identity of any participants, the nature of the data (sound recordings) means that it could be possible for an experienced listener to identify you.

Who is organising and funding the research?

The study is being organised by Miss Rebecca Vos, Dr Helena Daffern, and Professor David Howard. They work in the Department of Electronics at the University of York. Rebecca Vos is a PhD student, Helena Daffern is a Lecturer, and David Howard is the Head of the Audio Lab and

the department of Electronics.
The study is being funded by a grant from the AES Educational Foundation.

Who has reviewed the study?

This study was given a favourable ethical opinion by the Research Ethics Committee of the York Neuroimaging Centre.

Questionnaire

Before the experiment

1. Please provide your age in years _____

2. What is your voice type? _______________

3. What is your range? ______ to ______

4. Do you currently have any health issues that affect your singing?

____

5. What is your job?

____

6. For how many years have you been singing at your current level? _____

7. Where did you train, and for how long?

____

____

____

8. Do you have a method/technique for singing high notes? If so can you describe it?

____

____

____

9. Are you aware of making changes to the shape of your vocal tract when you sing high notes? If so, what are they? (eg soft palate, larynx, jaw position, tongue position)

____

____

____

10. Is this the same for different vowels? Are any of them more difficult to sing?

____

____

____

11. Do you know of any differences in the way you sing in different positions? Particularly standing up and lying down? If so what are they?

____

____

____

<u>After the experiment:</u>

12. Do you feel that you made changes to the shape of your vocal tract when you were singing high notes? If so, what? (eg soft palate, larynx, jaw position, tongue position)

____

____

____

____

13. Was this the same for different vowels? Were any of them more difficult to sing?

____

____

____

14. Was this the same for low and high notes?

____

____

____

15. Did you notice any differences in the way you sung in the different positions (standing up and lying down)? If so what were they?

____

____

____

16. How did you find singing whilst in the MRI machine?

____

____

____

17. Did you feel that you sung differently than you would in a “normal” standing position? If so how?

____

____

____

____

18. How did you find the whole experiment in general?

____

____

____

19. Any other comments?

____

____

____

____

____

Consent Form for Adult Participants

MRI Capture of the Female Vocal Tract

Participants should complete items 1 to 10 themselves

Please circle either
YES or NO

1. That I have read and understood the information sheet entitled 'MRI Capture of the Female Vocal Tract – Information for Adult Participants' **YES / NO**

2. That I have been given the opportunity to ask any questions I may have about my participation in the project and that these questions have been answered to my satisfaction. **YES / NO**

3. Who has explained the study to you? **YES / NO**

   Prof/Dr/Mr/Miss...

4. I understand that I am free to withdraw from this study: **YES / NO**

- At any time
- Without having to give a reason
- Without prejudice to my academic standing at the University of York

5. Do you agree to take part in the study? **YES / NO**

6. I understand that I can discuss the study with a researcher at any time, if I wish. **YES / NO**

7. I know that the research information which I will provide will be kept strictly confidential. When the results are published no individual person will be identified in any way without that person’s written agreement. **YES / NO**

8. I understand that although the data collected from me (audio and MRI) will be anonymised, due to its nature it may be possible for an experienced listener to identify me, but my identity will never be confirmed by the investigators. **YES / NO**

9. I understand that my anonymised data may be used for scientific research purposes, and published at conferences, in journals or online, either as raw data, or in secondary formats. **YES / NO**

10. If I have any questions or concerns about the research, I know I can contact Rebecca Vos at the Department of Electronics at rrv501@york.ac.uk

11. Participant:

Name (Block Letters): …………………………………………………………......

Signature …………………………………………………………………

Date …………………………………………………………………

12. Investigator:

I have explained the study to the above participant and he/she has indicated his/her willingness to take part.

Name (Block Letters): ……………………………………………………………......

Signature ……………………………………………………………………

Date ……………………………………………………………………

York NeuroImaging Centre
THE UNIVERSITY of York

Safety Questionnaire and Consent for Participants

Version 1 - 2014-01-01

Modality - MRI

Participant ID: R	**Surname:**	**Forename:**	**Date of Birth:**	**Weight:**	**Contact No:**
Address		**GP Address**			
Verified		**Verified**			

If your, or your GPs, address are different to those shown above please cross out the old address and write the new one next to it.
Please do not assume we will accept you for a research scan on the basis that you have been scanned elsewhere (regardless of whether it was for research or clinical reasons).
Please remove all eye make up prior to attending for your scan.

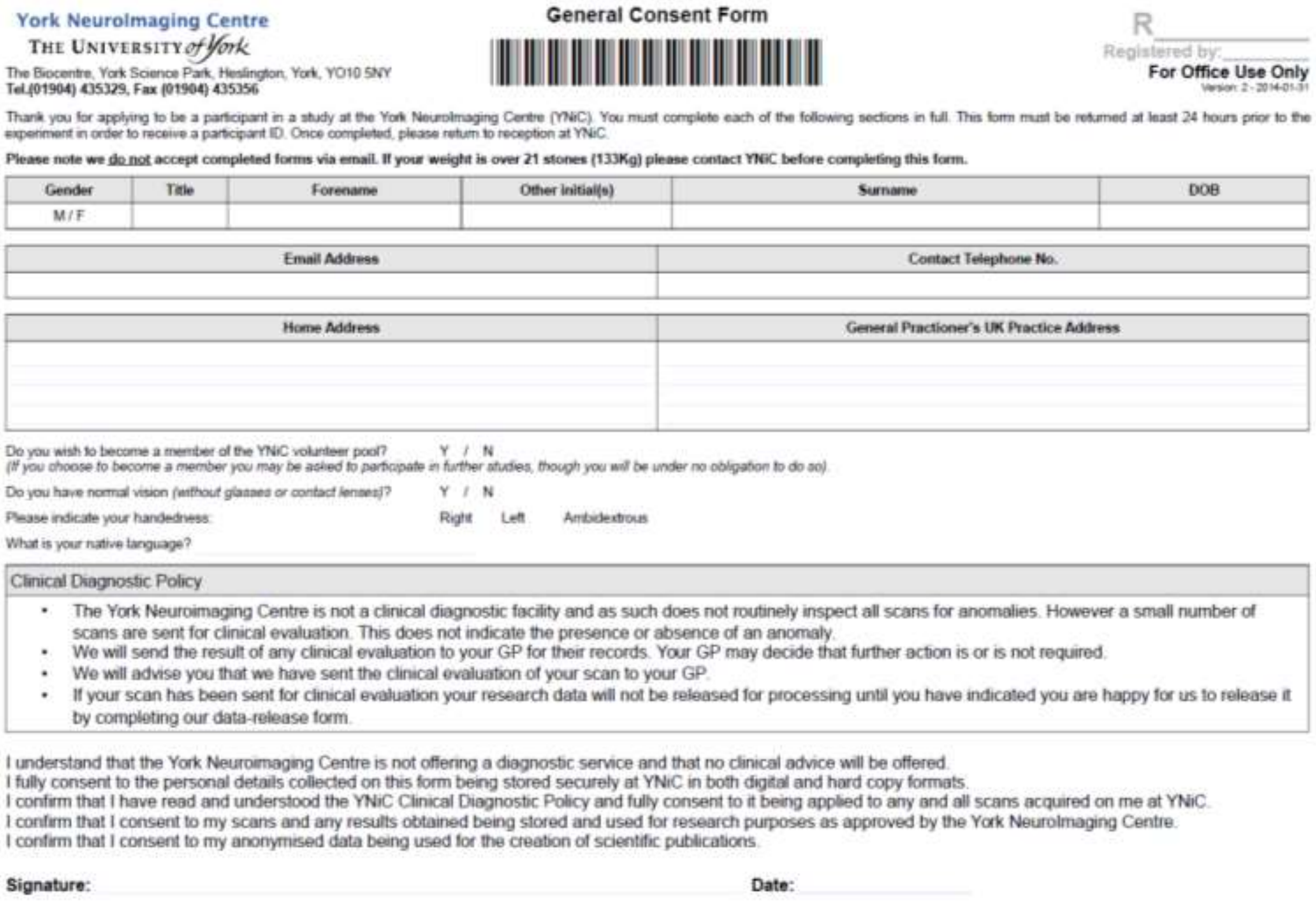
It is essential that all questions on this sheet are answered truthfully. This information is essential in order to ensure your safety and will be kept completely confidential.

Please answer the following questions accurately by ticking the appropriate box. If you answer YES to any of the safety questions please call YNIC on 01904 435329	**Participants Answers**	
Safety	**Yes**	**No**
Do you have a cardiac (heart) pacemaker or pacing wires?		
Have you ever had any operations on your heart, head or spine? *If yes please give details*		
Do you have or have ever had a spinal or other neuro stimulator		
Have you had any surgery which involved the use of medical implants? E.g. Hip or knee replacements, breast or penile implants or any procedure using metal stents e.g. coronary arteries. *If yes please give details*		
Do you have a programmable hydrocephalus shunt?		
Do you have a cochlear (ear) implant?		
Do you have a fixed dental brace?		
Have you any surgery in the last 3 months?		
Have you,at any time, had an injury to your eye involving metal fragments?		
If YES Did you see a doctor or get medical advice?		
If YES, did the doctor tell you that everything had been removed?		
Do you have any shrapnel in your body?		
Do you have any piercings?		
Do you have any tattoos?		
Do you have any medicinal patches? Including nicotine,hormone		
Are you a close relative of any of the investigators, i.e. Spouse, sibling, parent, grandparent, child or grandchild?		
Do you have epilepsy? Have you ever had a fit or seizure?		

Please answer the following questions accurately by ticking the appropriate box. If you answer YES to any of the safety questions please call YNIC on 01904 435329	**Participants Answers**	
Safety	**Yes**	**No**
Female Participants		
Do you have an Intra-Uterine Contraceptive Device? *if yes please give details (certain types are not safe at 3T)*		
Are you, or could you be pregnant?		
Consent		
I have fully understood and completed the safety section of this form.		
I have been fully informed and understand the nature of the procedures to be carried out		
I have been able to ask questions regarding the procedures		
I have had the emergency evacuation procedures explained to me		
I confirm that I give my full consent to MRI scans being performed on me		
I give my consent to my anonymised images from my scan being used for display		
I am aware that I may end the scan or procedure at any time by informing a member of staff		

Signatures

Project ID: -		Date
Participant		
Guardian (if under 18 years old)		
Approved Investigator		
Approved operator		



York NeuroImaging Centre
THE UNIVERSITY of York
The Biocentre, York Science Park, Heslington, York, YO10 5NY
Tel.(01904) 435329, Fax (01904) 435356

General Consent Form

R____________
Registered by:________
For Office Use Only
Version: 2 - 2014-01-31

Thank you for applying to be a participant in a study at the York NeuroImaging Centre (YNiC). You must complete each of the following sections in full. This form must be returned at least 24 hours prior to the experiment in order to receive a participant ID. Once completed, please return to reception at YNiC.

Please note we do not accept completed forms via email. If your weight is over 21 stones (133Kg) please contact YNiC before completing this form.

Gender	Title	Forename	Other Initial(s)	Surname	DOB
M / F					

Email Address	Contact Telephone No.

Home Address	General Practioner's UK Practice Address

Do you wish to become a member of the YNiC volunteer pool? Y / N
(If you choose to become a member you may be asked to participate in further studies, though you will be under no obligation to do so)

Do you have normal vision *(without glasses or contact lenses)*? Y / N

Please indicate your handedness: Right Left Ambidextrous

What is your native language?

Clinical Diagnostic Policy

- The York Neuroimaging Centre is not a clinical diagnostic facility and as such does not routinely inspect all scans for anomalies. However a small number of scans are sent for clinical evaluation. This does not indicate the presence or absence of an anomaly.
- We will send the result of any clinical evaluation to your GP for their records. Your GP may decide that further action is or is not required.
- We will advise you that we have sent the clinical evaluation of your scan to your GP.
- If your scan has been sent for clinical evaluation your research data will not be released for processing until you have indicated you are happy for us to release it by completing our data-release form.

I understand that the York Neuroimaging Centre is not offering a diagnostic service and that no clinical advice will be offered.
I fully consent to the personal details collected on this form being stored securely at YNiC in both digital and hard copy formats.
I confirm that I have read and understood the YNiC Clinical Diagnostic Policy and fully consent to it being applied to any and all scans acquired on me at YNiC.
I confirm that I consent to my scans and any results obtained being stored and used for research purposes as approved by the York NeuroImaging Centre.
I confirm that I consent to my anonymised data being used for the creation of scientific publications.

Signature: ________ **Date:** ________

D Standard text and song

This appendix includes the standard text, “Arthur the Rat”, and song “Once in Royal David’s city”.

Arthur the Rat

Once there was a young rat named Arthur, who could never make up his mind. Whenever his friends asked him if he would like to go out with them, he would only answer, "I don't know." He wouldn't say "yes" or "no" either. He would always shirk making a choice.

His aunt Helen said to him, "Now look here. No one is going to care for you if you carry on like this. You have no more mind than a blade of grass."

One rainy day, the rats heard a great noise in the loft. The pine rafters were all rotten, so that the barn was rather unsafe. At last the joists gave way and fell to the ground. The walls shook and all the rats' hair stood on end with fear and horror. "This won't do," said the captain. "I'll send out scouts to search for a new home."

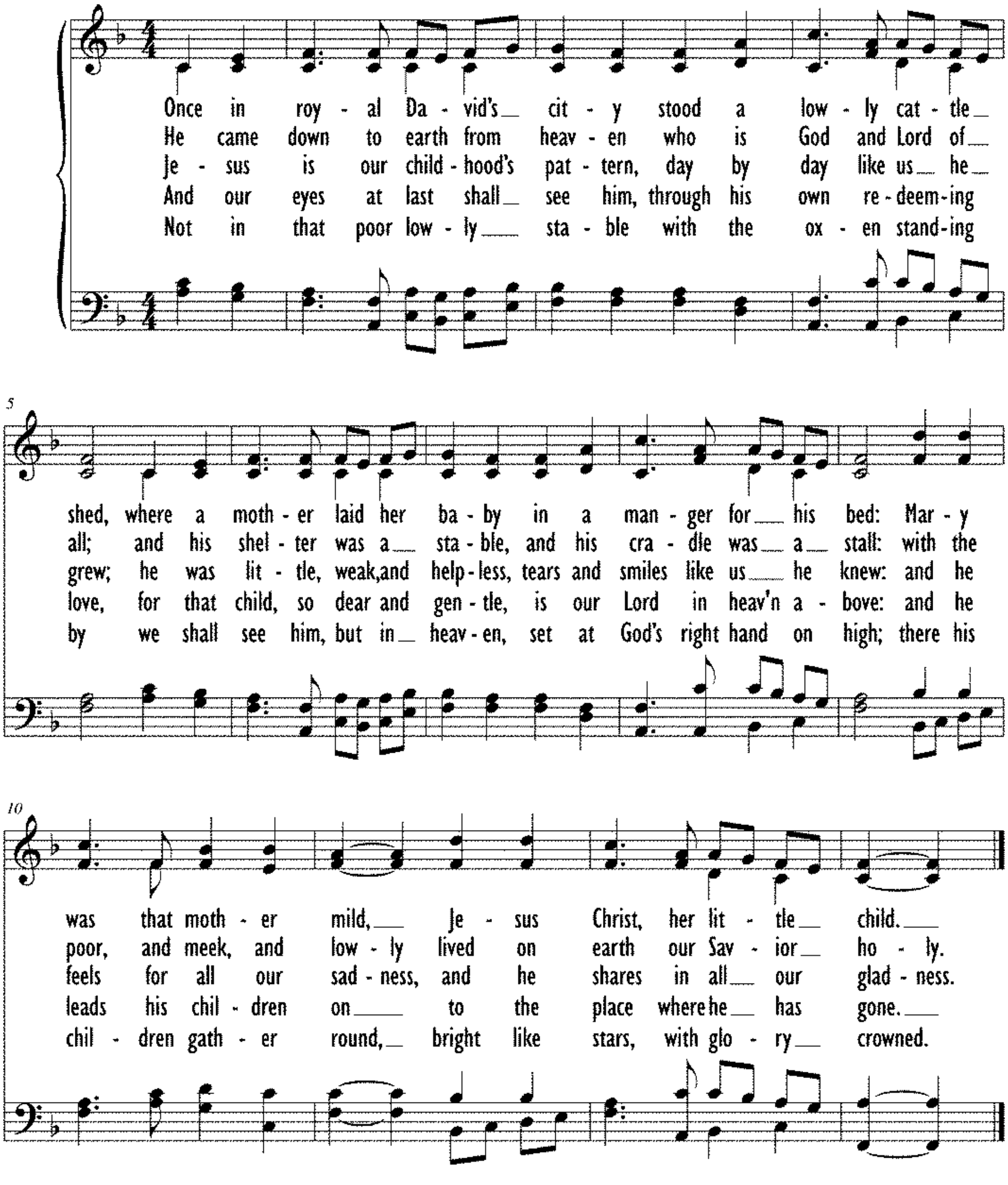
Within five hours the ten scouts came back and said, "We found a stone house where there is room and board for us all. There is a kindly horse named Nelly, a cow, a calf, and a garden with an elm tree." The rats crawled out of their little houses and stood on the floor in a long line. Just then the old one saw Arthur. "Stop," he ordered coarsely. "You are coming, of course?" "I'm not certain," said Arthur, undaunted. "The roof may not come down yet." "Well," said the angry old rat, "we can't wait for you to join us. Right about face. March!"

Arthur stood and watched them hurry away. "I think I'll go tomorrow," he calmly said to himself, but then again "I don't know; it's so nice and snug here."

That night there was a big crash. In the morning some men—with some boys and girls—rode up and looked at the barn. One of them moved a board and he saw a young rat, quite dead, half in and half out of his hole. Thus the shirker got his due.

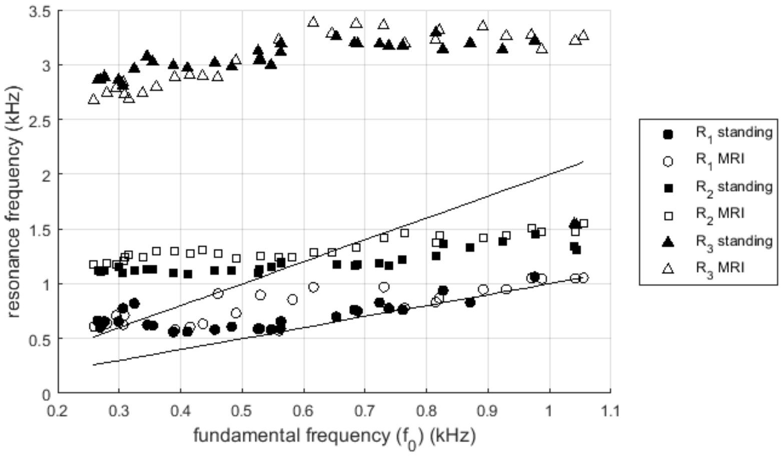
Once in Royal David's City

Cecil Frances Alexander

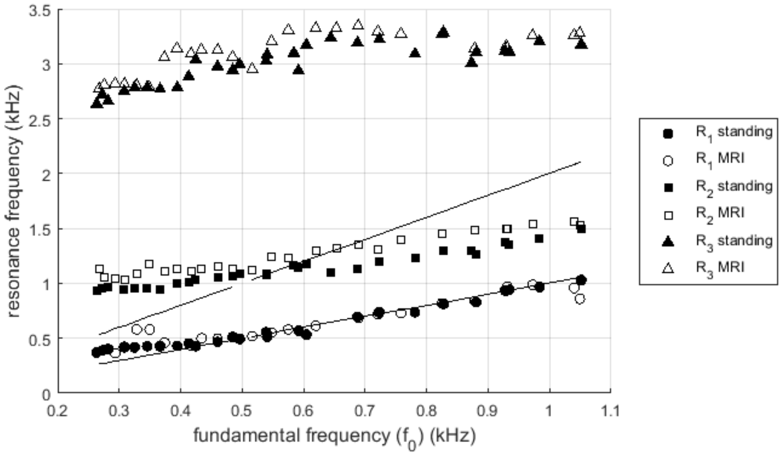
Henry John Gauntlett

1/17/2016

E Sopranos plots of resonances for all singers

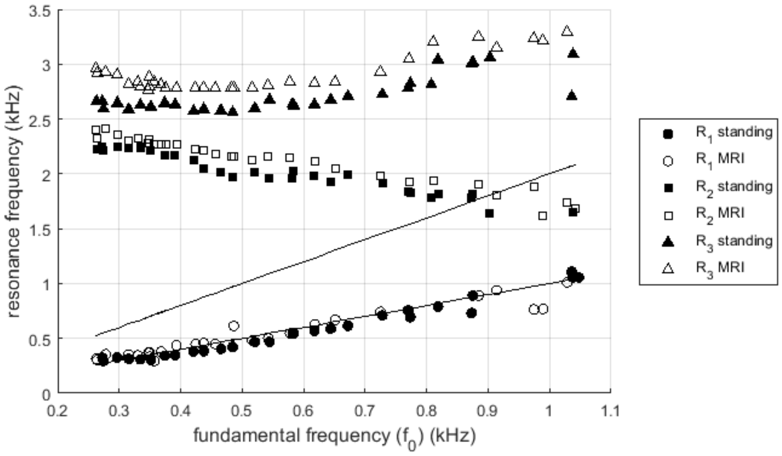
As discussed in section 5.3.1 and 6.2.1, this appendix includes plots of the first three resonances, for all singers, for all three vowels.



(a) /ɑ/ vowel

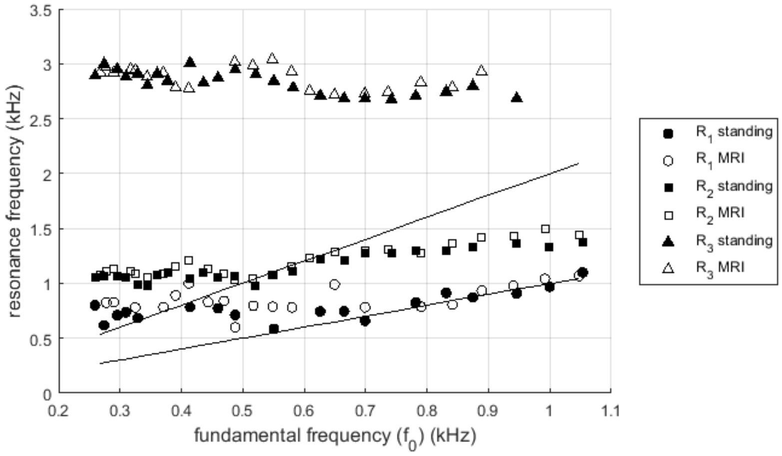


(b) /u/ vowel

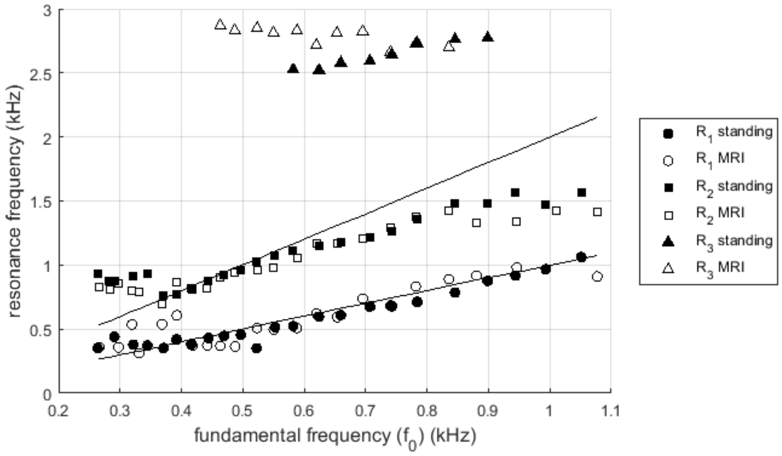


(c) /i/ vowel

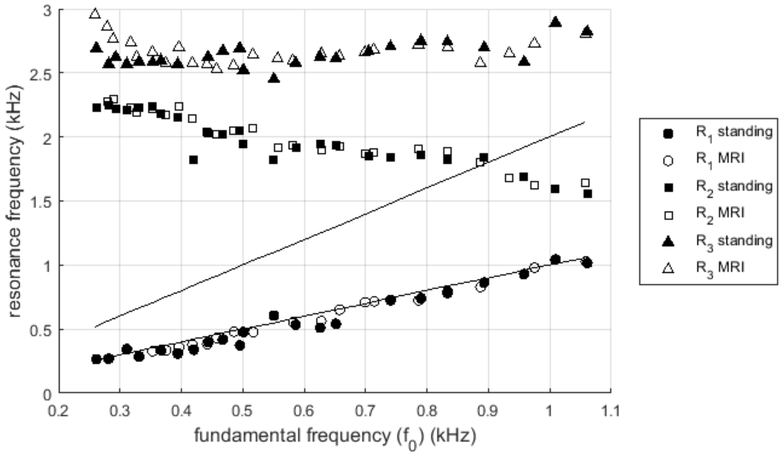
Figure 4: The first three resonances plotted against fundamental frequency, for subject 1, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom).



(a) /ɑ/ vowel

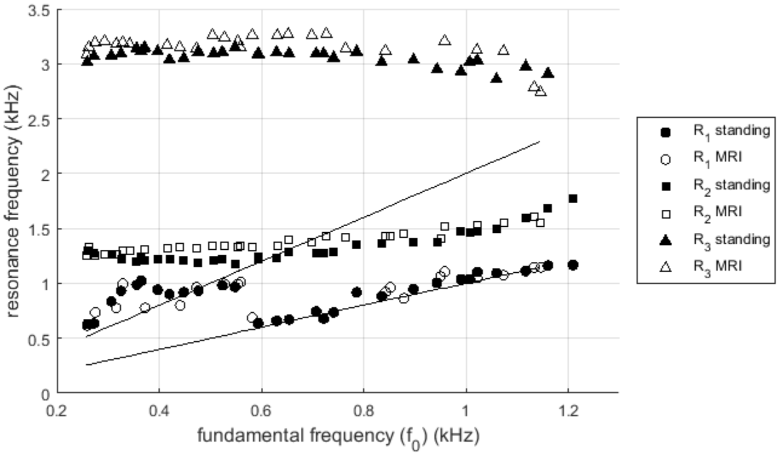


(b) /u/ vowel

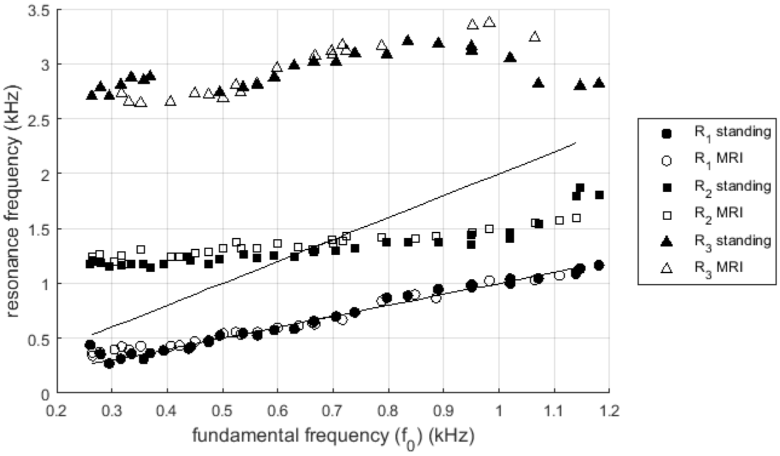


(c) /i/ vowel

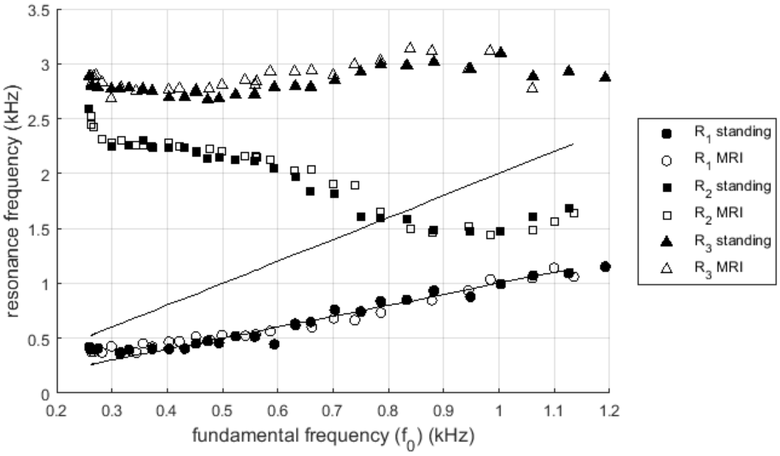
Figure 5: The first three resonances plotted against fundamental frequency, for subject 2, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom).



(a) /ɑ/ vowel

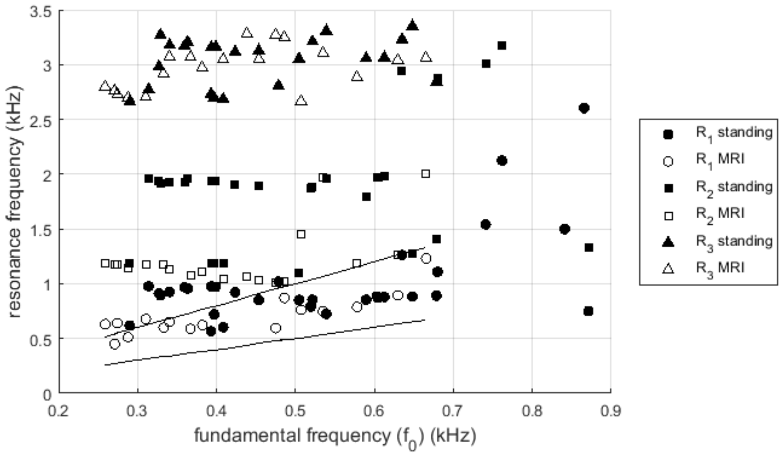


(b) /u/ vowel

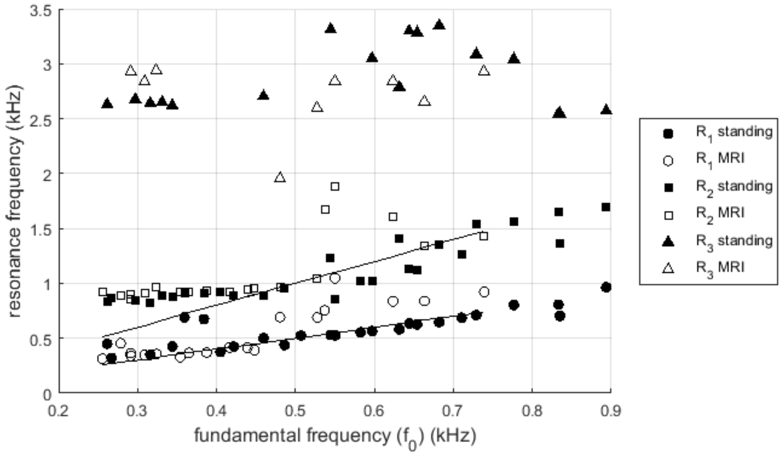


(c) /i/ vowel

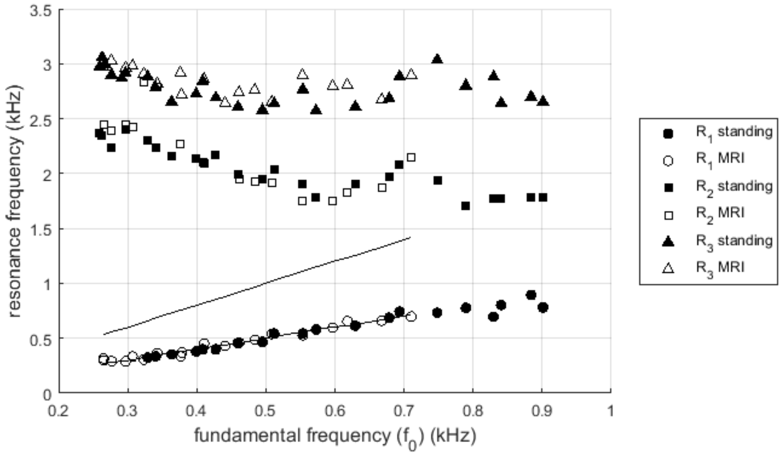
Figure 6: The first three resonances plotted against fundamental frequency, for subject 3, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom).



(a) /ɑ/ vowel

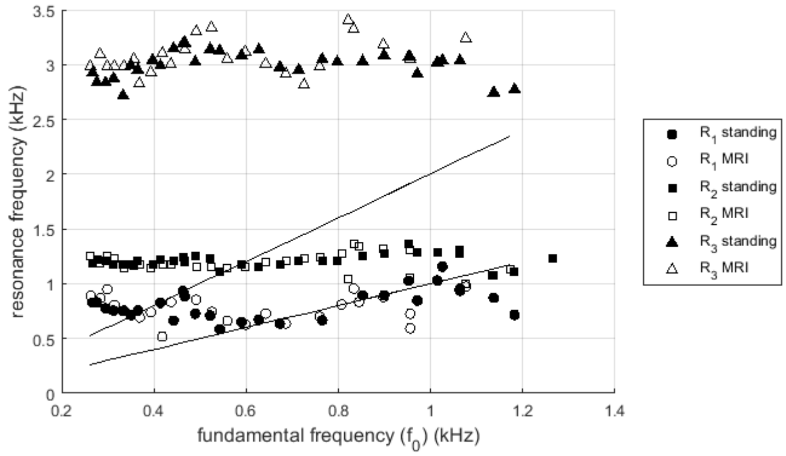


(b) /u/ vowel

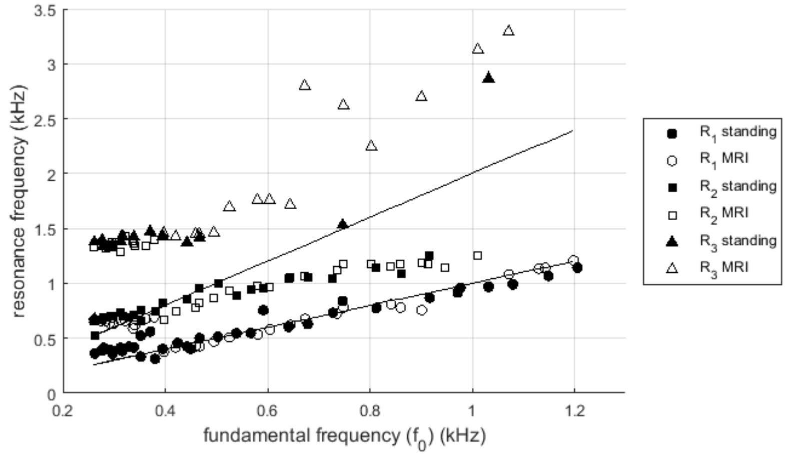


(c) /i/ vowel

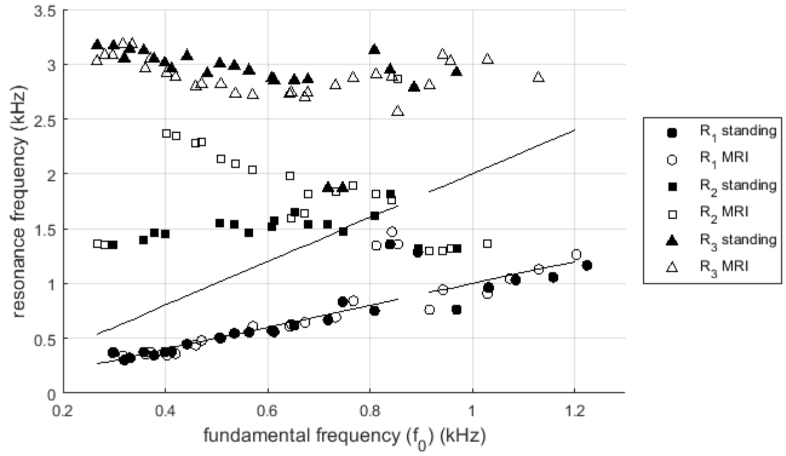
Figure 7: The first three resonances plotted against fundamental frequency, for subject 4, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom).



(a) /ɑ/ vowel

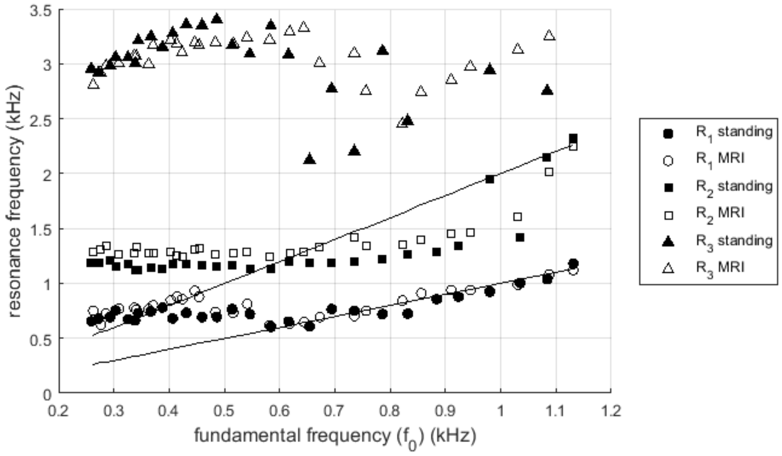


(b) /u/ vowel

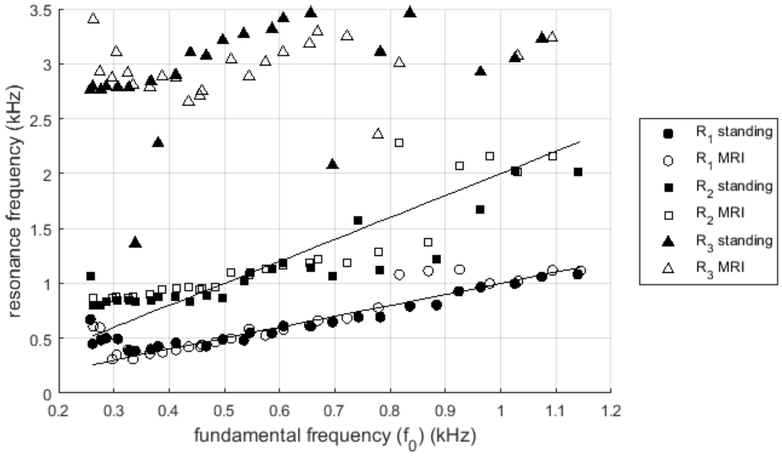


(c) /i/ vowel

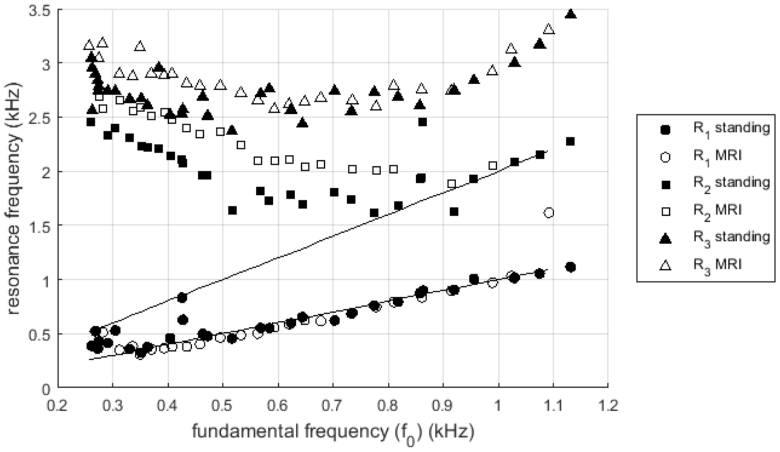
Figure 8: The first three resonances plotted against fundamental frequency, for subject 5, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom).



(a) /ɑ/ vowel



(b) /u/ vowel

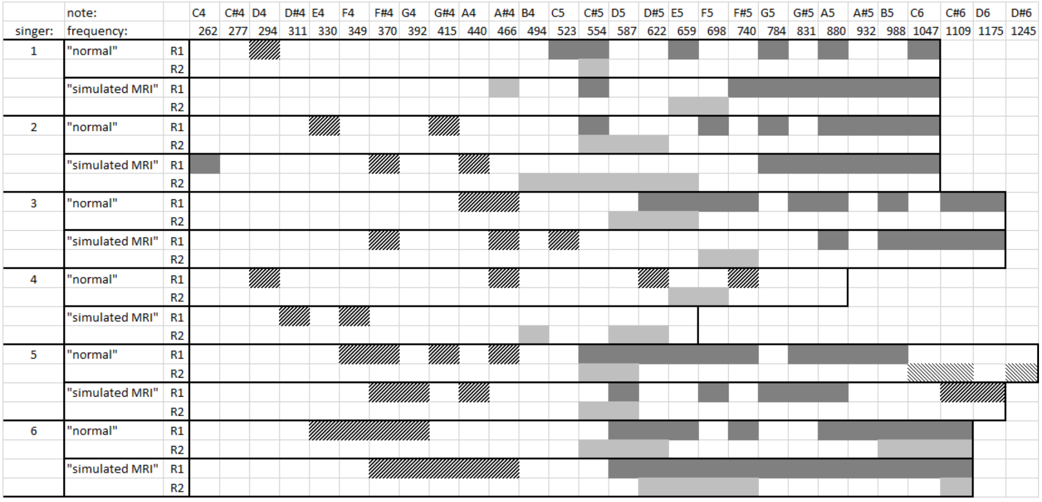


(c) /i/ vowel

Figure 9: The first three resonances plotted against fundamental frequency, for subject 6, for the /ɑ/ vowel (top), /u/ vowel (middle) and /i/ vowel (bottom).

F Sopranos plots of resonance tuning

As discussed in section 6.2.1, this appendix includes the resonance tuning results for all singers and vowels, in both “normal” and “simulated MRI” conditions.



	note:		C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5	G5	G#5	A5	A#5	B5	C6	C#6	D6	D#6
singer:	frequency:		262	277	294	311	330	349	370	392	415	440	466	494	523	554	587	622	659	698	740	784	831	880	932	988	1047	1109	1175	1245
1	"normal"	R1			▨										▓	▓			▓			▓		▓			▓			
		R2														░														
	"simulated MRI"	R1											░			▓					▓	▓	▓	▓	▓	▓	▓			
		R2																	░	░										
2	"normal"	R1					▨				▨					▓				▓		▓		▓	▓	▓	▓			
		R2														░	░	░												
	"simulated MRI"	R1	▓						▨			▨										▓	▓	▓	▓	▓	▓			
		R2												░	░	░	░	░	░											
3	"normal"	R1										▨	▨					▓	▓	▓	▓		▓	▓		▓		▓	▓	
		R2															░	░	░											
	"simulated MRI"	R1							▨				▨		▨									▓		▓	▓	▓	▓	
		R2																		░	░									
4	"normal"	R1			▨								▨					▨			▨									
		R2																	░	░										
	"simulated MRI"	R1				▨		▨																						
		R2												░			░	░												
5	"normal"	R1						▨	▨		▨		▨			▓	▓	▓	▓	▓	▓		▓	▓	▓	▓				
		R2														░	░										▧	▧		▧
	"simulated MRI"	R1							▨	▨		▨					▓			▓		▓	▓	▓				▨	▨	
		R2														░	░													
6	"normal"	R1					▨	▨	▨	▨							▓	▓	▓		▓			▓	▓	▓	▓	▓		
		R2														░	░	░								░	░	░		
	"simulated MRI"	R1							▨	▨	▨	▨	▨				▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓		
		R2																░	░	░	░							░		

Figure 10: The resonance tuning strategies employed by each subject, for the /ɑ/ vowel, in both “normal” and “simulated MRI” conditions. The tuning strategies observed were $R_1 : f_0$ tuning (dark grey), $R_1 : 2f_0$ tuning (dark stripes), $R_2 : f_0$ tuning (light stripes) and $R_2 : 2f_0$ tuning (light grey).

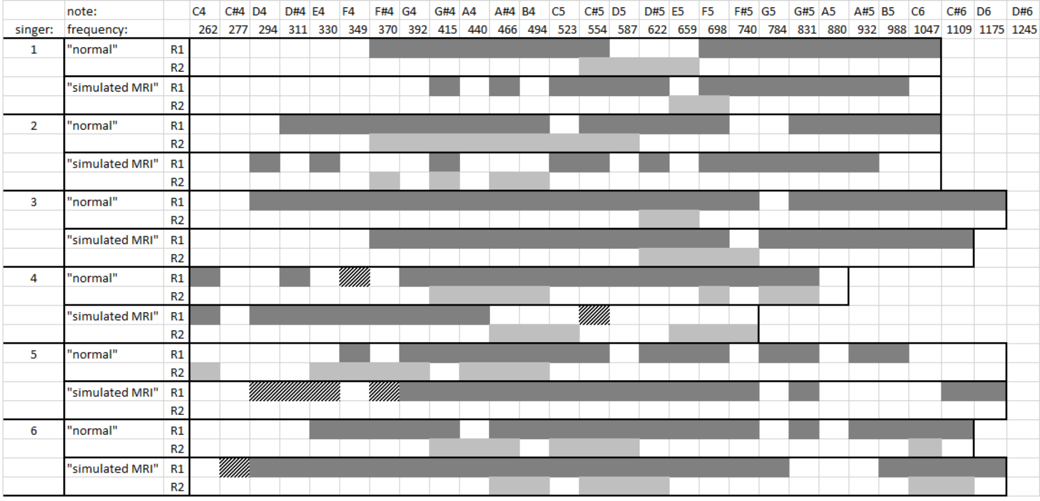


Figure 11: The resonance tuning strategies employed by each subject, for the /u/ vowel, in both “normal” and “simulated MRI” conditions. The tuning strategies observed were $R_1 : f_0$ tuning (dark grey), $R_1 : 2f_0$ tuning (dark stripes), $R_2 : f_0$ tuning (light stripes) and $R_2 : 2f_0$ tuning (light grey).

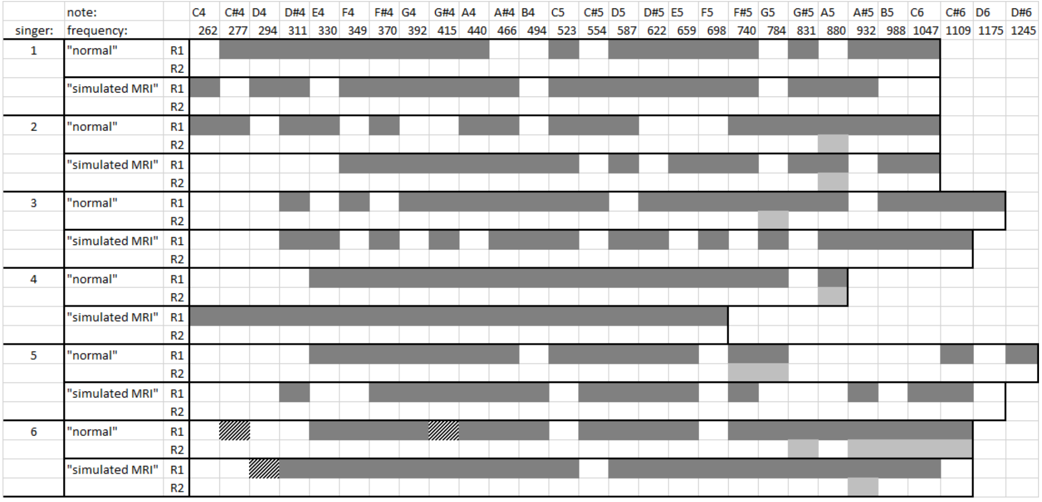
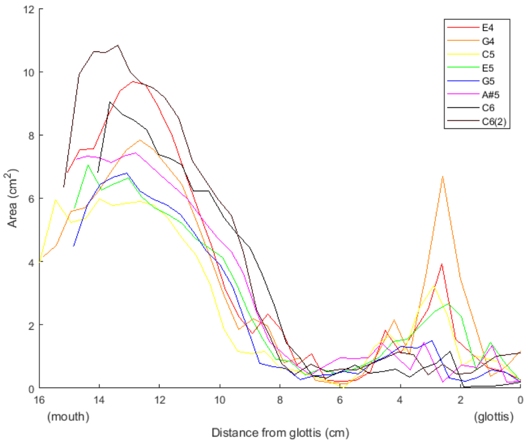


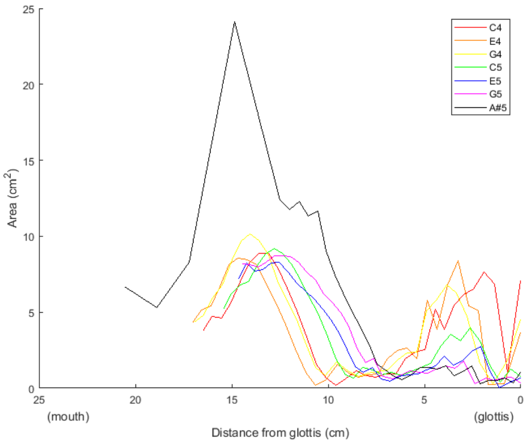
Figure 12: The resonance tuning strategies employed by each subject, for the /i/ vowel, in both “normal” and “simulated MRI” conditions. The tuning strategies observed were $R_1 : f_0$ tuning (dark grey), $R_1 : 2f_0$ tuning (dark stripes), $R_2 : f_0$ tuning (light stripes) and $R_2 : 2f_0$ tuning (light grey).

G Sopranos plots of area functions

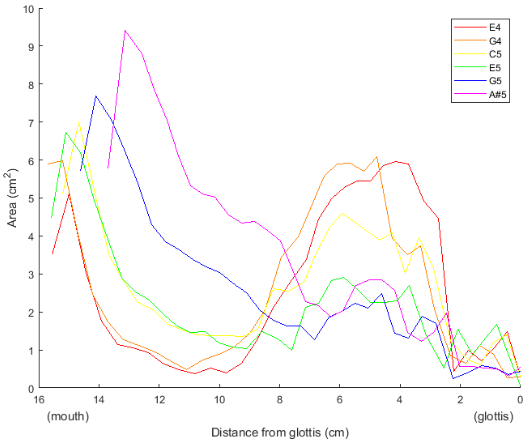
As discussed in section 6.2.3, this appendix includes plots of area functions for each singer, for each vowel and fundamental frequency investigated. For each Figure, the top subfigure (a) shows the area functions for all fundamental frequencies for the /ɑ/ vowel, the middle (b) for the /u/ vowel, and the bottom (c) for the /i/ vowel. Different fundamental frequencies are represented by different colours, as shown in the key for each Figure.



(a) /ɑ/ vowel

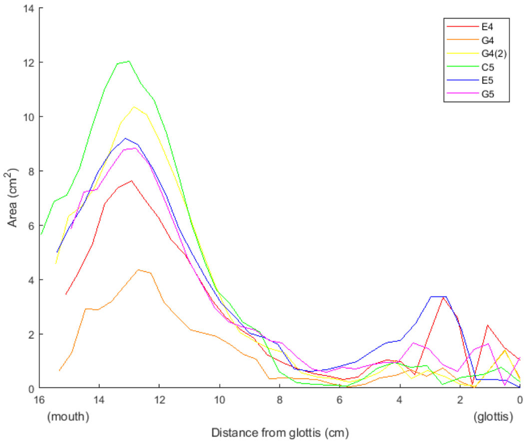


(b) /u/ vowel

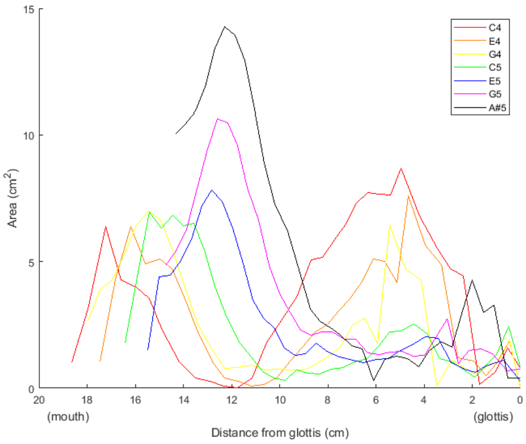


(c) /i/ vowel

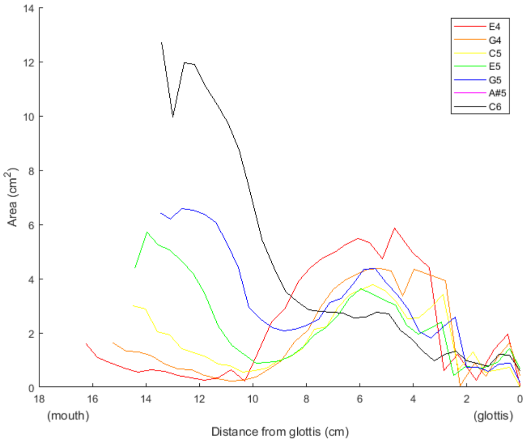
Figure 13: The area functions for all fundamental frequencies, for subject 1, for the /ɑ/ vowel (top), /u/ vowel (middle) and /(i)/ vowel (bottom).



(a) /ɑ/ vowel

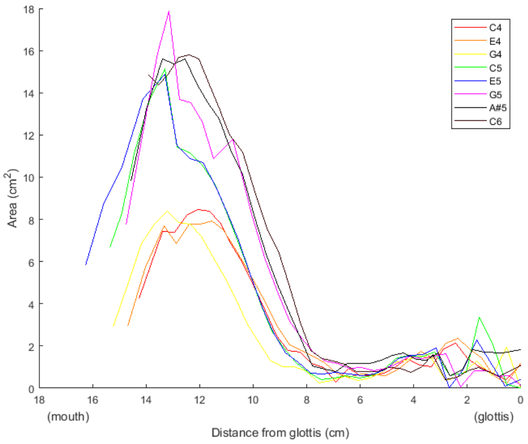


(b) /u/ vowel

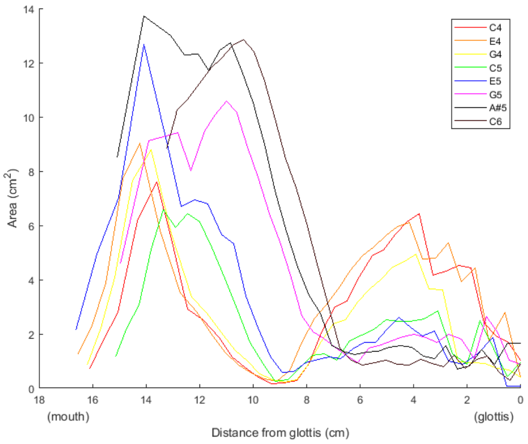


(c) /i/ vowel

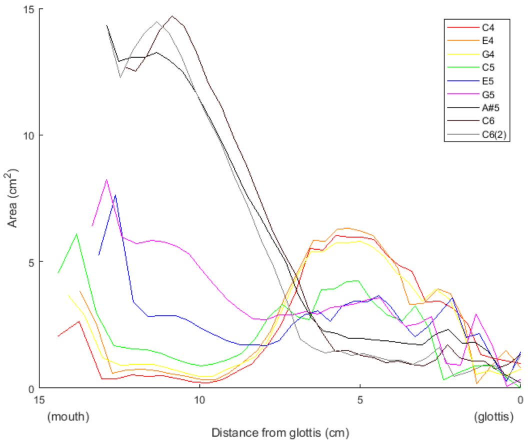
Figure 14: The area functions for all fundamental frequencies, for subject 2, for the /ɑ/ vowel (top), /u/ vowel (middle) and /(i)/ vowel (bottom).



(a) /ɑ/ vowel

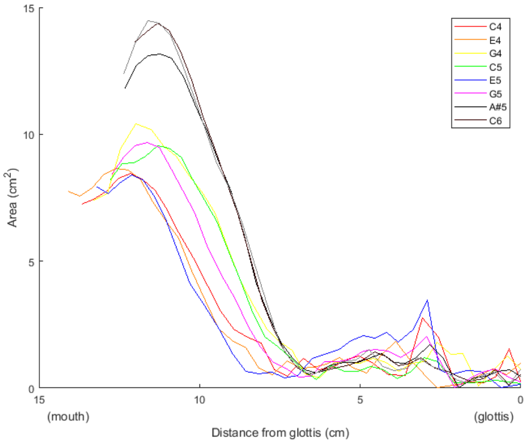


(b) /u/ vowel

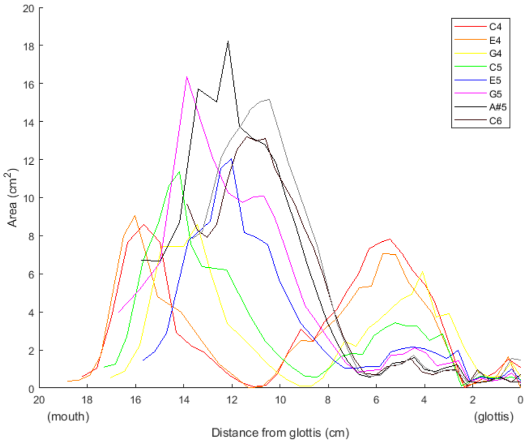


(c) /i/ vowel

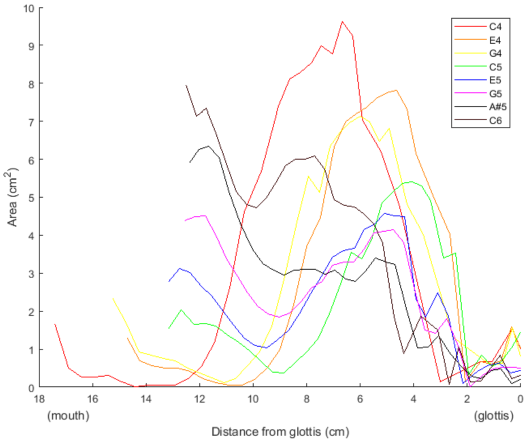
Figure 15: The area functions for all fundamental frequencies, for subject 3, for the /ɑ/ vowel (top), /u/ vowel (middle) and /(i)/ vowel (bottom).



(a) /ɑ/ vowel



(b) /u/ vowel

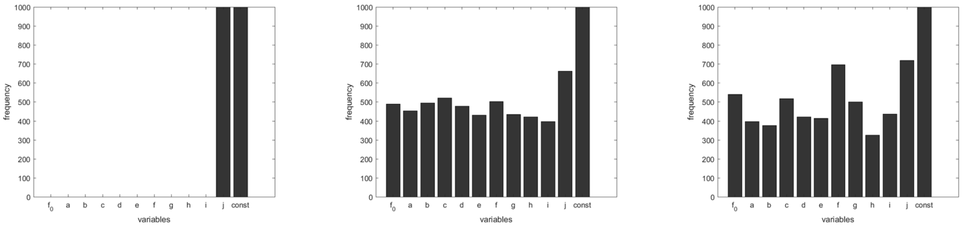


(c) /i/ vowel

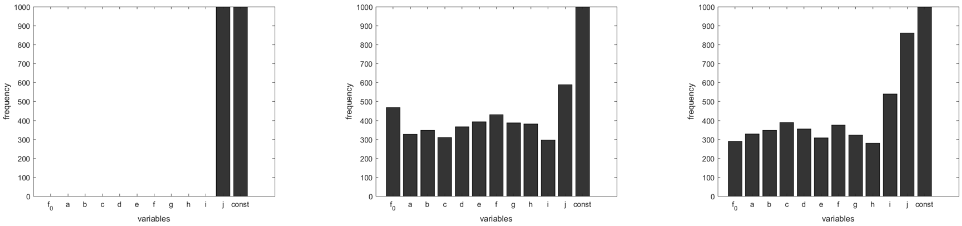
Figure 16: The area functions for all fundamental frequencies, for subject 6, for the /ɑ/ vowel (top), /u/ vowel (middle) and /(i)/ vowel (bottom).

H Feature selection plots for all singers

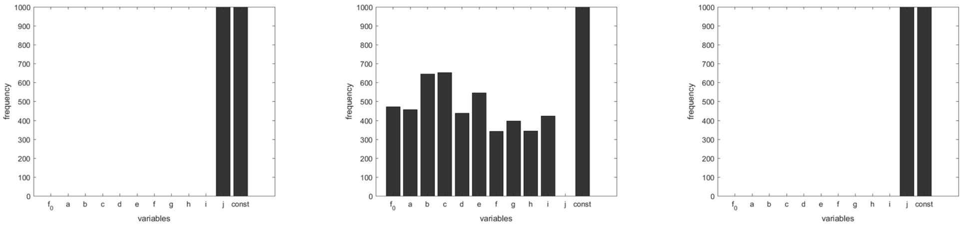
As discussed in section 6.3.3, this appendix includes the bar charts showing the results of the feature selection algorithm, split by vowel, resonance and singer.



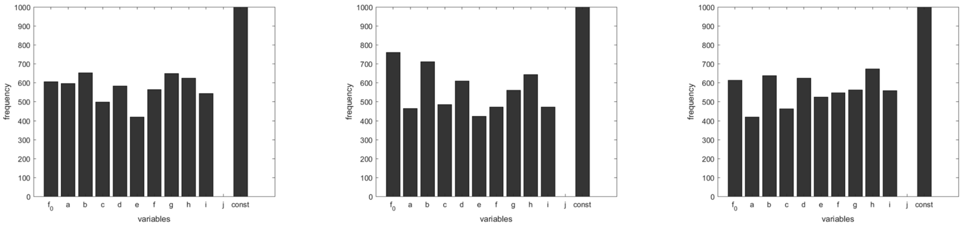
R_1 /ɑ/ Subject 1 R_2 /ɑ/ Subject 1 R_3 /ɑ/ Subject 1



R_1 /ɑ/ Subject 2 R_2 /ɑ/ Subject 2 R_3 /ɑ/ Subject 2

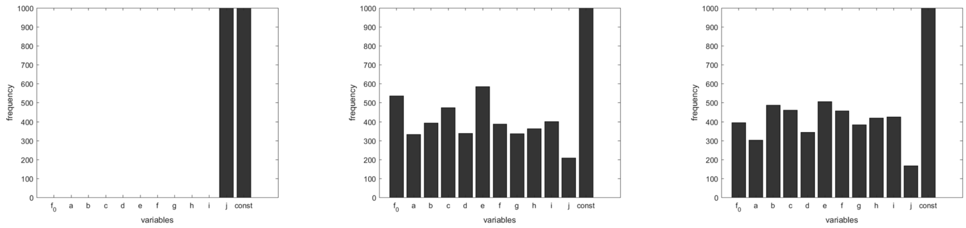


R_1 /ɑ/ Subject 3 R_2 /ɑ/ Subject 3 R_3 /ɑ/ Subject 3

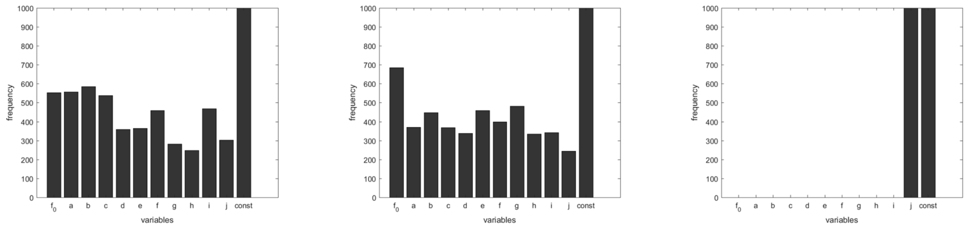


R_1 /ɑ/ Subject 6 R_2 /ɑ/ Subject 6 R_3 /ɑ/ Subject 6

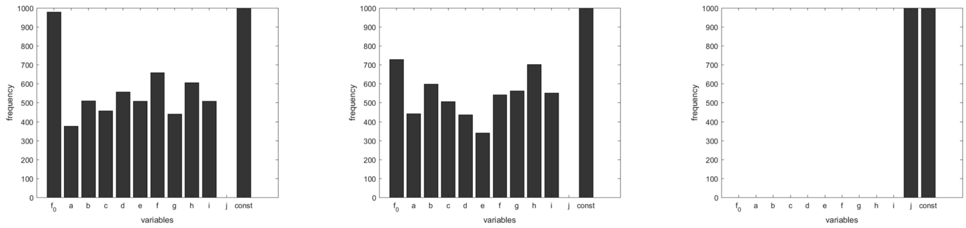
Figure 17: The variables chosen for the /ɑ/ vowel, for each resonance and subject.



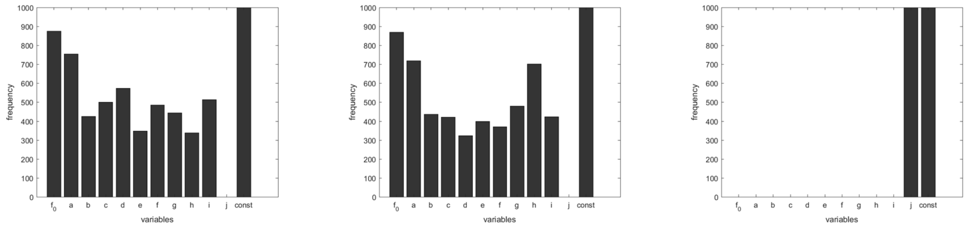
R_1 /u/ Subject 1 R_2 /u/ Subject 1 R_3 /u/ Subject 1



R_1 /u/ Subject 2 R_2 /u/ Subject 2 R_3 /u/ Subject 2

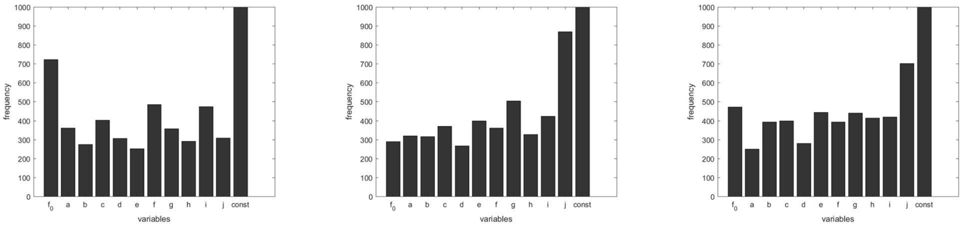


R_1 /u/ Subject 3 R_2 /u/ Subject 3 R_3 /u/ Subject 3

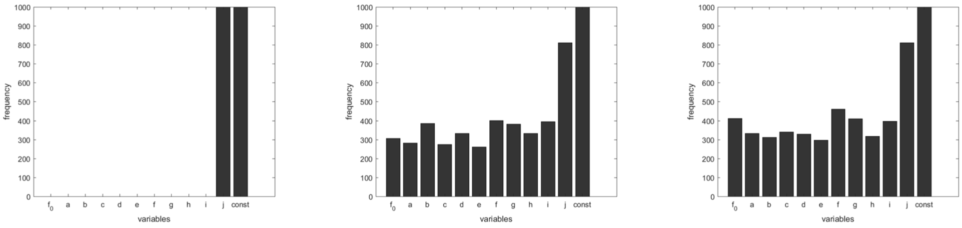


R_1 /u/ Subject 6 R_2 /u/ Subject 6 R_3 /u/ Subject 6

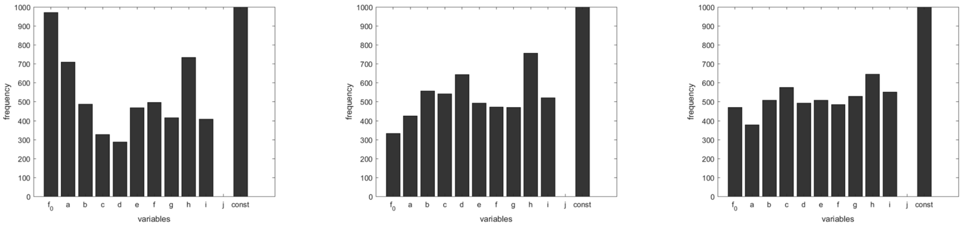
Figure 18: The variables chosen for the /u/ vowel, for each resonance and subject.



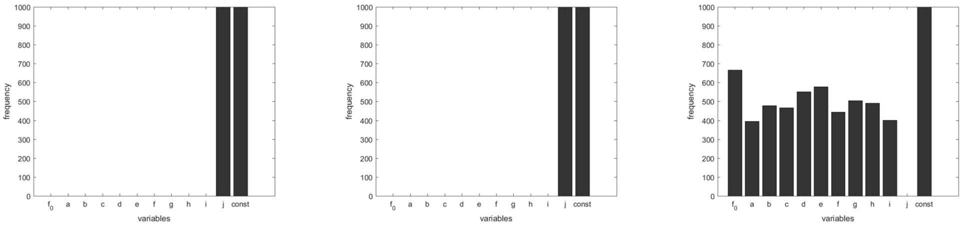
R_1 /i/ Subject 1 R_2 /i/ Subject 1 R_3 /i/ Subject 1



R_1 /i/ Subject 2 R_2 /i/ Subject 2 R_3 /i/ Subject 2



R_1 /i/ Subject 3 R_2 /i/ Subject 3 R_3 /i/ Subject 3



R_1 /i/ Subject 6 R_2 /i/ Subject 6 R_3 /i/ Subject 6

Figure 19: The variables chosen for the /i/ vowel, for each resonance and subject.

I LF model details

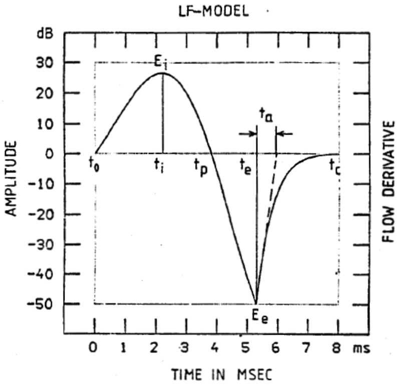


Figure 20: The parameters of the LF model.

As discussed in section 3.2.1, the Liljencrants-Fant Model [29] parameters used (setting $R_d = 1$) were:

$$F_a = 400Hz, \quad R_k = 0.30, \quad R_g = 1 \tag{1}$$

Where F_a is the cut-off frequency (accounting for the degree of spectral tilt), R_k specifies the relative duration of the falling branch from the peak at time T_p to the discontinuity point T_e, and R_g is a parameter which increases with a shortening f the rise time T_p.

$$R_a = t_a/t_0 \tag{2}$$

$$R_g = t_0/2t_p \tag{3}$$

$$R_k = (t_e - t_p)/t_p \tag{4}$$

$$OQ = t_e/t_0 \tag{5}$$

$$\begin{aligned} R_d = (t_d/t_0)(1/110) \\ = (U_0/E_0)(f_0/110) \\ \approx (0.5 + 1.2R_k)((R_k/4R_g) + R_a)/0.11 \end{aligned} \tag{6}$$

the parameters of the LF glottal model are calculated from the equations:

$$t_c = 1/f_0 \tag{7}$$

$$t_p = t_0/2R_g \tag{8}$$

$$t_a = 1/2\pi f_a \tag{9}$$

$$OQ = (1 + R_k)/2R_g \tag{10}$$

$$t_e = t_0(1 + R_k)/2Rg \tag{11}$$

References

[1] Manuel P Garcia, *A Complete Treatise on the Art of Singing*, 1984.

[2] Lelouda Stamou, “Plato and aristotle on music and music education: Lessons from ancient greece,” *International Journal of Music Education*, , no. 1, pp. 3–16, 2002.

[3] Richard Miller, *On the art of singing*, Oxford University Press, 1996.

[4] Malte Kob, Nathalie Henrich, Hanspeter Herzel, David Howard, Isao Tokuda, and Joe Wolfe, “Analysing and understanding the singing voice: Recent progress and open questions,” *Current Bioinformatics*, vol. 6, no. 3, pp. 362–374, 2011.

[5] Matthias Echternach, Johan Sundberg, Susan Arndt, Tobias Breyer, Michael Markl, Martin Schumacher, and Bernhard Richter, “Vocal tract and register changes analysed by real-time mri in male professional singersa pilot study,” *Logopedics Phoniatrics Vocology*, vol. 33, no. 2, pp. 67–73, 2008.

[6] Matthias Echternach, Johan Sundberg, Susan Arndt, Michael Markl, Martin Schumacher, and Bernhard Richter, “Vocal tract in female registersa dynamic real-time mri study,” *Journal of Voice*, vol. 24, no. 2, pp. 133–139, 2010.

[7] Erik Bresch and Shrikanth Narayanan, “Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing,”

The Journal of the Acoustical Society of America, vol. 128, no. 5, pp. EL335–EL341, 2010.

[8] Rebecca R Vos, Damian T Murphy, David M Howard, and Helena Daffern, “Determining the relevant criteria for three-dimensional vocal tract characterization,” *Journal of Voice*, vol. 32, no. 2, pp. 130–142, 2018.

[9] Nathalie Henrich, “Mirroring the voice from garcia to the present day: Some insights into singing voice registers,” *Logopedics Phonatrics Vocology*, vol. 31, no. 1, pp. 3–14, 2006.

[10] Johan Sundberg, “Research on the singing voice in retrospect,” *TMH-QPSR*, vol. 45, no. 1, pp. 11–22, 2003.

[11] Johan Sundberg, “Level and center frequency of the singer’s formant,” *Journal of voice*, vol. 15, no. 2, pp. 176–186, 2001.

[12] Rudolf Weiss, WS Brown Jr, and Jack Moris, “Singer’s formant in sopranos: fact or fiction?,” *Journal of Voice*, vol. 15, no. 4, pp. 457–468, 2001.

[13] Johan Sundberg, “Formant technique in a professional female singer,” *Acta Acustica united with Acustica*, vol. 32, no. 2, pp. 89–96, 1975.

[14] Nathalie Henrich, John Smith, and Joe Wolfe, “Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones,” *The Journal of the Acoustical Society of America*, vol. 129, pp. 1024, 2011.

[15] Maëva Garnier, Nathalie Henrich, John Smith, Joe Wolfe, et al., “The tuning of vocal resonances and the upper limit to the high soprano range,” in *Proceedings of the International Symposium on Music Acoustics ISMA 2010*, 2010.

[16] “Home feed — researchgate, [https://www.researchgate.net/],” .

[17] “Google scholar, [https://scholar.google.co.uk/],” .

[18] RD Kent and C Read, “The acoustic analysis of speech, 1992,” *Singular, San Diego.*

[19] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, 1987.

[20] David Martin Howard and Jamie Angus, *Acoustics and psychoacoustics*, Taylor & Francis, 2009.

[21] Nathalie Henrich, Christophe dAlessandro, Boris Doval, and Michèle Castellengo, “Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency,” *The Journal of the Acoustical Society of America*, vol. 117, no. 3, pp. 1417–1430, 2005.

[22] David M Howard, “Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers,” *Journal of Voice*, vol. 9, no. 2, pp. 163–172, 1995.

[23] Jonathan Harrington and Steve Cassidy, *Techniques in speech acoustics*, Kluwer Academic Publisher, 1999.

[24] Ray D Kent, “Vocal tract acoustics,” *Journal of voice*, vol. 7, no. 2, pp. 97–117, 1993.

[25] Gunnar Fant, *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*, vol. 2, Walter de Gruyter, 1971.

[26] Gunnar Fant, “The lf-model revisited. transformations and frequency domain analysis,” *Speech Trans. Lab. Q. Rep., Royal Inst. of Tech. Stockholm*, vol. 2, no. 3, pp. 40, 1995.

[27] Ingo R Titze, “Nonlinear source–filter coupling in phonation: Theory,” *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. 1902–1915, 2008.

[28] Ingo R Titze, "Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings," *Journal of Speech, Language and Hearing Research*, vol. 49, no. 2, pp. 448, 2006.

[29] Gunnar Fant, Johan Liljencrants, and Qi-guang Lin, "A four-parameter model of glottal flow," *STL-QPSR*, vol. 4, no. 1985, pp. 1–13, 1985.

[30] D Wong, J Markel, and A Gray Jr, "Least squares glottal inverse filtering from the acoustic speech waveform," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 27, no. 4, pp. 350–355, 1979.

[31] Donald G Childers and Chun-Fan Wong, "Measuring and modeling vocal source-tract interaction," *Biomedical Engineering, IEEE Transactions on*, vol. 41, no. 7, pp. 663–671, 1994.

[32] Robert Thayer Sataloff, *Voice Science*, Plural Publishing Inc. San Diego, Oxford, 2005.

[33] JW Van den Berg, JT Zantema, and P Doornenbal Jr, "On the air resistance and the bernoulli effect of the human larynx," *The journal of the acoustical society of America*, vol. 29, no. 5, pp. 626–631, 1957.

[34] Henry J Rubin, "The neurochronaxic theory of voice productiona refutation," *AMA archives of otolaryngology*, vol. 71, no. 6, pp. 913–920, 1960.

[35] Janwillem Van den Berg, "Myoelastic-aerodynamic theory of voice production.," *Journal of speech and hearing research*, , no. 1, pp. 227–44, 1958.

[36] James L Flanagan and Lois Landgraf, "Self-oscillating source for vocal-tract synthesizers," *Audio and Electroacoustics, IEEE Transactions on*, vol. 16, no. 1, pp. 57–64, 1968.

[37] Ingo R Titze, “The physics of small-amplitude oscillation of the vocal folds,” *The Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1536–1552, 1988.

[38] Brad H Story and Ingo R Titze, “Voice simulation with a body-cover model of the vocal folds,” *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1249–1260, 1995.

[39] Kenzo Ishizaka and James L Flanagan, “Synthesis of voiced sounds from a two-mass model of the vocal cords,” *Bell system technical journal*, vol. 51, no. 6, pp. 1233–1268, 1972.

[40] Ingo R Titze, “The human vocal cords: a mathematical model,” *Phonetica*, vol. 28, no. 3-4, pp. 129–170, 1973.

[41] Arend M Sulter, Harm K Schutte, and Donald G Miller, “Standardized laryngeal videostroboscopic rating: differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age,” *Journal of Voice*, vol. 10, no. 2, pp. 175–189, 1996.

[42] A-M Laukkanen, Ingo R Titze, Henry Hoffman, and Eileen Finnegan, “Effects of a semioccluded vocal tract on laryngeal muscle activity and glottal adduction in a single female subject,” *Folia Phoniatrica et Logopaedica*, vol. 60, no. 6, pp. 298–311, 2008.

[43] Richard Miller, *Training soprano voices*, Oxford University Press, 2000.

[44] Barrie Thorne, Cheris Kramarae, and Nancy Henley, *Language, gender and society*, Newbury House Rowley, MA, 1983.

[45] Christine C Bergan, Ingo R Titze, and Brad Story, “The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice,” *Journal of Voice*, vol. 18, no. 3, pp. 305–317, 2004.

[46] David M. Howard, Jenevora Williams, and Christian T. Herbst, “Ring in the solo child singing voice,” *Journal of Voice*, , no. 0, pp. –, 2013.

[47] Katherine Verdolini, Young Min, Ingo R Titze, Jon Lemke, Kice Brown, Miriam van Mersbergen, Jack Jiang, and Kim Fisher, "Biological mechanisms underlying voice changes due to dehydration," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 2, pp. 268–281, 2002.

[48] Christopher Barlow and Jeannette LoVetri, "Closed quotient and spectral measures of female adolescent singers in different singing styles," *Journal of Voice*, vol. 24, no. 3, pp. 314–318, 2010.

[49] David M Howard, "Quantifying developmental singing voice changes in children," in *1st International Conference on the Physiology and Acoustics of Singing*, 2002, pp. 1–16.

[50] Friederike Roers, Dirk Mürbe, and Johan Sundberg, "Predicted singers' vocal fold lengths and voice classificationa study of x-ray morphological measures," *Journal of Voice*, vol. 23, no. 4, pp. 408–413, 2009.

[51] Dr T Balu, "Anatomy of larynx by drtbalu," [Accessed: 17th March 2015].

[52] Emil Behnke, *The Mechanism of the Human Voice*, J. CURWEN & SONS Ltd, 1880.

[53] James Stark, *Bel canto: a history of vocal pedagogy*, University of Toronto Press, 2003.

[54] Manuel Garcia and Donald V Paschke, *A complete treatise on the art of singing: complete and unabridged*, Da Capo Pr, 1975.

[55] Bernard Roubeau, Nathalie Henrich, and Michèle Castellengo, "Laryngeal vibratory mechanisms: The notion of vocal register revisited," *Journal of Voice*, vol. 23, no. 4, pp. 425–438, 2009.

[56] John Smith, Nathalie Henrich, Maëva Garnier, and Joe Wolfe, "The acoustics of registers and resonances in singing," 2011.

[57] Donald Gray; Miller, *Registers in Singing. Empirical and Systematic Studies in the Theory of the Singing Voice*, University of Groningen, 2000.

[58] Elodie Joliveau, John Smith, and Joe Wolfe, “Acoustics: tuning of vocal tract resonance by sopranos,” *Nature*, vol. 427, no. 6970, pp. 116–116, 2004.

[59] Maëva Garnier, Nathalie Henrich, John Smith, and Joe Wolfe, “Vocal tract adjustments in the high soprano range,” *Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3771–3780, 2010.

[60] Charles Mayo Goss, “Gray’s anatomy of the human body,” *Academic Medicine*, vol. 35, no. 1, pp. 90, 1960.

[61] Keith Johnson, *Acoustic and auditory phonetics*, Karger Publishers, 2004.

[62] Rebecca Vos, “Vocal tract extraction from the vocal signal, for voice therapy and training,” M.S. thesis, University of Salford, 2013.

[63] Johan Sundberg et al., *The acoustics of the singing voice*, Scientific American, 1977.

[64] Gunnar Fant, *Acoustic Theory of Speech Production*, Mouton & Co, The Hague, Netherlands, 1960.

[65] Ingo R Titze, Ronald J Baken, Kenneth W Bozeman, Svante Granqvist, Nathalie Henrich, Christian T Herbst, David M Howard, Eric J Hunter, Dean Kaelin, Raymond D Kent, et al., “Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization,” *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 3005–3007, 2015.

[66] James R Sawusch, “Effects of duration and formant movement on vowel perception,” in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on.* IEEE, 1996, vol. 4, pp. 2482–2485.

[67] Johan Sundberg and Jöorgen Skoog, “Dependence of jaw opening on pitch and vowel in singers,” *Journal of Voice*, vol. 11, no. 3, pp. 301–306, 1997.

[68] “Matlab, the mathworks inc. natick, massachusetts, version r2016a.,” *2016*.

[69] Alan V Oppenheim, “Speech spectrograms using the fast fourier transform,” *Spectrum, IEEE*, vol. 7, no. 8, pp. 57–62, 1970.

[70] Paul Boersma and D Weenik, “Praat: a system for doing phonetics by computer. report of the institute of phonetic sciences of the university of amsterdam,” *Amsterdam: University of Amsterdam*, 1996.

[71] Maria Nordenberg and Johan Sundberg, “Effect on ltas of vocal loudness variation,” *Logopedics Phoniatrics Vocology*, vol. 29, no. 4, pp. 183–191, 2004.

[72] Timo Leino, Anne-Maria Laukkanen, and Vojtěch Radolf, “Formation of the actor’s/speaker’s formant: a study applying spectrum analysis and computer modeling,” *Journal of voice*, vol. 25, no. 2, pp. 150–158, 2011.

[73] Anders Löfqvist and Bengt Mandersson, “Long-time average spectrum of speech and voice analysis,” *Folia phoniatrica et logopaedica*, vol. 39, no. 5, pp. 221–229, 1987.

[74] Christopher Watts, Kathryn Barnes-Burroughs, Julie Estis, and Debra Blanton, “The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers,” *Journal of voice*, vol. 20, no. 1, pp. 82–88, 2006.

[75] T Leino, “Long-term average spectrum study on speaking voice quality in male actors,” in *SMAC93, Proceedings of the Stockholm Music Acoustics Conference*. The Royal Swedish Academy of Music Stockholm, 1993, vol. 28.

[76] Tadeus Nawka, Lutz Christian Anders, Mario Cebulla, and David Zurakowski, "The speaker's formant in male voices," *Journal of Voice*, vol. 11, no. 4, pp. 422–428, 1997.

[77] Elvira Mendoza, Nieves Valencia, Juana Muñoz, and Humberto Trujillo, "Differences in voice quality between men and women: use of the long-term average spectrum (ltas)," *Journal of Voice*, vol. 10, no. 1, pp. 59–66, 1996.

[78] Peta White, "Long-term average spectrum (ltas) analysis of sex-and gender-related differences in children's voices," *Logopedics Phoniatrics Vocology*, vol. 26, no. 3, pp. 97–101, 2001.

[79] John D Markel and AH Jr Gray, *Linear prediction of speech*, vol. 12, Springer Science & Business Media, 2013.

[80] Roy C Snell and Fausto Milinazzo, "Formant location from lpc analysis data," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 2, pp. 129–134, 1993.

[81] Matlab, "Polynomial roots - matlab roots - mathworks united kingdom," [Accessed: 12th February 2015].

[82] Rebecca Vos, Jamie AS Angus, and Brad H Story, "A new algorithm for vocal tract shape extraction from singer's waveforms," in *Audio Engineering Society Convention 136*. Audio Engineering Society, 2014.

[83] Randall B Monsen and A Maynard Engebretson, "The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction," *Journal of Speech, Language, and Hearing Research*, vol. 26, no. 1, pp. 89–97, 1983.

[84] Philippe Fabre, "Un procédé électrique percutané dinscription de laccolement glottique au cours de la phonation: glottographie de haute fréquence," *Bulletin de lAcadémie Nationale de Médecine*, vol. 141, pp. 66–69, 1957.

[85] Martin Rothenberg, "A multichannel electroglottograph," *Journal of Voice*, vol. 6, no. 1, pp. 36–43, 1992.

[86] Malte Kob and Tobias Frauenrath, "A system for parallel measurement of glottis opening and larynx position," *Biomedical Signal Processing and Control*, vol. 4, no. 3, pp. 221–228, 2009.

[87] Thomas Hézard, Thomas Hélie, Boris Doval, Nathalie Henrich, Malte Kob, et al., "Non-invasive vocal-folds monitoring using electrical imaging methods," in *Proceedings of the Workshop 100 years of electrical imaging*, 2012, pp. 1–4.

[88] Friederike Roers, Dirk Mürbe, and Johan Sundberg, "Voice classification and vocal tract of singers: a study of x-ray images and morphology," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 503–512, 2009.

[89] Amy Berrington de González and Sarah Darby, "Risk of cancer from diagnostic x-rays: estimates for the uk and 14 other countries," *The lancet*, vol. 363, no. 9406, pp. 345–351, 2004.

[90] Carstens Medizinelektronik GmbH, "Home- 3d electromagnetic articulograph — carstens medizinelektronik gmbh," [Accessed: 23rd January 2015].

[91] Q Meng, D Sen, S Wang, and L Hayes, "Impulse response measurement with sine sweeps and amplitude modulation schemes," in *Signal Processing and Communication Systems, 2008. ICSPCS 2008. 2nd International Conference on.* IEEE, 2008, pp. 1–5.

[92] Angelo Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108.* Audio Engineering Society, 2000.

[93] Osamu Fujimura and Jan Lindqvist, "Sweep-tone measurements of vocal-tract characteristics," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 541–558, 1971.

[94] Bertrand Delvaux and David Howard, "Using an exponential sine sweep to measure the vocal tract resonances," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 4206–4206, 2013.

[95] J Epps, JR Smith, and J Wolfe, "A novel instrument to measure acoustic resonances of the vocal tract during phonation," *Measurement Science and Technology*, vol. 8, no. 10, pp. 1112, 1997.

[96] "information/mri - ynic wiki, [https://www.ynic.york.ac.uk/information /mri]," .

[97] Yoon-Chul Kim, Jangwon Kim, Michael Proctor, Asterios Toutios, Krishna Nayak, Sungbok Lee, Shrikanth Narayanan, et al., "Toward automatic vocal tract area function estimation from accelerated three-dimensional magnetic resonance imaging," in *ISCA Workshop on Speech Production in Automatic Speech Recognition, Lyon, France*, 2013, pp. 2–5.

[98] Bertrand Delvaux and David Howard, "A new method to explore the spectral impact of the piriform fossae on the singing voice: benchmarking using mri-based 3d-printed vocal tracts," *PloS one*, vol. 9, no. 7, pp. e102680, 2014.

[99] Matthias Echternach, Peter Birkholz, Louisa Traser, Tabea V Flügge, Robert Kamberger, Fabian Burk, Michael Burdumy, and Bernhard Richter, "Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2586–2595, 2015.

[100] Shrikanth Narayanan, Krishna Nayak, Sungbok Lee, Abhinav Sethy, and Dani Byrd, "An approach to real-time magnetic resonance imaging for speech production," *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1771–1776, 2004.

[101] Hironori Takemoto, Kiyoshi Honda, Shinobu Masaki, Yasuhiro Shimada, and Ichiro Fujimoto, "Measurement of temporal changes in

vocal tract area function from 3d cine-mri data," *The Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1037–1049, 2006.

[102] Brad H Story, Ingo R Titze, and Eric A Hoffman, "Vocal tract area functions from magnetic resonance imaging," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 537–554, 1996.

[103] Hironori Takemoto, Tatsuya Kitamura, Hironori Nishimoto, and Kiyoshi Honda, "A method of tooth superimposition on mri data for accurate measurement of vocal tract shape and dimensions," *Acoustical science and technology*, vol. 25, no. 6, pp. 468–474, 2004.

[104] Ju Zhang, Kiyoshi Honda, and Jianguo Wei, "Tooth visualization in vowel production mr images for three-dimensional vocal tract modeling," *Speech Communication*, vol. 96, pp. 37–48, 2018.

[105] Chang-Sheng Yang, Hideki Kasuya, Sigeru Kano, and Toshihiko Sato, "Accurate measurement of vocal tract shapes using magnetic resonance imaging," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 78, no. 8, pp. 63–74, 1995.

[106] Mark Hasegawa-Johnson, Shamala Pizza, Abeer Alwan, Jul Setsu Alwan, and Katherine Haker, "Vowel category dependence of the relationship between palate height, tongue height, and oral area," *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 3, pp. 738–753, 2003.

[107] Tatsuya Kitamura, Hironori Nishimoto, Ichiro Fujimoto, and Yasuhiro Shimada, "Dental imaging using a magnetic resonance visible mouthpiece for measurement of vocal tract shape and dimensions," *Acoustical science and technology*, vol. 32, no. 5, pp. 224–227, 2011.

[108] Masahiko Wakumoto, Shinobu Masaki, Jianwu Dang, Kiyoshi Honda, Yasuhiro Shimada, Ichiro Fujimoto, and Yuji Nakamura, "Visualization of dental crown shape in an mri-based speech production study," *International Journal of Oral and Maxillofacial Surgery*, vol. 26, pp. 189–190, 1997.

[109] Silvia Olt and Peter M Jakob, “Contrast-enhanced dental mri for visualization of the teeth and jaw,” *Magnetic resonance in medicine*, vol. 52, no. 1, pp. 174–176, 2004.

[110] Eric J Hunter and Ingo R Titze, “Overlap of hearing and voicing ranges in singing,” *Journal of singing: the official journal of the National Association of Teachers of Singing*, vol. 61, no. 4, pp. 387–392, 2005.

[111] Miller, “Chapter 1, pavarotti: King of second formant tuning, [http://www.voiceinsideview.com/docs/miller%20chapter%201.pdf],” .

[112] Johan Sundberg, Filipa M.B. L, and Brian P. Gill, “Formant tuning strategies in professional male opera singers,” *Journal of Voice*, vol. 27, no. 3, pp. 278 – 288, 2013.

[113] Ingo R Titze, Sharyn Mapes, and Brad Story, “Acoustics of the tenor high voice,” *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1133–1142, 1994.

[114] DG Miller and Harm K Schutte, “Formant tuning in a professional baritone,” *Journal of Voice*, vol. 4, no. 3, pp. 231–237, 1990.

[115] Katrin Neumann, Patrick Schunda, Sebastian Hoth, and Harald A Euler, “The interplay between glottis and vocal tract during the male passaggio,” *Folia phoniatrica et logopaedica*, vol. 57, no. 5-6, pp. 308, 2005.

[116] Jennifer J Barnes, Pamela Davis, Jennifer Oates, and Janice Chapman, “The relationship between professional operatic soprano voice and high range spectral energy,” *The Journal of the Acoustical Society of America*, vol. 116, pp. 530, 2004.

[117] Gordon E Peterson and Harold L Barney, “Control methods used in a study of the vowels,” *The Journal of the Acoustical Society of America*, vol. 24, pp. 175, 1952.

[118] Arend M Sulter, Donald G Miller, Rienhart F Wolf, Harm K Schutte, Hero P Wit, and Eduard L Mooyaart, "On the relation between the dimensions and resonance characteristics of the vocal tract: a study with mri," *Magnetic resonance imaging*, vol. 10, no. 3, pp. 365–373, 1992.

[119] Thomas Baer, JC Gore, LC Gracco, and PW Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *The Journal of the Acoustical Society of America*, vol. 90, no. 2, pp. 799–828, 1991.

[120] Pierre Delattre and Donald C Freeman, "A dialect study of american rs by x-ray motion picture," *Linguistics*, vol. 6, no. 44, pp. 29–68, 1968.

[121] Joanne D Subtelny and N Oya, "Cineradiographic study of sibilants," *Folia Phoniatrica et Logopaedica*, vol. 24, no. 1, pp. 30–50, 1972.

[122] Sandra M Rua Ventura, Diamantino Rui S Freitas, Isabel Maria AP Ramos, and João Manuel RS Tavares, "Morphologic differences in the vocal tract resonance cavities of voice professionals: an mri-based study," *Journal of Voice*, vol. 27, no. 2, pp. 132–140, 2013.

[123] Friedemann Pabst and Johan Sundberg, "Tracking multi-channel electroglottograph measurement of larynx height in singers," *Logopedics Phonatrics Vocology*, vol. 18, no. 4, pp. 143–152, 1993.

[124] Helena Daffern, *Distinguishing characteristics of vocal techniques in the specialist performance of early music*, Ph.D. thesis, University of York, 2008.

[125] Gunilla Carlsson and Johan Sundberg, "Formant frequency tuning in singing," *Journal of Voice*, vol. 6, no. 3, pp. 256–260, 1992.

[126] Donald G Miller, Arend M Sulter, Harm K Schutte, and Rienhart F Wolf, "Comparison of vocal tract formants in singing andnonperiodic phonation," *Journal of Voice*, vol. 11, no. 1, pp. 1–11, 1997.

[127] Philippe Clément, Stéphane Hans, Dana M Hartl, Shinji Maeda, Jacqueline Vaissière, and Daniel Brasnu, “Vocal tract area function for vowels using three-dimensional magnetic resonance imaging. a preliminary study,” *Journal of Voice*, vol. 21, no. 5, pp. 522–530, 2007.

[128] J Steven Walker, “An investigation of the whistle register in the female voice,” *Journal of Voice*, vol. 2, no. 2, pp. 140–150, 1988.

[129] A Keilmann and F Michek, “Physiology and acoustic analysis of whistle voice of the woman,” *Folia phoniatrica*, vol. 45, no. 5, pp. 247–255, 1993.

[130] Theodore C Levin and Michael E Edgerton, “The throat singers of tuva,” *Scientific American*, vol. 281, no. 3, pp. 80–87, 1999.

[131] Nathalie Henrich, Mara Kiek, John Smith, and Joe Wolfe, “Resonance strategies used in bulgarian women’s singing style: A pilot study,” *Logopedics Phonatrics Vocology*, vol. 32, no. 4, pp. 171–177, 2007.

[132] James L Flanagan, “A difference limen for vowel formant frequency,” *The journal of the Acoustical Society of America*, vol. 27, no. 3, pp. 613–617, 1955.

[133] Paul Mermelstein, “Difference limens for formant frequencies of steady-state and consonant-bound vowels,” *The Journal of the Acoustical Society of America*, vol. 63, no. 2, pp. 572–580, 1978.

[134] James L Flanagan, “Difference limen for formant amplitude,” *Journal of Speech and Hearing disorders*, vol. 22, no. 2, pp. 205–212, 1957.

[135] Michael Kiefte, Teresa Enright, and Lacey Marshall, “The role of formant amplitude in the perception of/i/and/u,” *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2611–2621, 2010.

[136] Robert B Zajonc, “Mere exposure: A gateway to the subliminal,” *Current directions in psychological science*, vol. 10, no. 6, pp. 224–228, 2001.

[137] N Scotto di Carlo and Aline Germain, "A perceptual study of the influence of pitch on the intelligibility of sung vowels," *Phonetica*, vol. 42, no. 4, pp. 188–197, 1985.

[138] Martha S Benolken and Charles E Swanson, "The effect of pitch-related changes on the perception of sung vowels," *The Journal of the Acoustical Society of America*, vol. 87, pp. 1781, 1990.

[139] Rebecca R Vos, Damian T Murphy, David M Howard, and Helena Daffern, "The perception of formant tuning in soprano voices," *Journal of Voice*, 2017.

[140] Joe Wolfe, Maëva Garnier, and John Smith, "Vocal tract resonances in speech, singing, and playing musical instruments," *HFSP journal*, vol. 3, no. 1, pp. 6–23, 2009.

[141] Johan Sundberg, "Acoustic and psychoacoustic aspects of vocal vibrato," *STL-QPSR*, vol. 35, no. 2–3, pp. 45–68, 1994.

[142] John W Hawks and James D Miller, "A formant bandwidth estimation procedure for vowel synthesis [43.72. ja].," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1343–1344, 1995.

[143] G. Fant, "The acoustics of speech," in *In proceedings of the 3rd International Congress on Acoustics Stuttgart, Elsevier, New York, NY, 1961. volume 1, pages 188-201.*, 1961.

[144] Jessica Gillard and Michael Schutz, "The importance of amplitude envelope: Surveying the temporal structure of sounds in perceptual research," in *Proceedings of the Sound and Music Computing Conference.* Citeseer, 2013, pp. 62–68.

[145] Dik J Hermes, "Synthesis of breathy vowels: Some research methods," *Speech communication*, vol. 10, no. 5-6, pp. 497–502, 1991.

[146] Ingo R Titze, "A theoretical study of f_0 - f_1 interaction with application to resonant speaking and singing voice," *Journal of Voice*, vol. 18, no. 3, pp. 292–298, 2004.

[147] Qualtrics, "[computer program] provo, utah, usa, copyright 2015," .

[148] Michael Schoeffler, Fabian-Robert Stöter, Harald Bayerlein, Bernd Edler, and Jürgen Herre, "An experiment about estimating the number of instruments in polyphonic music: A comparison between internet and laboratory results.," in *ISMIR*, 2013, pp. 389–394.

[149] ITUR Recommendation, "Bs. 1534-1. method for the subjective assessment of intermediate sound quality (mushra)," *International Telecommunications Union, Geneva*, 2001.

[150] John Smith and Joe Wolfe, "Vowel-pitch matching in wagners operas: Implications for intelligibility and ease of singing," *J. Acoust. Soc. Am*, vol. 125, pp. 196–201, 2009.

[151] Rebecca R Vos, Helena Daffern, and David M Howard, "Resonance tuning in three girl choristers," *Journal of Voice*, vol. 31, no. 1, pp. 122–e1, 2017.

[152] Johan Sundberg, "Vocal tract resonance in singing," *National Association of Teachers of Singing Journal*, vol. 44, no. 4, pp. 11–31, 1988.

[153] Randall S Moore, "Comparison of children's and adults' vocal ranges and preferred tessituras in singing familiar songs," *Bulletin of the Council for Research in Music Education*, pp. 13–22, 1991.

[154] Graham F Welch and David M Howard, "Gendered voice in the cathedral choir," *Psychology of Music*, vol. 30, no. 1, pp. 102–120, 2002.

[155] NHS UK, "Stages of puberty: What happens to boys and girls - nhs.uk," .

[156] John R Smith, "Phasing of harmonic components to optimize measured signal-to-noise ratios of transfer functions," *Measurement Science and Technology*, vol. 6, no. 9, pp. 1343, 1995.

[157] Microsoft, "Microsoft excel," 2016.

[158] Tatsuya Kitamura, Hironori Takemoto, Kiyoshi Honda, Yasuhiro Shimada, Ichiro Fujimoto, Yuko Syakudo, Shinobu Masaki, Kagayaki Kuroda, Noboru Oku-Uchi, and Michio Senda, "Difference in vocal tract shape between upright and supine postures: Observations by an open-type mri scanner," *Acoustical Science and Technology*, vol. 26, no. 5, pp. 465–468, 2005.

[159] Louisa Traser, Michael Burdumy, Bernhard Richter, Marco Vicari, and Matthias Echternach, "The effect of supine and upright position on vocal tract configurations during singinga comparative study in professional tenors," *Journal of Voice*, vol. 27, no. 2, pp. 141–148, 2013.

[160] Maureen Stone, G Stock, Kevin Bunin, Kausum Kumar, M Epstein, Chandra Kambhamettu, Min Li, Vijay Parthasarathy, and J Prince, "Comparison of speech production in upright and supine position," *The Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 532–541, 2007.

[161] Matthew David Adam Speed, *Voice Synthesis Using the Three-Dimensional Digital Waveguide Mesh*, Ph.D. thesis, Department of Electronics, University of York, 2012.

[162] Olov Engwall, "Are static mri measurements representative of dynamic speech? results from a comparative study using mri, epg and ema.," in *INTERSPEECH*, 2000, pp. 17–20.

[163] Olov Engwall, "Assessing magnetic resonance imaging measurements," *Speech Production: Models, Phonetic Processes, and Techniques*, p. 300, 2013.

[164] Meribeth Bunch and Janice Chapman, "Taxonomy of singers used as subjects in scientific research," *Journal of Voice*, vol. 14, no. 3, pp. 363–369, 2000.

[165] "Laryngograph ltd, laryngograph ltd, 78 manor road, wallington, greater london, sm6 0ab, united kingdom, [http://www.laryngograph.com/," .

[166] Duncan Markham and Valerie Hazan, "Speech, hearing and language: work in progress," *The UCL Speaker Database*, vol. 14, pp. 1–17, 2002.

[167] MD Burkhard and RM Sachs, "Anthropometric manikin for acoustic research," *The Journal of the Acoustical Society of America*, vol. 58, no. 1, pp. 214–222, 1975.

[168] "Audacity(r) software is copyright (c) 1999-2012 audacity team. [web site: http://audacityteam.org/. it is free software distributed under the terms of the gnu general public license.] the name audacity(r) is a registered trademark of dominic mazzoni."," .

[169] Erik Bresch, Jon Nielsen, Krishna Nayak, and Shrikanth Narayanan, "Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 1791–1794, 2006.

[170] JA Peacock, "Two-dimensional goodness-of-fit testing in astronomy," *Monthly Notices of the Royal Astronomical Society*, vol. 202, no. 3, pp. 615–627, 1983.

[171] Etienne Lombard, "Le signe de l'elevation de la voix," *Ann. Mal. de L'Oreille et du Larynx*, pp. 101–119, 1911.

[172] Thomas F Cleveland, Johan Sundberg, and RE Stone, "Long-term-average spectrum characteristics of country singers during speaking and singing," *Journal of voice*, vol. 15, no. 1, pp. 54–60, 2001.

[173] version R2016a. The MathWorks Inc. Natick, Massachusetts, "Linear or rank correlation - matlab corr - mathworks united kingdom," .

[174] James D Evans, *Straightforward statistics for the behavioral sciences*, Brooks/Cole, 1996.

[175] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.

[176] C. Law J. Ahrens, B. Geveci, *ParaView: An End-User Tool for Large Data Visualization, Visualization Handbook, ISBN-13: 978-0123875822*, Elsevier, 2005.

[177] Eric A Hoffman and Warren B Gefter, “Multimodality imaging of the upper airway: Mri, mr spectroscopy, and ultrafast x-ray ct.,” *Progress in clinical and biological research*, vol. 345, pp. 291, 1990.

[178] MathWorks, “Feature selection - matlab & simulink,” [Accessed: 2nd April 2018].

[179] Loren Shure, “Subset selection and regularization - loren on the art of matlab - matlab & simulink,” .

[180] F Joseph, J Hair, W Black, B Babin, and R Anderson, *Multivariate data analysis (7th Edition)*, United States: Pearson Education Ltd., 2014.

[181] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[182] Matlab, “Run test for randomness - matlab runstest - mathworks united kingdom,” .

[183] MathWorks, “Principal component analysis of raw data - matlab pca - matlab & simulink,” [Accessed: 2nd April 2018].

[184] Björn EF Lindblom and Johan EF Sundberg, “Acoustical consequences of lip, tongue, jaw, and larynx movement,” *The Journal of the Acoustical Society of America*, vol. 50, no. 4B, pp. 1166–1179, 1971.

[185] John Smith, Joe Wolfe, Nathalie Henrich, and Maëva Garnier, “Diverse resonance tuning strategies for women singers,” in *Stockholm Music Acoustics Conference SMAC 2013*, 2013, pp. 306–310.

[186] N Scotto Di Carlo and Denis Autesserre, "Movements of the velum in singing," *J. Res. Sing*, vol. 11, pp. 3–13, 1987.

[187] Shinji Maeda, "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech production and speech modelling*, pp. 131–149. Springer, 1990.

[188] Reza Zolfaghari, Nicolas Epain, Craig T Jin, Joan Glaunès, and Anthony Tew, "Generating a morphable model of ears," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on.* IEEE, 2016, pp. 1771–1775.

[189] "Phonetic symbols for english," https://www.phon.ucl.ac.uk/home/wells/phoneticsymbolsforenglish.htm, [Accessed 21-07-18].