The Acoustic and Psychoacoustic Characteristics of Barbershop Singing

Peter James Cookson

Doctor of Philosophy

University of York

Physics, Engineering and Technology

November 2025

Abstract

For a cappella singing groups to produce a memorable listener experience they need to be able to generate a broad palette of sounds. This capability requires an understanding of the key factors which affect their acoustic output. The advent of computers capable of rapid real time analysis has enabled the development of objective metrics for identifying and classifying sounds. In parallel, perceptual descriptions of sounds have evolved so practitioners can communicate their intent to performers and listeners can describe their experiences. Having these two ways of delineating sounds raises the question of whether the acoustic metrics and the perceptual descriptions correlate and what are the key variables affecting their values.

This thesis describes investigations into the objective acoustic metrics and the perceived acoustic features of Barbershop singing. A research programme comprising an analysis of published quartet recordings, the development of a computer model to simulate a Barbershop quartet, live testing with a Barbershop quartet in an audio laboratory and perception tests with a panel of Barbershop judges identified the primary influences of compositional choices (chord structure, chord pitch and vowel sung) and singing characteristics which affected the acoustic metrics and the perceptual features. The distinctive qualitative properties of Barbershop have been described as Lock, Ring and Expanded Sound. From the research it is concluded Lock correlates with intonation accuracy. Ring correlates with a high spectral centroid in the frequency domain and Expanded Sound does not correlate with spectral spread, nor is it a result of combination tones, but is more likely to correlate with increasing spectral amplitude as a result of vocal efficiency. The effects can be enhanced in sustained chords as the singers make real time adjustments to their vocal techniques. Practical applications of the research which could help practitioners achieve these performance goals are presented.

Contents

List of Figures

List of Tables

Acknowledgements

This research would never have started, or concluded, if it had not been for the enthusiasm and encouragement of my supervisor Professor Helena Daffern. From the first day I tentatively expressed an interest in doing research on a cappella singing, through the following years of learning and thinking, her knowledge, interest, stimulating discussions and thoughtful suggestions have made my research journey a rewarding experience for which I will always be grateful.

My thanks are also due to Professor Gavin Kearney who has contributed complementary support throughout, always being willing to provide advice and guidance when needed. Andrew Chadwick, the Technical Team Manager in the AudioLab, was invaluable in providing extensive technical support to make the laboratory testing with the Barbershop quartet effective and efficient.

I am especially grateful to the Barbershop quartet Sound Hypothesis (Arran Bayliss-Chalmers, Conrad Godfrey, Hugh Blayney, Alex Moore) who kindly volunteered their time to perform in the laboratory tests which constitute a key component of this research.

The online perception tests were only possible because judges from both the British Association of Barbershop Singers and the Ladies Association of British Barbershop Singers were willing to share their expertise and time.

Finally, I am truly indebted to my wife, Dr Claire Cookson, for her patience, support and encouragement while I consumed valuable days of our retirement years to indulge in my research interests and for displaying those same attributes through all the intervening years since we married in 1974 when we were both PhD students.

Funding

The laboratory testing at the AudioLab at the University of York was generously supported by SEMPRE (Society for Education, Music and Psychology Research) to cover the costs of travel and subsistence for the quartet Sound Hypothesis.

Declaration of Authorship

I 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 acknowledged as references.

I also declare that some of the work in this thesis has been presented previously at conferences:

"Factors affecting the timbral characteristics of barbershop singing", P J Cookson, H Daffern, presented at the 14th Pan-European Voice Conference (PEVOC-14), Tallinn, Estonia, 24th - 27th August 2022

Chapter 1

Introduction

A desirable skill for singing ensembles is being able to create a diverse palette of sounds so they can enrich listener experience. To develop this capability composers, arrangers, lyricists, directors and performers need to understand the factors within their control which can affect an ensemble's acoustic output. The advent of computers capable of rapid real-time analysis has stimulated the development of objective measurements which can be used to identify and classify sounds (Peeters et al. 2011). These *absolute* acoustic measures can be used for comparative studies of the variables which affect an ensemble's sound output. It is equally beneficial to have consistent descriptions for the perception of the sounds so practitioners can communicate their intent to performers and listeners can describe their experiences. Historically the description of sounds has been by qualitative terms such as loud, quiet, bright, dark (Helmholtz 1877) and the way this vocabulary has evolved is largely subjective and derived empirically.

Having these two ways of delineating sounds raises the question of whether there are direct correlations between the absolute acoustic measures and the perceptual descriptions to the extent that when making an intentional adjustment to a composition or performance technique a predictable change in listener perception will result.

The research described in this thesis is a study of both the objective acoustic metrics and the perceived acoustic features of the musical genre of "Barbershop" singing. Barbershop can be considered to be a subset of a cappella singing with 'unique' differentiating features. It is primarily based on homophonic chords, precisely tuned using just intonation, with no vibrato or tremolo. Barbershop chords are strictly limited to four notes sung by the Tenor, Lead, Baritone and Bass voice parts. The Tenor and the Bass almost without exception sing the highest and the lowest notes in a chord respectively with the Lead singing the melody line. The Baritone may sing above or below the Lead. Chords are in close harmony resulting in the frequency spectra of the individual voices substantially overlapping thereby producing strong reinforcement of the harmonics in each voice part.

Barbershop singing offers a number of benefits for studying a cappella singing;

- Barbershop chords are frequently sustained for significant durations allowing high resolution, time-averaged (steady-state) spectral analysis to be performed
- an ensemble size of four singers (a Barbershop quartet) with one performer per voice part avoids the complexity of multiple voice choirs
- the music form is relatively simple to model mathematically for computer studies as each chord is at most four notes tuned in just intonation
- a large database of recordings made under competition conditions using standard recording practices is available
- a set of perceptual terms has been developed which judges are trained to use to give consistency when describing Barbershop sounds

In this research a set of objective acoustic metrics based on key spectral features are proposed for differentiating the sounds and the physical factors influencing their values are investigated. The correlation between these metrics and commonly used perceptual descriptions are then evaluated. Although the research described here is focused on Barbershop singing the methodology, results and conclusions have relevance for a cappella singing in general; the computer model can be used to investigate the influence of using different chord structures, tuning systems, mixed voice ensembles and the identification of resonance tuning opportunities for multiple voice ensembles.

1.1 Objective acoustic characteristics

An extensive catalogue of absolute acoustic measurements which are suitable for classifying sounds in both the time domain and the frequency domain is now available (Peeters et al. 2011; Giannakopoulos and Pikrakis 2014). In addition to pitch, loudness and duration, timbre is an important characteristic which is frequently used to describe a sound, even though it lacks a universally accepted definition (Siedenburg, Saitis, et al. 2019). The American National Standards Institute (ANSI 1997) defines timbre as the perceptual attribute which distinguishes two sounds that have the same pitch, loudness, and duration. Earlier Scholes (1970) gave a narrow definition of timbre as a tone quality in which "the one and only factor which conditions it is the presence or absence, or relative strength or weakness, of overtones". Scholes, like Helmholtz, considered that timbre can be defined by the frequency spectrum of a sound in a steady state. In this classical approach no allowance was made for the way spectral components vary with time. At the simplest level musical tones comprise three stages: attack, steady state, and decay and it is now accepted temporal variations are an important component in the concept of timbre (Risset and Wessel 1999). However, for practical reasons, the scope of this research is restricted to the timbre of sounds which are in a steady state.

Siedenburg, Fujinaga, et al. (2016) compared the differences between descriptions of timbre used for music information research (MIR) and music psychology. MIR uses a multitude of audio descriptors whereas music psychology adopts a very restricted set to describe the physical correlates of timbre perception. In MIR key measures are extracted using short-term Fourier transforms (STFT). By using this methodology the timbre of a sound in a steady state will have a strong association with the way the acoustic energy is distributed across the frequency range. At a basic level this energy profile can be represented by the absolute amplitude of the harmonics, the spectral centroid and the spectral spread. The spectral centroid is the centre of 'gravity' of the spectrum. The spectral spread is the second central moment of the spectrum.

Each voice in an ensemble has its own timbre. The ensemble will also possess a timbre which, to paraphrase Sandell (1995), results from the "fusion of the timbres of the individual voices into a single timbral image". The timbre of an ensemble results from combining the spectra of the individual voices. The net result will be affected by the accuracy of intonation between the voices and their relative intensities. Timbre also has a particular importance in singing because the timbres of the different phonemes not only affect the acoustic 'aesthetics' but also the intelligibility of the lyric.

1.2 Psychoacoustic characteristics

Psychoacoustics relates to the perception of sound by the human auditory system in contrast to acoustics which concerns the mechanisms involved in the physical transfer of sound.

In Barbershop singing primary psychoacoustic goals which distinguish the genre are idiomatically described as "Lock, Ring and Expanded Sound". Barbershop competition judges look for strong evidence of these features when determining their scores. The Barbershop Harmony Society (2023) explains these perceptual descriptors as:

- Lock "is the feeling associated with a justly in tune chord, whose quality is determined by the degree of intonation achieved in and between the individual voice parts"
- Ring "is the sound resulting from the production and reinforcement of harmonics in the composite voice parts, derived from the ringing quality contained in the individual voices"
- Expanded Sound is "the effect resulting from the combined interaction of voices singing with accurate intonation with uniform word sounds in good quality with proper volume relationships that reinforce the more compatible harmonics and combination tones and with precision all producing an effect greater than the sum of the individual voices"

These explanations largely revolve around the way the frequency spectra of the individual voices in a singing ensemble combine. The aggregation of the participating voices will determine the overall distribution of the acoustic energy across the harmonics in the frequency spectrum of the ensemble. It therefore seems reasonable, although it has not been previously established, to postulate there will be some correlation for a singing ensemble between the objective spectral metrics described above and these perceptual descriptions.

1.3 Statement of research hypothesis

The hypothesis being tested in this research is:

The distinctive sound of Barbershop singing can be defined by the absolute acoustic measures of accurate tuning in just intonation and high spectral centroid and broad spectral spread in the frequency domain, which are the physical correlates of the perceived qualities of Lock, Ring and Expanded Sound.

And as a corollary:

For a given chord structure, pitch and vowel determined at the point of composition/musical arrangement, a Barbershop ensemble can optimise this unique sound by adjusting, in real time, their individual vocal techniques.

1.4 Novel contributions

As a result of this research novel contributions to the subject area are:

- The development of a research framework for analysing ensemble singing using multiple data sources
- The assembly and analysis of a database of published Barbershop recordings to identify suitable acoustic metrics for defining the Barbershop genre
- The creation of a computer model simulating a Barbershop quartet for investigating the key variables affecting the acoustic output
- Capture of a unique data set of measurements of a high quality Barbershop quartet performing under controlled experimental conditions, revealing the differences in vocal characteristics between the singers and adjustments they make in real time when performing together
- Evaluation of the causal effects of compositional choices (chord structure, vowel, pitch) and individual vocal techniques on the acoustic output of Barbershop ensembles
- The results of perception tests with a panel of qualified Barbershop judges to assess the correlation between acoustic objective measurements and the idiomatic descriptions of Lock, Ring and Expanded Sound
- Providing practical guidance based on the research results for composers, arrangers, lyricists and performers to realise their acoustic objectives in Barbershop performances

1.5 Structure of the thesis

The research programme undertaken has four components:

- An analysis of sound samples extracted from published recordings of Barbershop quartets to obtain an indication of the range of acoustic metrics produced by different pitches, chord structure and vowels
- The development of a computer programme to model a Barbershop quartet to run parametric studies on pitch, chord structure and vowel

- Experimental tests with a live quartet performing prescribed exercises under laboratory conditions
- Perception tests with a panel of certified Barbershop judges using samples from each of the previous components to investigate the correlation between acoustic measures and perceptual descriptions

To set the context for the research previous relevant work on the acoustics of ensemble singing, computer modelling of the voice and perception of sounds is presented in Chapters 2, 3 and 4 respectively. In Chapter 5 the results from analysing published recordings of Barbershop quartets are presented which give an indication of the range of the selected metrics realised when using different chord structures, root frequencies and vowels. In Chapter 6 the construction of a computer model written to represent a Barbershop quartet is described. Parametric studies using different chord structures, root frequencies and vowels were run and the predicted acoustic characteristics are compared with the data derived from the quartet analysis described in Chapter 5. In Chapter 7 practical experiments undertaken with a live Barbershop quartet to further investigate the acoustic parameters in a controlled acoustic environment are presented. The extent to which the singers adjusted their vocal techniques to optimise the ensemble's output was assessed. In Chapter 8 the results of perception tests undertaken by trained Barbershop judges are presented. The perception tests were to establish if there is a correlation between the acoustic metrics and the perceptual descriptions commonly used by the Barbershop community. Chapter 9 draws together the key findings obtained from the different research components. The original research hypothesis is restated and conclusions are drawn on whether the objectives of the research have been achieved. Areas for future work which have been identified throughout this thesis are presented. Finally in Chapter 10 practical applications of the research results which may assist composers, arrangers, lyricists and performers are presented.

1.6 Terminology

In the literature on voice, the same terminology is sometimes used for items which have different meanings. A glossary is given in Appendix A but some key terms are highlighted here because of their particular importance to the subject matter.

1.6.1 Loudness, volume, power and intensity

Loudness is a perceptual quality assessed by the auditory system. Volume is a qualitative description in relative terms (eg pp,p,mp,mf,f,ff). Acoustic power is an objective physical measure of the amount of energy radiated into the air per second. Intensity is a measure of power per unit area and is measurable using a sound level meter.

1.6.2 Fundamental frequencies, harmonics, phonation frequencies, chord root frequencies and pitch

The fundamental frequency (F_0) is the lowest frequency in a periodic waveform and is also called the first harmonic. A harmonic is a component in a periodic waveform whose frequency is an integer multiple of the fundamental. The phonation frequency is the fundamental frequency of the voice, determined by the number of vibrations per second of the vocal cords. The root note of a chord is the fundamental note upon which a chord is built and establishes the chord's harmonic foundation. Pitch is a perceived property which is related to the fundamental frequency of a complex tone.

1.6.3 Formants and resonant frequencies

The terms formant and resonant frequency are often used interchangeably when describing vocal characteristics but they are conceptually different. Titze, Baken, et al. (2015) have made recommendations on the definition and notation of these terms to encourage consistency. In this document:

- Formants are peaks in the spectral envelope P(f) of the sound output of a singer and are labelled F_n for n = 1,2,3...
- Resonance frequencies are peaks in the transfer function T(f) of the vocal tract of a singer and are labelled R_n for n = 1,2,3...

In practice the frequencies of the first two formants and the first two vocal tract resonances of a singer will be close, but not necessarily coincidental. Whereas the "singers formant", which occurs around 3 kHz for male singers, is a clustering of the third, fourth and fifth resonances of the vocal tract (Sundberg 1974) and the formant and resonances are distinctly different. When modelling the vocal tract mathematically, the poles of the digital filters are taken to be the same as the vocal tract resonant frequencies.

1.6.4 Notation of chords

Where the notes in a chord are listed they are always in the sequence Bass:Baritone:Lead:Tenor with the frequency ratios between the voice parts in the same order. The scale degrees of the individual notes are given relative to the root note of the chord. For example a major triad chord with 10th voicing C3-G3-C4-E4 has frequency ratios 2:3:4:5 and scale degrees 1-5-8-10.

1.6.5 Categorisation of quartet samples

Three different sources of samples of Barbershop quartet singing are used in this research. In the text the sources are classified as:

Archive samples which are extracted from publicly available CDs and internet recordings. **Synthesised samples** which are created using the computer programme written to model a Barbershop quartet as part of this research. **AudioLab samples** which are recordings of Sound Hypothesis, a UK Barbershop gold medal quartet, performing under laboratory conditions at the AudioLab at the University of York as part of this research.

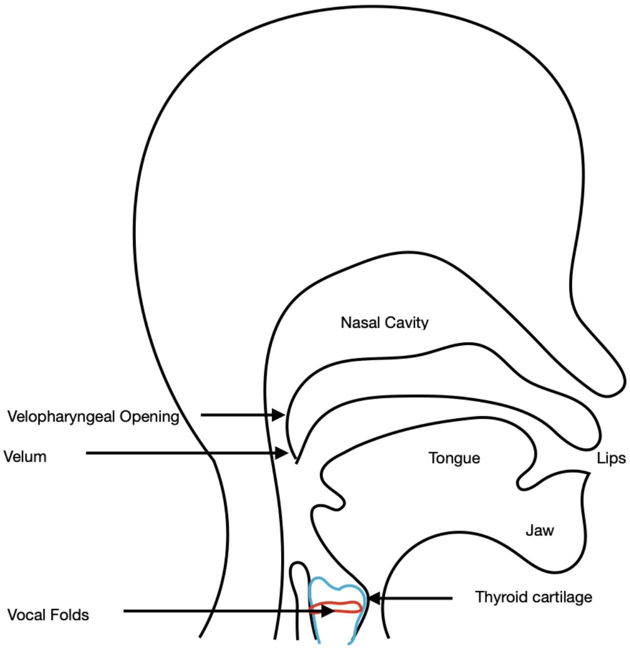
Chapter 2

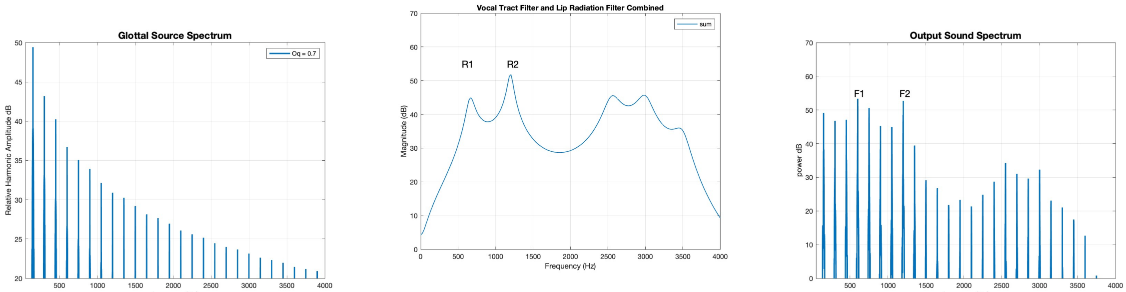
The acoustic characteristics of a cappella ensembles

A very substantial body of research into solo singing encompassing voice acoustics, sound perception and human physiology has been undertaken over the years, the results of which underpin current vocal pedagogy and performance; for example Sundberg (1988), Titze (2004), Dayme (2009). A comprehensive overview of progress in understanding solo singing was provided by Kob et al. (2011) encompassing features of the voice, modelling of the singing mechanism and methods of assessment. In comparison, and perhaps unsurprisingly because of the practical constraints, less research has been undertaken with choral ensembles. The knowledge derived from the research on solo singing is equally applicable to the choral singer but additional complexity is introduced when combining multiple voices in diverse acoustic environments. The overall acoustic output of an a cappella ensemble is not solely determined by the musical and lyrical choices of the composer or arranger; every participating singer can adjust their vocal production and affect the combined acoustic results. Before reviewing the previous research on choral singing, an overview of key aspects of solo voice production is given in the next section which will inform the later discussions.

2.1 Key aspects of solo voice production

The commonly adopted model to represent the acoustic production of a solo voice derives from the early work of Fant (1971) and is based on a source/filter concept. The model has three principal stages which can be explained by reference to Figure 2.1.





(a) source spectrum

(b) combined vocal tract and lip radiation filter

(c) Output spectrum

Figure 2.1: Source/filter model for voice production (a) the harmonic spectrum of the glottal flow (b) the envelope of the vocal tract filter and lip radiation combined (c) the spectrum of the resulting sound produced

A slowly varying flow of air driven from the lungs causes vibration of the vocal folds which generates a quasi-periodic glottal flow which has a frequency spectrum as shown in Figure 2.1(a). The strength of the harmonics in this spectrum decrease at around 6 dB/ octave. The glottal wave is then conditioned by the oral and nasal tracts and the acoustic impedance at the lips and nose before radiating into the space around the singer. When the velopharyngeal opening (VPO) is closed by raising the velum all the air and sound are directed through the mouth. The spectral envelope of the combined filtering effect of the vocal tract and lip radiation is illustrated in Figure 2.1(b). The shape of the filter's envelope affects the perceived timbre of the sound produced. The filter shape is determined by the positioning of the "articulators"; the lip and jaw openings, the tongue shape which affects the cross-sectional area of the tract along its length, and the vertical larynx position which affects the length of the tract. The peaks in the envelope of the filter define the vocal tract resonances (R1, R2 ...). The resulting output spectrum also has peaks which are defined here as the formants (F1, F2 ...). Each vowel sound is associated with a particular range of (F1, F2). In much of the literature on the voice the terms vocal tract resonances and formants are used interchangeably but they are conceptually different. The distinction is important for the research that follows. The first resonant frequency (R1) lies between 0 and 1 kHz and the second (R2) between 1 and 2 kHz. Generally the values of (R1, R2) and (F1, F2) will be similar but it is possible for the note being sung to be above the value of R1 which would not therefore produce a peak F1.

Being able to modify the vocal tract resonances leads to the singing practice of "resonance tuning" where a singer can adjust the articulators so the tract resonances (R1 and/or R2) match a harmonic in the glottal wave thereby increasing the loudness of the output sound. Similar effects can be achieved using the practice of "vowel modification" (Bozeman 2014) where changes in the values of R1 and R2 result from changing the perception of the vowel being sung. However, fine tuning of the resonances can be realised without necessitating substantial modification of the vowel. Vowel recognition relates to the relative positioning of the formant peaks, not their absolute values. Vowels are also context dependent; the perceived vowel may change depending on spoken language of the listener and the consonant/vowel combinations (Gregg and Scherer 2006).

2.2 Key aspects of choral singing

Ternström (2003) summarised research on choirs which included studies on intonation accuracy, the dynamic range of choir singing, the influence of singer positioning in a choir and the impact of the acoustic environment. A review of research on singing ensembles by Daffern and D'Amario (2022) encompassed intonation, vibrato, sound intensity and timbre. The bulk of the previous research has used isolated notes, chords, phonemes and durations as the inputs. Comparatively little investigation has been undertaken considering the influence of the compositional choices on the acoustic results.

How each piece of research contributes to a holistic picture of the multi-dimensional nature of choral singing can be viewed conceptually as illustrated in Figure 2.2. Starting with a defined musical composition each singer's contribution is determined by their individual vocal technique and conditioned by environmental factors such as room acoustics, singer positioning and electronic processing. A feedback mechanism may also be an integral part of the system whereby singers adjust their contribution in real time depending on the output they hear. Key outputs for choral singing are intensity (loudness), intonation (tuning) and timbre. More than one of the selected outputs may be affected concurrently by a change in a variable. The model is time dependent but for ease of analysis sound samples can be selected which have reached stable conditions and can be considered to be essentially in a "steady state". For comparative purposes methods of quantifying the outputs need to be specified, whether as absolute acoustic metrics or qualitative descriptions.

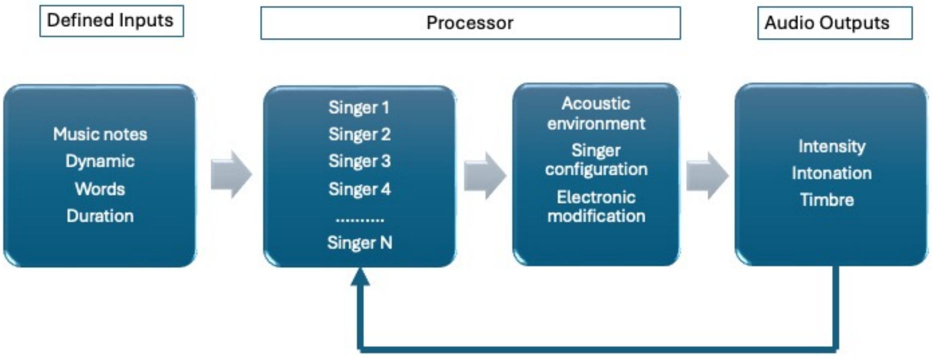


Figure 2.2: Conceptual model for ensemble singing

For Barbershop singing key outputs are accuracy of intonation, intensity and timbre. In the following sections previous research which relates specifically to these key objectives is described in detail.

2.2.1 Intonation

When performing a piece of music maintaining the correct pitch and tonal centre (horizontal tuning) is important for both solo singers and choirs. Choral singers have the additional demands of matching the tuning within a voice part and between the voice parts (vertical/ interval tuning). In music there is no absolute measure of accuracy of tuning in a choir with equal temperament, just intonation and Pythagorean tuning being options.

Western musical scales are primarily based on octaves whose frequency ratios are 2:1, with each octave divided into 12 semitones. In equal tempered tuning the frequency ratio between each pair of adjacent semitones is constant ($r = \sqrt[12]{2} = 1.0595$). As all semitones and tones have the same frequency ratios equal temperament has the advantage the music can be played in all keys but has the disadvantage the intervals between notes are not precise enough for singing chords requiring matching harmonics. An alternative way of dividing the octave is to use just intonation which is based on the intervals making up the major triads: the octave (2:1), the perfect fifth (3:2) and the major third (5:4). A diatonic major scale can be constructed in just intonation using three major triads (Appendix B). However a scale constructed this way results in two different sized whole tones which makes it less suitable for melody lines. A third scale can be constructed (Pythagorean tuning) based on repeatedly using the frequency ratio of the fifth (3:2) then transposing in octaves so all the notes are in the same octave. The scale using Pythagorean tuning has perfect octaves and perfect fifths, very sharp major thirds, very sharp major sixths and sevenths. In summary there is no absolute musical scale which is perfect for harmony singing and some degree of compromise is required. Richards (2001) opines Barbershop singers probably use a different scale for singing melody than they use for harmony.

Research has been undertaken on the tuning scales used by singers. Vurma and Ross (2006) investigated professional singers performing three isolated intervals: minor second, tritone and fifth and found the singers intonation was very close (20 cents or less) to equally tempered tuning, but with a tendency to expand the wider fifth and tritone intervals and compress the minor second interval slightly. Sundberg (2019) analysed the intonation accuracy of solo singers performing in a musical context (commercial operatic recordings) and found a much greater deviation (from 42 cents sharp to 42 cents flat) in equal temperament. Lottermoser and F. Meyer (1960) studied the interval sizes between the notes in chords sung by three choirs. They found the mean deviation from pure intonation varied from 2 to 60 cents but was typically in the range 20 to 30 cents. They found the choirs tended to make major thirds rather wide and minor thirds quite narrow whereas octaves and fifths were sung very close to just intonation.

The aim of the harmony parts in Barbershop singing is to tune with the specific objective of aligning the harmonics in the individual voices. Hagerman B and Sundberg J (1980) investigated the intonation of Barbershop quartets and found that accuracy in phonation frequency was very good and did not seem to depend on vowel. The absolute magnitudes of the standard deviations of F_0 ranged from 4 to 17 cents. They noted the singers found it is easier to tune intervals which shared many harmonics than intervals with few common harmonics. They also observed major and minor 3rd intervals deviated from the values according to just or Pythagorean intonation but without giving rise to beats. Howard (2007) investigated the problem of pitch drift when using just intonation which can develop through the course of a performance. It can result from resolution around the circle of fifths or key modulation.

There are a number of different methods available for measuring accuracy of intonation. Devaney et al. (2012) developed an Automatic Music Performance Analysis and Comparison Toolkit (AMPACT), which is a MATLAB package for determining the performance data of timing, pitch, and dynamics from existing recordings. From analysing four three part vocal groups they found the groups avoided pitch drift by tuning to equal temperament. They also found there was variation in tuning major and minor third intervals where just intonation, as well as equal temperament and Pythagorean tuning were used. Intonation accuracy of live performers was investigated by D'Amario and Daffern (2017) by analysing data captured using electrolaryngographs. They observed individuals in a singing quintet tended towards equal temperament although harmonic major thirds were found to be tuned closer to just intonation while minor thirds remained closer to equal temperament.

In summary, choosing a reference framework for quantifying the intonation accuracy of choral ensembles is difficult because equally valid alternative tunings are available depending on the acoustic goal. Equal temperament is often utilised to avoid pitch drift but intervals such as major and minor thirds appear to be subject to singer preference. For Barbershop singing the acoustic objective of Lock is determined by the precision of tuning in just intonation between the individual voice parts and this goal underpins the assessment of intonation accuracy adopted in this research.

2.2.2 Intensity

Vocal intensity is a key output for individual singers and for an ensemble. The dynamic range of a singer can be represented by a Voice Range Profile (VRP) which plots the vocal intensity range of a singer against fundamental frequency. Titze and Maxfield (2017) applied the concept of a VRP to a choir to investigate the acoustic factors affecting the dynamic range. Using six different levels of choir dynamics (*pp, p, mp, mf, f, ff*) over two octaves of fundamental frequency in a choir section they concluded overall choir size had negligible effect on the dynamic range. Their conclusions were based on the assumption that the individual sound intensities of singers are incoherent and can be added linearly. For a significant number of participating singers this assumption is valid as the random phase differences between singers will average out superposition effects. However, for a small number of singers such as a quartet, phase differences between the singers could have a discernible impact on the ensemble's net acoustic output and should be considered when interpreting research results with small groups.

Singers can dramatically boost their vocal intensity by modifying the shape of their vocal tracts to move its resonant frequencies closer to a harmonic in the note being sung (Henrich, J. Smith, et al. 2011). The gain in amplitude of that harmonic enables the singer to deliver more acoustic power without requiring an increased airflow, thereby improving acoustic efficiency. Kalin (2005) investigated this effect with a Barbershop quartet, describing it as "formant frequency adjustment". He determined each singer's resonant frequencies using inverse filtering, although found it difficult to establish the formants for the tenor when singing falsetto. He concluded the singers intentionally "strived to separate their formants" even when singing the same vowel. Achieving this would equate to "vowel modification" which would enhance the power produced but at the expense of lyric uniformity between the voice parts (Bozeman 2014).

For singers in an environment away from reflective surfaces the sound intensity generated by an ensemble can be quantified by measuring the Sound Pressure Level (SPL). It is noted that the SPL level is an objective acoustic measurement of the physical pressure at a point in space whereas loudness is a perceptual quality affected by the frequency, duration, and other factors.

2.2.3 Timbre

The timbre of an ensemble is not a simple superposition of the timbres of the individual singers in a group. Performers do modify their usual vocal techniques when singing with others. Goodwin (1980) observed solo singers adjust both their individual intensity and vibrato when singing in a group as they perceive it as helping the blend. Ternström (1991) investigated the physical and acoustic factors which affected how a singer contributes to the ensemble sound. From using two acoustic signals, a singer's own voice (the feedback) and the rest of the choir (the reference) intonation errors were found to increase when there were large level differences between the feedback and the reference and the magnitude of the errors was also indirectly related to room acoustics. Daugherty (1999) used different choir spacing and formations and concluded there were clear preferences in choral sound for certain performer spacings for both singers and auditors. These results were supported by Aspaas et al. (2004) and separately by Basinger (2006) who found choral formation, both in terms of voice matching and placement, was an important factor "contributing to improved choral blend" as it affected a choir member's ability to hear their own voice and the singers in different vocal sections.

Not only is it important for a singer to be able to hear their own voice relative to the other singers but Marshall and J. Meyer (1985) identified *how* singers hear their own voice is also significant. Audio feedback to a singer can arrive in two ways in a reverberating room; early reflections which arrive in less than 100 msec and the reverberation sound resulting from multiple reflections, diminishing over time. The loudness of the reverberations was most important to singers. The reverberation time was found to be of little significance. Irrespective of the reverberation time the singers preferred early reflections arriving within a time span of 15-35 msec.

As explained in the introduction, timbre is a difficult attribute to quantify objectively because it combines many acoustic features and lacks an absolute definition. Despite this limitation its strong association with the acoustic energy distribution in the frequency domain suggests metrics which categorise the shape of the frequency spectrum can be adopted for differentiating the timbre of sounds. As such any changes which affect intonation accuracy and sound intensity will concurrently have an impact on timbre.

2.3 Summary

A substantial body of research into the solo voice now exists which informs our understanding of its behaviour and management. In comparison ensemble singing, which is a more complex multidimensional activity, has not received as much research attention as the solo voice. To render the research task more tractable it is necessary to be selective when investigating key outputs. The research to date on choral ensembles has established:

- voice parts adopt different tuning references depending on the note interval being sung
- singers participating in group singing modify their performances depending on what they can hear of their own voice relative to what they hear of the rest of the ensemble
- singers adjust their sound level and voice usage depending on the acoustics of their environment

It is necessary to be cognisant of the differences between solo and group singing when interpreting the results of ensemble research.

Chapter 3

Computer modelling of voices

Mathematical modelling is an effective way of investigating multivariable systems when they cannot be extensively evaluated using live participants. In voice science a substantial volume of work has been done on modelling the solo voice, for both speech and singing, but comparatively little work has been done on modelling combinations of voices.

In Chapter 2, the model for a single voice was described which is based on a source/filter concept as illustrated in Figure 3.1. The model has three principal stages: a slowly varying flow of air from the lungs causes vibration of the vocal folds, which generates a quasi-periodic glottal wave, which is then

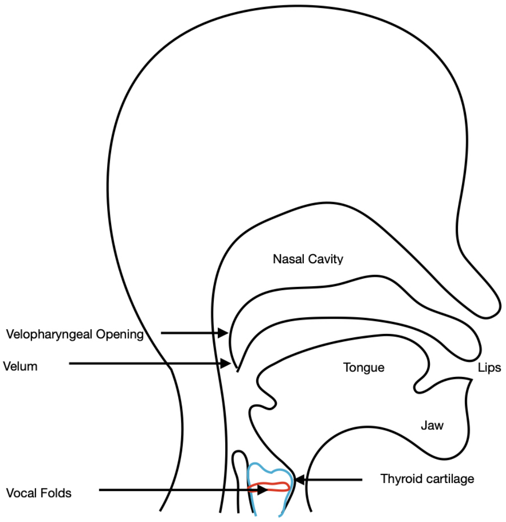


Figure 3.1: Source/Filter Model for voice production

filtered through spectral conditioning by the oral and nasal tracts before the production of a radiated wave which is determined by the acoustic impedance at the lips and nose. When the velopharyngeal opening is closed all the air and sound are directed through the mouth.

If the vocal tract configuration is maintained then an output wave y(t) in the time-domain can be expressed as the convolution of the glottal input wave x(t) and the impulse response of the system h(t) as:

$$y(t) = h(t) * x(t) \tag{3.1}$$

Equation (3.1) can also be expressed in the frequency domain as:

$$Y(\omega) = H(\omega) * X(\omega) \tag{3.2}$$

The filter H(ω) results from the product of the vocal tract transfer function T(ω) and the radiation characteristics at the mouth and nose R(ω).

$$H(\omega) = T(\omega) * R(\omega) \tag{3.3}$$

In its simplest form the system is treated as linear with interaction between the glottal source and vocal tract assumed to be negligible. For the vowel sounds analysed in this research the model is further simplified by assuming the velopharyngeal opening is closed so the voiced sound is non-nasal. The radiation impedance at the lips is approximated by a differentiating filter which is considered to be valid for frequencies up to 5 kHz. (Arnela et al. 2013)

3.1 Modelling the solo voice

3.1.1 Glottal source wave

Alternative formulations for the glottal flow and its derivative were reviewed by Doval et al. (2006) who demonstrated that the most commonly used glottal flow models (GFMs) can be described by "a set of 5 time domain parameters; the fundamental period, the maximum excitation, the open quotient, the asymmetry coefficient and the return phase coefficient. All the GFMs are equivalent with regard to the first three parameters". They compared the spectral properties of the different GFMs. The simplest model mathematically is the KLGLOTT88 model (D. Klatt and L. Klatt 1990) which only has 4 degrees of freedom as the asymmetry coefficient is fixed. A key parameter when modelling the glottal wave is the glottal open quotient O_q which Henrich, d'Alessandro, et al. (2005) found to depend on the laryngeal mechanism. For male singers the open quotient ranges from 0.3 to 0.8 for the chest, modal and head register and from 0.5 to 0.95 for falsetto. A frequent assumption in voice research is that the vocal tract filter and lip radiation filter are linear and time invariant so they can be commuted. The glottal source wave and the radiation filter can then be combined to form a "derivative glottal wave".

3.1.2 Vocal tract resonances

The vocal tract can be represented by a set of digital resonators connected in series or parallel. D. H. Klatt (1980) considered five resonators to be sufficient for simulating a vocal tract with a length of about 17 cm which is the length of a typical male vocal tract. J. O. Smith (2007) used three biquadratic filters in parallel to avoid noise quantisation and Holmes (1983) also expresses a strong preference for parallel synthesisers. Selecting the resonant frequency and bandwidth for each resonator to represent different vowel sounds has been the subject of several investigations. Early work by Peterson and Barney (1952) identified the first, second and third formant frequencies for sustained spoken vowels. Figure 3.2 illustrates the range of first and second formant frequencies for a given vowel as identified by listeners. It is evident that the perception of a vowel cannot be precisely classified by fixed formant values.

3.1.3 Lip radiation

Several mathematical models have been proposed to represent the radiation at the lips. A piston set in a spherical baffle (Morse et al. 1970) is the most physically representative but computationally complex. A simpler approximation, which is the commonly adopted model, assumes the sound radiates like a circular piston in an infinite plane baffle which can be represented by a first-order time derivative (Airaksinen et al. 2014). However Titze and Palaparthi (2018) demonstrated deviations from this model can occur at higher frequencies with harmonics significantly above 1 kHz being affected for larger mouth openings. The deviation from the simpler models was analysed in more detail by Arnela et al. (2013) using finite element modelling to investigate the effects of head and lip geometry on acoustic radiation. They concluded the piston in a spherical baffle model is valid up to 5 kHz, but in the 5-10 kHz range the lips play a significant part in the overall transfer function.

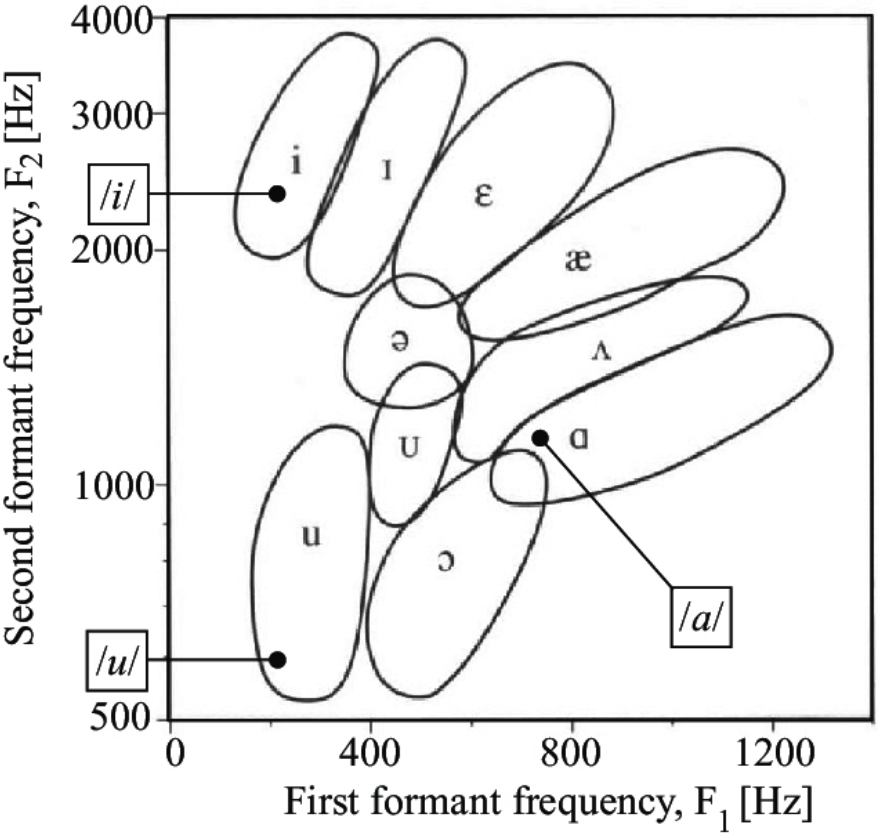


Figure 3.2: Formant frequency ranges for different vowels reproduced from Peterson and Barney (1952)

3.1.4 Nasalisation

The velopharyngeal valve is a structure which comprises the soft palate and walls of the throat and separates the mouth from the nose. It is necessary for producing most speech sounds. When the velopharyngeal valve is closed air and sound are directed through the mouth. A velopharyngeal opening (VPO) arises when the soft palate relaxes and airflow and sound can then pass through both the nasal and oral cavities (Sundberg 1987).

Whether singers benefit from having a VPO has been the subject of some debate. Gill et al. (2020) investigated the acoustic consequences of a large, a narrow, and no VPO and concluded a wide VPO seems disadvantageous as it decreases the overall sound level and produces a nasalized vowel quality. No VPO appears to be disadvantageous as it reduces the intensity of higher partials in the area of the singers' formant cluster. In contrast a slight VPO can be beneficial, since it offers singers the possibility of increasing the levels of the higher spectrum partials without requiring increase of vocal loudness. Havel et al. (2023) investigated the effect of a VPO on the vocal tract transfer function and concluded a VPO introduces a dip in the transfer function at the main resonance of the nasal tract.

3.1.5 Source/Filter interaction

The source/filter model generally assumes the vocal fold vibrations are independent of the vocal tract with the consequence that possible enhancements are not taken into account. Titze and Story (1997) conducted a theoretical study of possible interactions of the voice source with the lower vocal tract and concluded the inertance of the vocal tract can assist vocal fold vibration by lowering the oscillation threshold pressure. In a later study Titze (2004) concluded nonlinear source filter interaction can boost the power output of the voice when the fundamental frequency F_0 is below the first vocal tract resonant frequency R_1. When using computer modelling for comparative studies the absolute values of the phonation pressures are not used. Also when modelling voice production in a steady state it seems reasonable to assume the glottal source wave, which is an approximation in any event, has incorporated the effects of source/filter interaction.

3.2 Modelling an Ensemble

Very little previous research has been done on synthesising choral ensembles. Ternström et al. (1988) conducted preliminary studies by adding multiple representations of single voices. The work was mainly on unison sounds. Firstly they calculated a "fundamental frequency contour" for a choir singer based on measuring the random pitch variations of a single singer. A choir sound was then created by adding the contours of multiple voices using the same vowel. They were able to create effective ensemble simulations by building a chord and adding reverberation. They found a clear difference in the sound between one, two and three voices but hardly any difference between more than three voices.

Constructing a mathematical model of a singing ensemble requires the superposition of the mathematical representations of each voice. Combining two sounds p_1(t) and p_2(t) with the same frequency but different amplitudes and phases β, where $p_i = p_i cos(wt + \beta_i)$, i = 1,2, the time average total pressure squared $\langle p_t^2 \rangle$, which is proportional to the combined energy of the two sounds, is

$$\langle p_t^2 \rangle = \langle p_1^2 \rangle + \langle p_2^2 \rangle + 2\langle p_1 p_2 \rangle cos(\beta_1 - \beta_2) \tag{3.4}$$

For two sources of the same strength with the phase difference varying from π to 0 the total energy will range from 0 to 4 x $\langle p_1^2 \rangle$. With N sources (including reverberations) the sounds will have random relative phases, and as N increases the inner product terms will average out. The sources can then be added as scalar quantities on a linear energy (pressure squared) basis (Bies et al. 2023) and the limiting form of equation (3.4) becomes

$$\langle p_t^2 \rangle = \langle p_1^2 \rangle + \langle p_2^2 \rangle + ... + \langle p_N^2 \rangle \tag{3.5}$$

The above analysis is for pure tones but the same logic can be extended to the superposition of complex tones which are individually represented by Fourier series in the form $p_i = \sum_{m=1}^{M} A_m cos(wt + \varphi_m)$ where A_m is the amplitude and φ_m is the phase shift of the m^{th} harmonic. When the number of sound sources is small averaging out the inner product terms will not apply and the phase differences between the harmonics in the contributing sources will have a perceptible effect (Andersen and Jensen 2004). In addition to phase differences between the matching harmonics of the singers additional phase shifts are introduced by the asynchronicity of the individual voices. For Barbershop singing, which by definition aims to accurately match the harmonics in the contributing voices, phase differences could be significant, especially when the number of singers is small. For a quartet, where only 2 or 3 voices may align on specific harmonics, the impact of phase differences should be considered.

3.3 Summary

Mathematical modelling offers an effective way of investigating the impact changing key parameters has on the outputs of a multivariable system. Of necessity, simplifying assumptions are made to render the analysis and interpretation more tractable. To represent a vocal ensemble by using a combination of source/filter models some key choices have to be made; the glottal flow/ derivative model to use, the resonance frequencies and bandwidths of the vocal tract filters, the extent (if any) of nazalization, the extent of vibrato, the representation of lip radiation, the interaction between the glottal source and the vocal tract, and the phase differences between the harmonics and the voice parts.

Chapter 4

Perception of musical sounds

Musical psychoacoustics is the study of the psychological response of humans to the physical properties of acoustic stimuli. Rasch and Plomp (1999) considered the most important elements of musical psychoacoustics to be the subjective properties of musical tones (pitch, loudness, timbre) and the phenomena which result when several tones occur simultaneously. Much of the early research on perception used pure tones played separately or simultaneously. Pitch relates to the frequency of a simple tone and the fundamental frequency of a complex tone. Loudness relates to intensity, usually expressed as a sound pressure level (SPL) in dB. When tones (simple and complex) occur simultaneously small frequency differences between the tones partials can create beats or roughness in the combined sound (R. Plomp and Levelt 1965). Ensemble singing, by its nature, is the aggregate result of superimposing multiple complex tones and the frequency differences between the partials will affect the perception of the resulting sound.

While the relationships between pitch and fundamental frequency and between loudness and sound pressure level (SPL) are relatively straightforward, at least for simple tones, the relationship between timbre and physical acoustic metrics is more complex. The multi-dimensional nature of timbre necessitates limiting the number of features to be investigated at any one time to be able to isolate key causal agents. In this Chapter acoustic phenomena which are particularly significant to the genre of Barbershop singing are described in some detail.

4.1 Key aspects of the hearing mechanism

To give context to the following discussions about the perception of ensemble sounds a brief description of a key component of the hearing mechanism is presented (Howard and Angus 2017). It is primarily a result of the ear's capacity to function as a spectral analyser that we are able to differentiate the acoustic properties of complex sounds. The cochlea in the inner ear is divided into two parts along its entire length by the basilar membrane. Each point along the length of this membrane vibrates with maximum amplitude at a specific frequency so any periodic sound is resolved into its frequency components. For unimpaired human hearing the discernible frequency range is from 20 Hz to 20,000 Hz. The smallest interval between two frequencies which can be distinguished, if two tones are played *separately*, is the Just Noticeable Difference (JND) and is approximately 1/12 of a semitone. When two tones (or partials) are played *simultaneously* the "critical bandwidth" is the smallest frequency difference between the two tones such that each can be heard separately. As the two tones move together within the critical bandwidth the displacements of the basilar membrane caused by the individual tones increasingly overlap and the sound becomes "rough". As the frequency difference is reduced further beats are heard until the two sounds fuse into unison when the frequencies match. Critical bands within the ear are not fixed areas but are created during the experience of sound. Any audible sound can create a critical band centred on it. The critical bands are wider for higher frequencies than for lower and are generally in the range of two to four semitones wide. The magnitude of the critical bandwidth is significant as it relates to our experience of consonance and dissonance which are discussed in more detail below.

4.2 Perceptual features of single, complex and simultaneous tones

4.2.1 Loudness

The physical correlate of loudness is intensity and several loudness scales have been proposed. The phon scale has loudness levels expressed in dB where the loudness level of a sound in phons at different frequencies is equal to the sound pressure of a 1000 Hz tone assessed as having the same loudness in comparison tests (Bies et al. 2023). When considering the relationship between sound pressure level (SPL) and loudness a distinction must be made between sounds with all the spectral energy concentrated in one critical band and sounds where the spectral energy is spread over more than one critical band. If all the sound energy is limited to one critical band the loudness increases monotonically with intensity. However Zwicker et al. (1957) demonstrated when sound energy is distributed across more than one critical band the total loudness is greater than when the same amount of energy is concentrated within one critical band. They found that the loudness perception of complex tones depended on the frequency spacing between the components and they also showed that uniform spacing produces greater loudness than non-uniform spacing. They calculated the audibility by assessing each tone component in a critical band one by one and summing the total. An alternative to summation is to assess the total tonal content holistically where the identified tones are weighted together. This is the methodology used by Nyman B (2025) to study the loudness perception of sustained complex tones with an increasing number of harmonics as well as non-harmonic tone components. The results showed a positive linear relationship between perceived loudness and increasing number of harmonics, even with decreasing level of amplitude of harmonic (-6 dB/oct). Moving the second harmonic out of tune decreased the perceived loudness. Moving the third harmonic out of tune did not affect loudness perception. They found the perceived loudness increased more than the spectral summation method indicated and concluded the subjective estimation of loudness of complex tones should be made holistically. The implication of this research is that for the complex harmonic composition of sustained sung chords, simply measuring the total SPL on its own is not a sufficient measure to define the perceived loudness. How the acoustic energy in a chord is distributed between the harmonics will also influence the result.

4.2.2 Accuracy of intonation

In Chapter 2 alternative tuning systems were considered as the reference for being “in tune”, with just intonation being adopted as the benchmark for Barbershop singing. There is a large body of research that considers the perception of musical sounds and classification of in-tune and pitch matching. Sundberg (2019) described tests undertaken with six experienced professional music listeners who were asked to judge the accuracy of 10 commercial recordings of Franz Schubert’s Ave Maria. The judges were provided with the score and circled all tones they perceived to be out of tune. In most cases the intonation values perceived as in tune by all the experts lay within a band of ± 10 cents but the results showed considerable scatter with some tones which none of the judges considered out of tune varying between 35 cents sharp and 20 cents flat. Vurma and Ross (2006) analyzed listener accuracy of pitch perception. 13 professional singers performed a vocal exercise consisting of three ascending and descending melodic intervals (minor second, tritone, and perfect fifth). Good correspondence was found between interval tuning and the listeners’ responses with the data suggesting that melodic intervals may be 20 to 25 cents out of tune and still be estimated as correctly tuned by expert listeners. These conclusions apply when the different voice parts are singing the same vowels. If two or more voice parts in a quartet simultaneously sing different vowels different pitches are perceived, even though the fundamental frequencies being sung are maintained (Howard 2023). These intrinsic pitch variations are a result of the different vocal tract filters and the singers need to make adjustments from just intonation to maintain the perception of accurate intonation.

4.2.3 Masking

An important psychoacoustic phenomenon which can affect the perceived timbre of a sound is masking which occurs when two tones sound together and one tone (the masker) makes the other harder to perceive. This effect essentially equates to a shift in the hearing threshold for the masked tone (Sundberg 1987). Much of the research on masking has used separate pure tones. However, with complex sounds the harmonics within the sound can mask each other, which makes it more difficult to differentiate between the harmonic which is the masker and that which is being masked.

Sundberg and Gauffin (1974) developed a simple graphical method for predicting the harmonic masking effects in a vowel spectrum for a single voice. Several harmonics above the peak harmonic would not be audible as they fall below the masked threshold and consequently do not contribute to the perceived timbre of the sound. This phenomenon raises the question how much masking occurs between the individual voices in an ensemble. Applying Sundberg's methodology to an ensemble would suggest the higher harmonics need to reinforce each other to be audible above a masking threshold. Masking could also overshadow the contribution combination tones may make to the perceived timbre of an ensemble. The potential effect of masking highlights the importance of the reinforcement of harmonics between voices to the timbre of an ensemble's sound.

4.2.4 Combination tones

Combination tones are an acoustic phenomenon where additional tones can be created by pure tones of different frequencies being played simultaneously at a sufficient intensity level. Helmholtz (1877) identified two types of combination tones; differential tones, whose frequency is the difference between the frequency of the generating tones and summation tones, which are the sum of the generating frequencies. Helmholtz stated "summation tones are difficult to hear because they are subjected to the masking effect of the primary tones whereas most of the difference tones are not masked because they are lower than the primary tones". Ever since they were identified by the violinist Tartini in

1754, there has been debate about the origin of combination tones; do they exist outside the ear, are they produced in the ear by non-linearities in the hearing mechanism or are they purely psychological (White and Grieshaber 1980; Lohri et al. 2011). In any event, whether they exist in the environment or not, they can be perceived.

An explanation for combination tones was given by Rasch and Plomp (1999) predicated on a general transmission function based on two pure tones with the same amplitude and in phase:

$$p(t) = sin(2\pi * ft) + sin(2\pi * gt)$$

Based on experiments Rasch and Plomp (1999) found the most important combination tone frequencies are the difference tone (f - g) Hz, the second order difference tone (2f - g) Hz and the third order difference tone (3f - 2g) Hz. They found the summation tones of (f + g) Hz and (2f + g) Hz are seldom heard, even when in the audible range. Goldstein J L (1970) found the intensities of combination tones of the type f - k(g -f) Hz were very dependent on the relative intensities and phase relations of the tones. The research on combination tones has largely used pure tones and White and Grieshaber (1980) consider investigation of combination tones between complex sounds is needed.

On the basis of the above observations the combination tone most likely to be audible in Barbershop singing is the "undertone" (a frequency below the notes being sung by the ensemble) which is in an octave relationship with the root of the chord being sung.

For example for a major triad chord with 10th voicing the voice part frequency ratios bass:baritone:lead:tenor are 2:3:4:5. For a chord root frequency of 2f the frequency combinations are:
(baritone - bass) = (3f - 2f) = f, the missing fundamental
(lead - baritone) = (4f - 3f) = f
(tenor - lead) = (5f - 4f) = f
(bass(2nd harmonic) - baritone) = (2(2f) - 3f) = f
(baritone(2nd harmonic) - tenor) = (2(3f) - 5f) = f

For a Barbershop 7th chord with medium voicing and the bass on the 5th the note ratios are 3:4:5:7. The frequency combinations are:

(baritone - bass) = (4f - 3f) = f, the missing fundamental
(lead - baritone) = (5f - 4f) = f
(bass(2nd harmonic) - lead) = (2(3f)-5f) = f

(baritone(2nd harmonic) - tenor) = (2(4f) - 7f) = f

Additional combinations of harmonics can be computed but apart from the undertone they will be difficult to discern because many coincide with the harmonics of the notes the quartet is singing. For combination tones to be audible the singers need to sing with relatively high intensity, be precisely tuned in just intonation and have their fundamental frequencies in phase. Even when these conditions are met the higher harmonics in the complex tones produced by singers will have very different amplitudes and are therefore unlikely to generate audible combination tones.

4.2.5 Consonance and dissonance

The combined sound of simultaneously occurring tones may subjectively be considered to be pleasant (consonant) or unpleasant/rough (dissonant). This perception has been labelled as tonal consonance by R. Plomp and Levelt (1965). When two simple tones are played with a frequency separation that is very small or greater than the critical bandwidth the tones sound consonant. Critical bands are wider for higher frequencies than for lower and are generally in the range of two to four semitones wide. At frequency separations within the critical bandwidth the sound will be dissonant. The greatest perceived dissonance occurs when the frequency separation is about a quarter of the critical bandwidth, which generally is a little less than a semitone (R. Plomp and Levelt 1965). A major third comprises four semitones so notes played simultaneously in a third generally occupy separate critical bands and will be considered consonant. Seconds, which fall within the same critical band, are considered dissonant. At low frequencies the critical bands span more than a third so thirds begin to lose their consonance.

For complex tones an assessment of the consonance of intervals can be derived from the simple tone combinations. The dissonance of all combinations of neighbouring partials are determined and added to give the total dissonance. The degree of consonance of the musical interval between two complex tones with a defined ratio between the fundamental frequencies is dependent on the simplicity of the ratio. Rasch and Plomp (1999) considered intervals with frequency ratios that can be expressed using small integer numbers (less than 6) to be relatively consonant because the most important components of the two tones are either widely apart or coincide. When the frequency ratios are less simple the difference in frequency between a number of the partials is small and these partials give rise to dissonance. Intervals with the number 7 in their frequency proportions (7/4, 7/5, ...) are on the borderline between consonance and dissonance. Bernini and Talamucci (2014) developed the method of R. Plomp and Levelt (1965) to assess the consonance of complex tones with harmonics of different intensity. Using their methodology they computed a relative ranking of the dissonance of different chord structures and inversions. In summary, even when the notes in a chord are accurately tuned in just intonation, certain chord structures will still be perceived as dissonant.

4.3 Perceptual features of Barbershop singing

In Barbershop singing defining perceptual descriptions are the terms "Lock", "Ring" and "Expanded Sound" which are used to describe desirable attributes of sung chords, particularly when they have reached a steady state. An explanation of these terms is given here together with an assessment of the current reasoning supporting these concepts. A primary objective of this research is to identify physical correlates for these terms and ways of measuring them.

A way of ranking Barbershop chords in terms of "lockability" was proposed by Richards (2001) who derived a "fidelity" rating based on the "combination tones" for each possible pair of harmonics in a chord using just intonation. When the difference between paired harmonics was zero, or a sum tone or difference tone supported one of the fundamental notes of the chord it was considered a "hit". Each time the pairing did not meet these requirements it was considered a "miss". The greater number of hits the higher the rating. A fidelity rating was then calculated using an empirical logarithmic formula based on the number of "hits" and "misses". Richards posited that the sense of "Lock" is less pronounced for chords with less consonant intervals, ie where the first occurrence of consonance of harmonics is higher up the harmonic stack. This approach to rating Lock assumes each harmonic in a voice has equal importance when in practice they will have different strengths, affected by the pitch (root frequency) of the chord and the vowel being sung. The method also assumes there is no phase differences between harmonics.

Titze and Story (1997) describe Ring as "the percept of a bell like constant background tone, which spectrally shows up as a prominence of acoustic energy in the 2500 - 3000 Hz range independent of vowel". Sundberg (1974) called this spectral peak the singer's formant. Ring has also been described as the "brightness" in a sound which has been shown to correlate with an increased power in the higher frequencies (Schubert and Wolfe 2006). Richards (2001) opined "voices with very little resonance generally do not ring chords very well". These assessments suggest Ring is associated with a concentration of acoustic energy in the higher frequencies and that Ring for an ensemble will be determined by the distribution of acoustic energy in the individual voices.

Richards (2001) explanation of Expanded Sound is that "the four notes in a chord create their own set of harmonics and each pair of harmonics creates two additional notes from the sum and difference frequencies. Each chord could create hundreds of combination tones. When these created frequencies are sorted we discover they form a new harmonic series of their own. A listener will often hear a chord's undertone frequency even though no one voice is producing the sound. Undertones are created as the result of combined singing voices". Unquestionably undertones below the fundamental frequencies of the individual voices can often be heard. Whether other combination tones contribute to the perception of sound expansion is debatable as each harmonic in a participating voice will have different strengths, affected by the fundamental frequency of the note and the vowel being sung and there will be phase differences between the voices. Consequently the potential for sound expansion will not be determined by chord structure alone.

4.4 Summary

Defining how sounds are perceived is a diverse, complex subject in its own right. Of necessity research must focus on selected features. In singing key differentiating attributes for a listener are pitch, loudness and timbre. For Barbershop singing the idiomatic descriptions of Lock, Ring and Expanded Sound are declared desirable properties, albeit currently lacking unambiguous definitions. In the past it has been postulated "combination tones" are the source of some of the perceived attributes, especially Expanded Sound. While combination tones may explain the perception of "undertones" in ensemble singing the different amplitudes and phases of the harmonics in the contributing voices and the effects of masking render it unlikely that combination tones can be the sole source of Expanded Sound. In this research alternative explanations for the phenomena are explored.

Chapter 5

Analysis of published Barbershop quartet recordings

Important acoustic characteristics for differentiating the sounds of sustained chords performed by a vocal ensemble are accuracy of tuning (absolute pitch and intervals between the voices), intensity (loudness) and timbre. Each of these characteristics is related to the distribution of the acoustic energy across the chord's harmonics in the frequency spectrum. When Barbershop chords are accurately tuned the fundamental frequencies of the contributing voice parts should be in small integer ratios to ensure the highest consonance. For many Barbershop chords the frequency ratios of the notes are precisely defined by perfect intervals between the voice parts in just intonation but for some chord structures there are alternative tunings which are equally valid. In this chapter investigations of some of the factors affecting these key attributes are described, based on an analysis of publicly available recordings of Barbershop quartets. The procedure adopted was approved by the University of York Physical Sciences Ethics Committee. (Approval reference Cookson310120).

5.1 Methodology

5.1.1 Database of quartet samples analysed

A database of 50 samples of Barbershop quartets was created using segments extracted from CDs and recordings of live performances available on the internet for thirteen different male Barbershop quartets; 48 of the samples were top-scoring quartets from international Barbershop competitions, chosen to represent good quality singing. Two additional samples which were significantly out of tune were selected, although one was later omitted as the vowel was found to change mid sample. This database is referred to hereafter as the "archive samples". The chord structures, chord root

Sample No	Voicing	Vowel	Root Hz	Freq Ratios	Centroid Hz	Spread Hz
1	BBS7th spread voicing	i	109	2:5:3:7	242	299
2	BBS7th medium voicing	ɔ	123	4:7:6:10	383	396
3	Major triad 12th voicing	e	184	2:5:4:6	1402	1203
4	BBS7th medium voicing	e	117	4:7:6:10	768	948
5	Major triad 10th voicing	a	172	2:3:4:5	1023	652
6	BBS7th close voicing	u	146	4:5:6:7	256	210
7	BBS7th med voicing Bass 5th	o	87	3:5:4:7	397	253
8	BBS7th close voicing	a	154	4:5:6:7	408	511
9	Major triad 10th voicing	ʌ	208	2:3:4:5	1976	1307
10	Major triad 10th voicing	ʌ	199	2:3:4:5	1174	583
11	BBS7th spread voicing	i	111	2:3:5:7	289	204
12	BBS7th medium voicing	ɔ	127	4:7:6:10	465	563
13	Minor 6th (alt)	a	195	6:9:10:14	1270	956
14	Major triad 12th voicing	e	145	2:4:5:6	1239	923
15	Major triad 2 octave voicing	u	84	2:5:6:8	243	87
16	Minor 6th (alt)	e	141	6:10:9:14	342	320
17	BBS7th close voicing	a	150	4:5:6:7	823	454
18	Minor 6th (alt) close	u	197	6:7:9:10	369	473
19	BBS7th med voicing Bass 5th	e	138	3:5:4:7	1301	1306
20	BBS7th med voicing Bass 5th	e	138	3:5:4:7	1028	757
21	BBS7th medium voicing	e	118	4:7:6:10	470	604
22	Major triad 12th voicing	e	187	2:5:4:6	868	490
23	Major triad 12th voicing	e	187	2:4:5:6	1301	807
24	BBS7th medium voicing	e	121	4:7:6:10	438	561
25	Minor 6th (alt)	a	201	6:9:10:14	1369	923
26	Major triad 12th voicing	e	149	2:4:5:6	1045	591
27	Minor 6th (alt)	a	196	6:9:10:14	1191	856
28	Major triad octave voicing	ɔ	183	4:5:6:8	933	623
29	Major triad 10th voicing	ɔ	181	2:3:4:5	1081	536
30	Major triad 10th voicing	e	172	2:3:4:5	1175	807
31	Major triad 2 octave voicing	e	86	2:5:6:8	294	163
32	Major triad octave voicing	i	123	4:5:6:8	342	407
33	BBS7th medium voicing	ı	130	3:4:5:7	150	1272
34	Major triad octave voicing	u	152	4:5:6:8	249	154
35	Major triad 10th voicing	e	122	2:4:3:5	1255	1076
36	Major triad 10th voicing	u	153	2:4:3:5	366	456
37	Major triad 12th voicing	ı	78	2:5:4:6	983	1247
38	Major triad octave voicing	e	129	4:6:5:8	573	356
39	Major triad 10th voicing	e	139	2:3:4:5	533	423
40	Major triad 10th voicing	o	154	2:3:4:5	952	1075
41	Minor triad 10th voicing	a	149	6:9:12:14	1045	703
42	Major triad octave voicing	e	178	4:5:6:8	596	495
43	BBS7th medium voicing	i	153	4:7:6:10	738	890
44	Major triad 10th voicing	a	178	2:3:4:5	1006	668
45	Major triad octave voicing	a	144	4:6:5:8	504	580
46	Major triad 12th voicing	ı	146	2:4:5:6	413	490
47	Major triad 10th voicing	ɔ	130	2:4:3:5	498	317
48	Major triad 10th voicing	a	87	2:3:4:5	302	203
49	Minor triad octave voicing	a	170	6:9:7:12	1094	1272
50	Major triad 10th voicing	a	174	2:3:4:5	1829	1432

Table 5.1: Details of archive database samples

frequency and vowels sung in these samples are shown in Table 5.1. The target frequency ratios used are those proposed by Richards (2001) to give maximum consonance based on the lowest possible combination of ratio numbers for the entire chord. The minor triad is more consonant when the sub-minor third (7:6) is used. The minor 7th chord contains two dyads of a perfect 5th offset from one another by a minor third (6:5). The minor 7th chord is the most consonant when the minor 7th interval dyad is tuned in the ratio 9:5. The Barbershop 7th chord (BBS7th) achieves its best consonance when the harmonic minor seventh interval is tuned as 7:4. The (alt) tuning is when the sub-minor third (6:7) is used in place of the classic (6:5) minor third.

5.1.2 Method of analysis

For the spectral analysis a segment of 32,768 samples was extracted from the audio file at a rate of 44,100 samples/sec giving a 743 msec duration and a frequency resolution of 1.35 Hz. (Using a power of 2 for the number of samples significantly improves the computational efficiency of the Fourier transform). The segment was taken from the middle of a sung chord to give a *steady state* thereby avoiding the transient effects of vocal onset and offset and minimising any singer adjustments made during the sample. A 'window', part Hamming for the first and last 10 percent and a value of 1 for the middle section, was applied prior to spectral analysis. The outputs were normalised relative to the strongest harmonic in the sample.

5.1.3 Metrics for evaluating the acoustic characteristics of samples

In Barbershop singing the Lead has the primary responsibility for singing the melody and maintaining the tonal centre. The other voice parts create the chords by singing harmony in tune with the Lead. The *intonation accuracy* achieved by a harmony part relative to the Lead's note can be measured by comparing the fundamental frequency being sung with the ideal target frequencies to create maximum consonance. The fundamental frequency being sung by a voice part was determined from the singer's harmonic peak in the frequency spectrum. For a harmony part singing at fundamental frequency F$_h$, the deviation 'c' cents of their actual frequency from the target frequency can be defined as:

$$c = 1200 * log_2 \left[\frac{F_h}{r * F_l} \right] \tag{5.1}$$

where 'r' is the ideal frequency ratio relative to the Lead fundamental frequency F$_l$.

The *intensity* and *timbre* of a chord in a steady state is reflected in the line spectrum and cumulative line spectrum in the frequency domain. The line spectrum displays the energy in each harmonic contributing to the sound. The cumulative line spectrum is a running total of the energy in each harmonic across a frequency range from 0% at 0 Hz up to 100% at a defined upper frequency. Fine frequency resolution is required in the line spectrum to accurately identify the spectral peaks. The lowest frequency sung by a male bass in Barbershop is around 87 Hz (F2) where a semitone interval is 5 Hz. At this pitch a frequency resolution of 1.5 Hz is 30 cents. The cumulative line spectrum gives a visual indication of the distribution of energy in frequency 'bins' which are defined discrete intervals along the frequency axis. Examples of these spectra for two different chord samples are given in Figure 5.1. The chord in Figure 5.1a was well tuned and shows precise steps in the cumulative line spectrum. In contrast the chord in Figure 5.1b was not stable which is reflected in the noisy spectrum and lack of precise steps in the cumulative line spectrum. To be able to compare the spectra of different chords the dB levels of the harmonics have been normalised relative to the peak harmonic in the chord and the energy distribution has been proportioned to give a total figure of 100 percent in the chord.

Metrics related to the position and shape of the spectra can be computed, which provide absolute acoustic measures suitable for differentiating between sounds (Giannakopoulos and Pikrakis 2014). Two key measures are the *spectral centroid* and the *spectral spread* defined as:

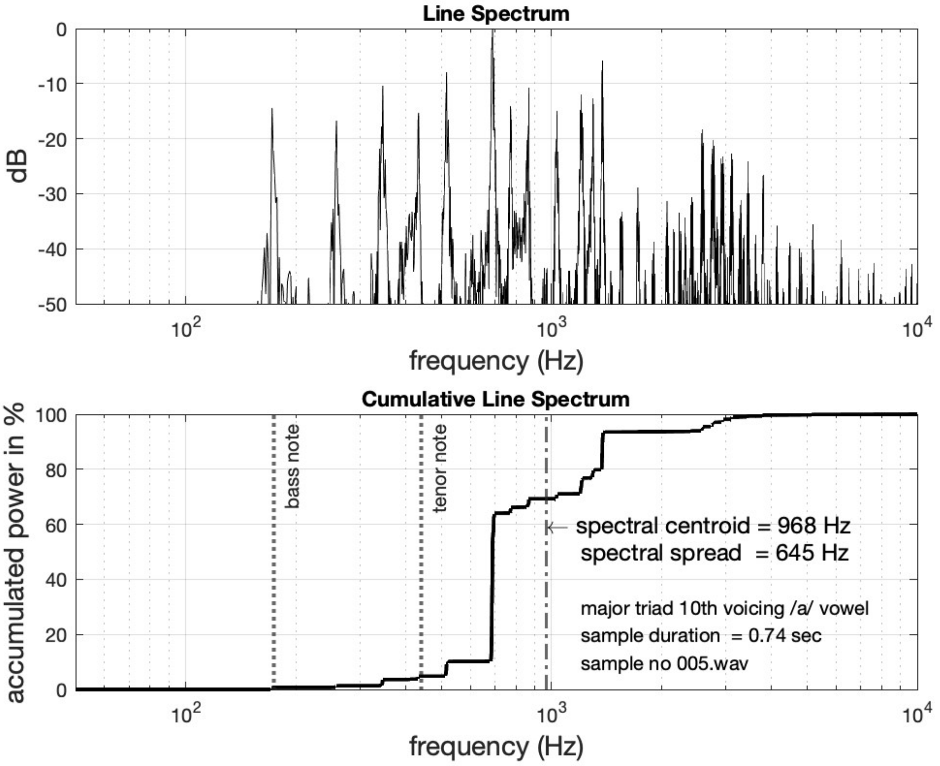
Spectral Centroid

$$f_{sc} = \frac{\sum_{k=1}^{N/2} A_k^2 * f_k}{\sum_{k=1}^{N/2} A_k^2} \tag{5.2}$$

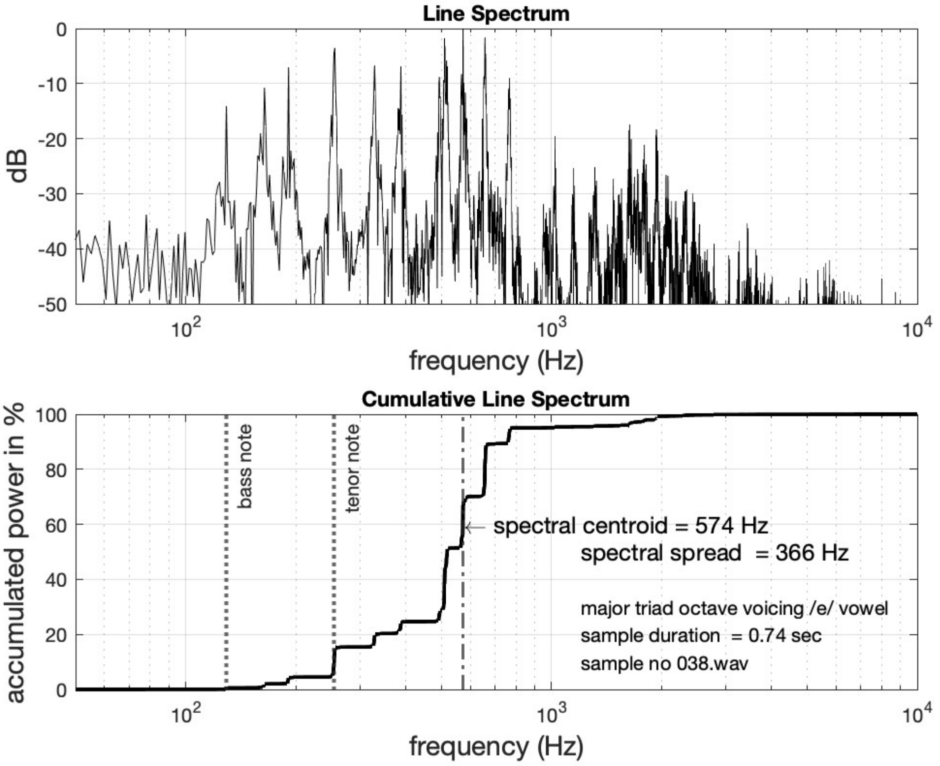
Spectral Spread

$$f_{ss} = \sqrt{\frac{\sum_{k=1}^{N/2} A_k^2 * (f_k - f_{sc})^2}{\sum_{k=1}^{N/2} A_k^2}} \tag{5.3}$$

where A_k is the absolute amplitude of the harmonic at frequency f_k and N is the number of samples used in the spectral analysis. Conceptually the spectral centroid can be thought of as the *centre of gravity* of the spectrum (Peeters 2004). The greater the amount of the acoustic energy in the higher harmonics the higher the spectral centroid, which has been correlated with the perceptual quality of brightness (Schubert and Wolfe 2006). The spectral spread is a measure of the distribution of the energy in the spectrum about the centroid and is an indication of the *bandwidth* of the sound. It is noted here, because of relevance to later discussion, that normalising the amplitude of the harmonics does not affect the values of the centroid and spectra, it is simply a scaling of the spectrum whereas



(a) sample 005



(b) sample 038

Figure 5.1: Illustrative examples of line spectrum and cumulative line spectrum of two different chord samples (a) 005 accurately tuned and (b) 038 poorly tuned

the *absolute* amplitude of the harmonics will affect the perceived intensity of a sound. The centroid and spread of a chord can be affected by the chord root frequency, the chord structure, the balance between the voices, the vowel being sung and the individual singer's vocal techniques all of which affect the strength of the different harmonics.

5.2 Results

The intonation accuracy of each voice part for every sample is plotted in Figure 5.2. The Lead note in the chord was always taken as the reference frequency, even if it had strayed from the specified pitch, and therefore had a deviation of 0 cents. The majority of the harmony parts lie within ±20 cents of their target frequency but with a few samples considerably outside this range (samples 11, 18, 38 and 41). The computed spectral centroid and spectral spread of each sample are given in Table 5.1 and plotted against the chord's root frequency in Figure 5.3 differentiated by vowel sung. The values of centroid range from 200 Hz to 1600 Hz, well above the fundamental frequencies for the singers which have a maximum of around 500 Hz. Notably the /u/ vowel centroids remain low for all chord root frequencies. The spread ranges from around 100 Hz to 1300 Hz and the /u/ values remain low across all root frequencies.

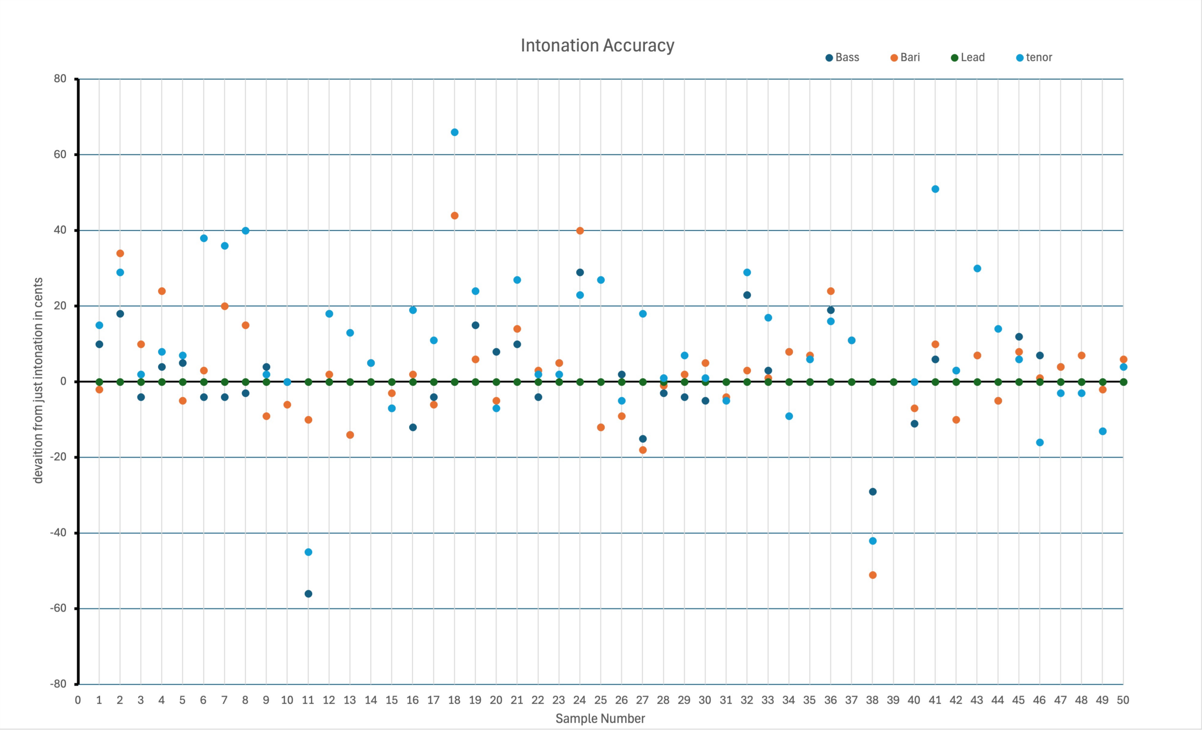
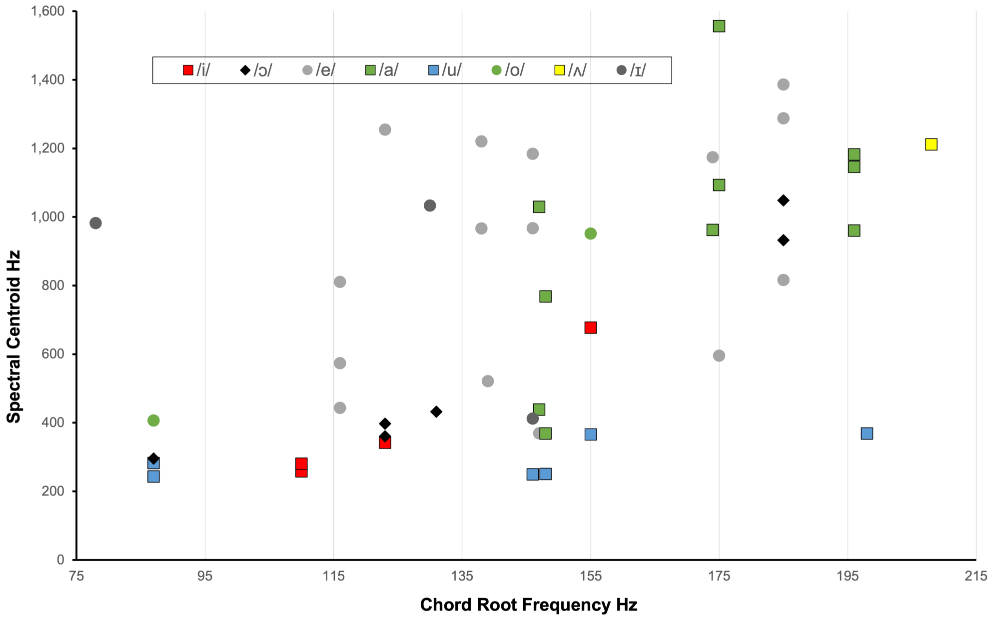
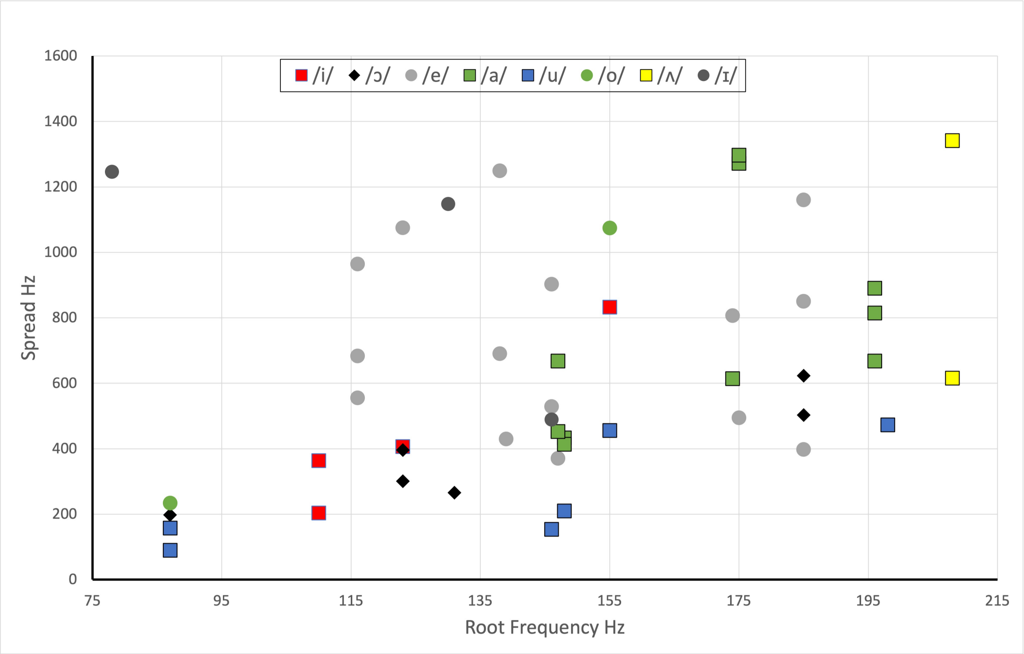


Figure 5.2: Intonation accuracy of each voice part for each database sample



(a) centroid



(b) spread

Figure 5.3: Acoustic metrics vs chord root frequency differentiated by vowel (a) centroid vs chord root frequency (b) spread against chord root frequency

5.3 Discussion

Overall the singers demonstrated a good degree of interval intonation accuracy with most being within ±20 cents of their target frequency. These results compare favourably with the interval sizes measured by Lottermoser and F. Meyer (1960) in chords sung by three choirs. They found the mean deviation from pure intonation varied from 2 to 60 cents but was typically in the range 20 to 30 cents. However some of the archive database samples showed significant deviation from the target frequency ratios. Sample 11 is a BBS7th with spread voicing (2:5:3:7). The phonation frequencies of the bass and lead are a perfect fifth at 110 Hz and 165 Hz but are well below the first formant of 300 Hz for an /i/ vowel with the result that the root and 5th were not precisely tuned. As a result accurately tuning the tenor note on the 7th would be difficult. Sample 18 is a minor 6th chord which does not have a high degree of consonance with frequency ratios 6:7:9:10. It will be more difficult to tune because of the lack of matching harmonics. The singers may also be aiming for frequency ratios which are different to the target ones assumed. Sample 38 is a major triad chord with an octave voicing. The frequency ratios are 4:6:5:8 with the baritone singing above the lead who is singing the third. It was intentionally selected as a sample to test the analysis methodology as it was audibly out of tune, as can be seen from the spectrum which is presented in Figure 5.1. Sample 41 is a minor triad with frequency ratios 6:9:12:14. The bass and baritone were within 10 cents of the target values but the tenor, on the minor third, was 50 cents above target frequency.

For the spectral metrics of centroid and spread clear trends are difficult to discern because of the number of different variables in the samples. The data set, which used 12 different chord structures, 12 different root frequencies and 7 different vowels, potentially has over 1000 combinations. Nevertheless the results for the 50 chosen samples demonstrated the substantial range of centroid and spread values which can be produced illustrating the variety in timbre which can be realised. In particular the vowel had a significant impact on the results. For the /u/ vowel both the centroid and the spread remained low (both around 300 Hz) across the full range of root frequencies. For the /i/ vowel both the centroid and spread are in the region of 300 Hz until the root frequency moves above 135 Hz when significantly higher values are produced. For the /a/ vowel above a root frequency of 135 Hz values of both the centroid and spread are significantly higher. For intermediate vowels like /e/ there is considerable scatter. Explanations for these different behaviours were investigated by developing and running a computer model of a Barbershop quartet which is the subject of the next chapter.

5.4 Limitations

There is no absolute measure for accurate tuning of note intervals in chords. The benchmark adopted here for Barbershop singing is that the frequency ratios between the voice parts should be small integer ratios to give maximum consonance, with the Lead note as the reference frequency. As the sound samples analysed were extracted from CDs and the internet it was not possible to identify the variables which affected the absolute intensity of the sound produced or to know the nature of any editing applied to the recordings. Only the quartet sound was available for analysis so the relative contribution of individual voices could not be isolated. Furthermore the acoustic environment of the recordings was not known which could add another variable to the resulting sound. To address these limitations laboratory tests with a live quartet under controlled conditions were carried out. These tests are the subject of Chapter 7.

5.5 Summary

Accuracy of intonation and the spectral values of centroid and spread can be useful absolute acoustic measures for differentiating between the sounds of Barbershop chords. An analysis of 50 samples taken from quartet recordings illustrated a wide range of values for these metrics can result from changing the chord structure, the chord root frequency and the vowel sung. To obtain more insight into the way these variables affect the acoustic characteristics of an ensemble a computer model of a Barbershop quartet was created to run parametric studies. This model is the subject of the next chapter.

Chapter 6

Computer Model for Synthesised Quartet

The analysis of 50 different published recordings of Barbershop quartets presented in the previous chapter demonstrated that a wide range of values for the selected acoustic metrics of accurate intonation, spectral centroid and spectral spread can result from using different chord structures, chord root frequencies and vowels. However it was not possible in the analysis to isolate the causal effects these different inputs had on the spectral metrics nor identify how the different characteristics of the contributing voices affected the ensemble's sound. To gain greater insight into these factors a computer model was written in Matlab (Mathworks 2021) which could be used for parametric studies. A synthesised quartet was created by combining four separate source/filter models of the form described in Chapter 3. The resulting output of the synthesised ensemble was analysed using the same methodology and metrics as used in Chapter 5 for the samples of published recordings. The defined inputs were chord structure, chord root frequency, vowel sung and the vocal characteristics of each singer. The model did not aim to be a lifelike simulation of a real quartet but it needed to be sufficiently representative to reflect the effect of changing key parameters. Features such as jitter, shimmer and room reverberation were not included. The computer model predictions were compared with the results from the equivalent archive samples described in Chapter 5.

6.1 Methodology

6.1.1 Modelling a single voice

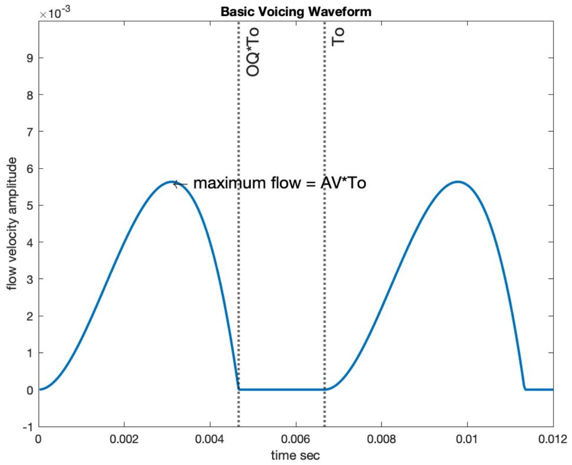
In much of the previous research on the solo voice the mathematical models used the glottal flow *derivative*, assuming it to be equivalent to a convolution of the glottal flow with a lip radiation filter. In the model used here the glottal flow and the lip radiation functions are kept separate to allow independent variation of their properties and to correspond more closely with the physical reality of the vocal mechanism.

For each voice in the quartet the glottal air flow $U_g(t)$ was based on the KLGLOTT88 model (D. Klatt and L. Klatt 1990). The simple form of the equations 6.1 in this model have the benefit of speed in computational iterations and it is simple to alter each singer's characteristics and vocal power without the resulting spectra differing significantly from other glottal flow models (Doval et al. 2006).

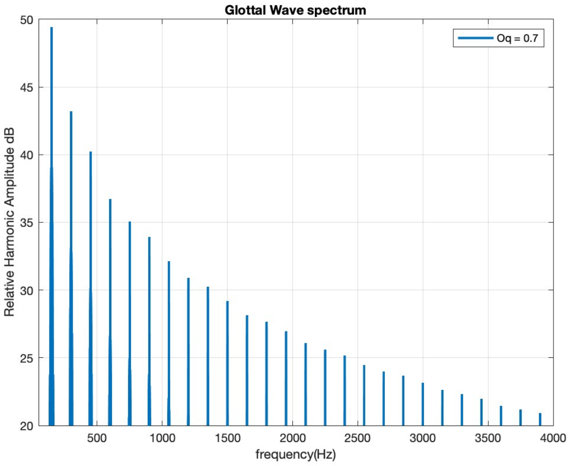
$$\begin{aligned} U_g(t) &= (at^2 - bt^3) \quad 0 \leq t \leq O_q T_0 \\ &= 0 \qquad\qquad\quad O_q T_0 \leq t \leq T_0 \end{aligned} \tag{6.1}$$

where a $= \frac{27}{4}\frac{AV}{O_q^2 T_0}$, b $= \frac{27}{4}\frac{AV}{O_q^3 T_0^2}$, T_0 is the period of the wave, the fundamental frequency $F_0 = 1/T_0$, the peak flow amplitude is $AV*T_0$, and O_q is the open quotient (the ratio of the time the vocal folds are open to the fundamental period of the wave). Graphical representations of the glottal flow and its associated spectrum for a fundamental frequency of 150 Hz with an open quotient of 0.7 are shown in Figures 6.1 (a) and (b). The spectral envelopes for different values of the open quotient O_q, with the same glottal power, are shown in Figure 6.1 (c). As O_q is increased, moving from a 'pressed' to a 'breathy' voice, the roll off rate (dB/octave) in the strength of the harmonics increases. As will be evident later the decrease in harmonic strengths can be particularly significant for vowels which have high formant frequencies.

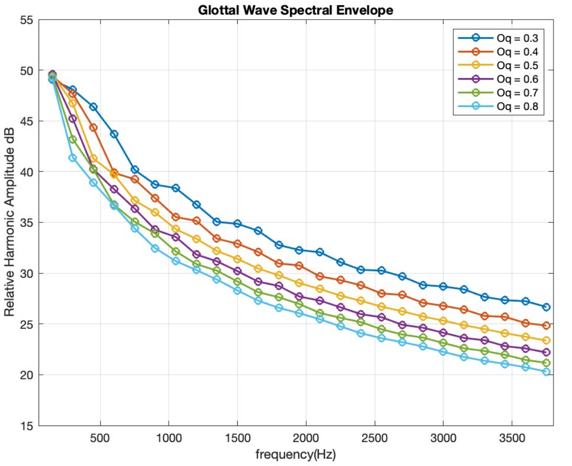
The glottal wave is passed through a filter bank comprising seven biquadratic filters in parallel which represent the resonances of the vocal tract. The poles of the vocal tract transfer function correspond to the vocal tract resonances (R1,R2,R3...) with their associated bandwidths (B1,B2, B3...). Different values of R_n and B_n are used to represent each vowel as shown in Table 6.1. The values are based on the figures D. H. Klatt (1980) estimated for the first three formants (F1, F2, F3) and bandwidths (B1, B2, B3). Sundberg (1988) observed the fourth formant frequency is generally in the region 2500 Hz to 4000 Hz and the fifth is in the vicinity of 3000 Hz to 4500 Hz. For the



(a) glottal flow waveform



(b) glottal spectrum



(c) glottal envelopes

Figure 6.1: KLGLOTT88 glottal flow model (a) waveform (b) spectrum (c) spectral envelope for different open quotients

model used here R4 and R5 were fixed at 3000 Hz and 3500 Hz respectively each with a bandwidth of 200 Hz. A 6th and 7th resonance were added at 5500 Hz and 6500 Hz to reduce abrupt cut off at higher frequencies. The vocal tract filter, when applied to the glottal flow wave, introduces different phases for each harmonic.

Vowel	Word	R1	R2	R3	B1	B2	B3
i	Heed	310	2020	2960	45	200	400
ɪ	Hid	400	1800	2570	50	100	140
e	Hay	480	1720	2520	70	100	200
ɛ	Head	530	1680	2500	60	90	200
æ	Hair	620	1660	2430	70	150	320
ɑ	Father	700	1220	2600	130	70	160
ɔ	Hall	600	990	2570	90	100	80
ʌ	Hut	620	1220	2550	80	50	140
o	Hoe	540	1100	2300	80	70	70
ʊ	Hood	450	1100	2350	80	100	80
u	Hoot	350	1250	2200	65	110	140
ə	Again	470	1270	1540	100	60	110
a	Had	660	1200	2550	100	70	200

Table 6.1: Resonant frequencies and bandwidths used for different vowels in the synthesized quartet model

It is important to note that in the analysis presented in this chapter the same filter parameters for a given vowel were used for each voice in the quartet and at all phonation frequencies. In reality, even when nominally singing the same vowel, the vocal tract resonances and bandwidths will differ between singers for physiological reasons, the pitch being sung and as a result of singers making adjustments to optimise their vocal output while performing (Henrich, J. Smith, et al. 2011). An example of the seven filters combined for an /a/ vowel is shown in Figure 6.2).

The modified flow is then passed through a lip radiation filter. Several mathematical models have been proposed to represent the radiation at the lips. A piston set in a spherical baffle (Morse et al. 1970) is the most physically representative but computationally complex. A simpler approximation, which is a commonly adopted model, assumes the sound radiates like a circular piston in an infinite plane baffle which can be represented by a first order time-derivative (Airaksinen et al. 2014) and is a reasonable approximation for frequencies up to 5 kHz. A parameter $\alpha = 0.99$ is included to give computational stability without significantly changing the effects.

$$R(z) = (1 - \alpha z^{-1}) \tag{6.2}$$

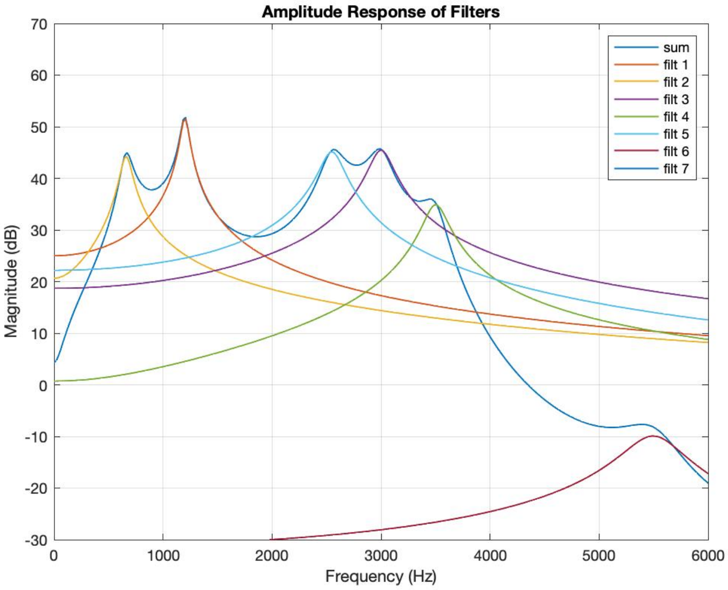


Figure 6.2: Vocal tract filter for an /a/ vowel

The resulting output spectrum for a single voice singing an /a/ vowel is shown in Figure 6.3 using the glottal wave input shown in Figure 6.1.

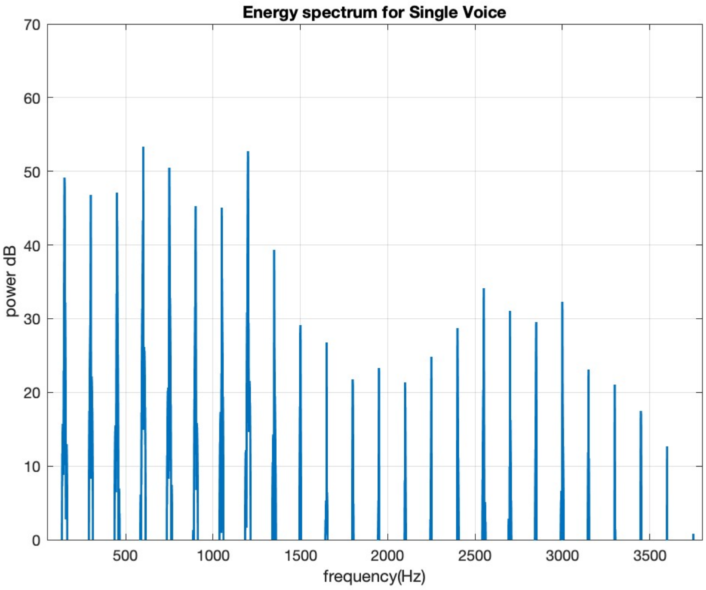


Figure 6.3: Output spectrum of single voice on an /a/ vowel

6.1.2 Superposition of four voices

For the KLGLOTT88 model the time averaged glottal source power can be shown to be proportional to $(AV * T_o)^2 * O_q$. It was assumed each singer in the quartet generated the same glottal power.

The fundamental frequencies of the four voices were selected so they were in just intonation, using the same frequency ratios as assumed for the archive samples shown in Table 5.1. The part Hamming window used in the archive analysis was applied to each computer generated voice prior to the spectral analysis. When combining the four voices they were treated as being incoherent, making no allowance for phase differences between the harmonics. The time averaged combined power of the quartet was obtained by adding the energy contributed by each voice at each frequency as scalar quantities to give a total sound power level L_w in dB

$$L_w = 10 * log_{10} \sum_{k=1}^{N/2} \sum_{m=1}^{4} A_{k,m}^2 / A_o^2 \tag{6.3}$$

where $A_{k,m}$ is the absolute amplitude of the harmonic in voice m at frequency f_k, A_o is a reference amplitude, m is the number of voices and N is the number of samples used in the spectral analysis.

6.1.3 Running the model

Each quartet chord sample was 743 ms duration, created by using 32,768 samples at a rate of 44,100 samples/sec, which gave a frequency resolution of 1.35 Hz in the spectral analysis. For the parametric studies a glottal open quotient O_q of 0.7 was used for each voice. For each synthesised chord the line spectrum, cumulative line spectrum, spectral centroid and spread were computed. As an example the output from the computer model corresponding to archive sample 005 (Figure 5.1) is shown in Figure 6.4. The computer programme was run using the same chords, root frequencies and vowels as the 50 archive samples to allow comparison.

The frequency resolution of the discrete Fourier transform used in the sample analysis was 1.35 Hz across the whole frequency range. It was chosen to give precision when calculating intonation accuracy at low frequencies. However this mathematically fine resolution does not truly represent the way the human ear analyses sound. Consequently an alternative spectral analysis method, more akin to human hearing, was also used for some analyses to assess how the changes in the model inputs could affect the perceived qualities of the sounds. For this alternative spectral analysis a '1/24 octave' filter bank of band pass filters, of order 16, was used which divided the frequency range into 'bins' of ± 25 cents width about a note's centre frequency. The 1/24 octave filter bank was used to assess the impact intonation inaccuracy and voice phase differences could have on the amplitude of the harmonics of the quartet sound.

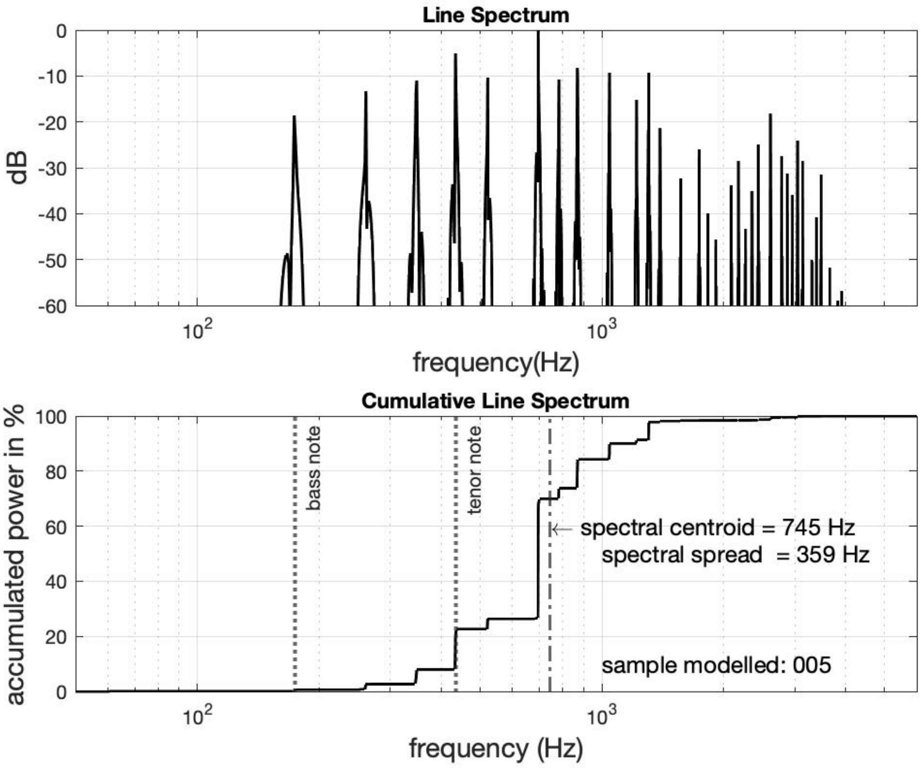


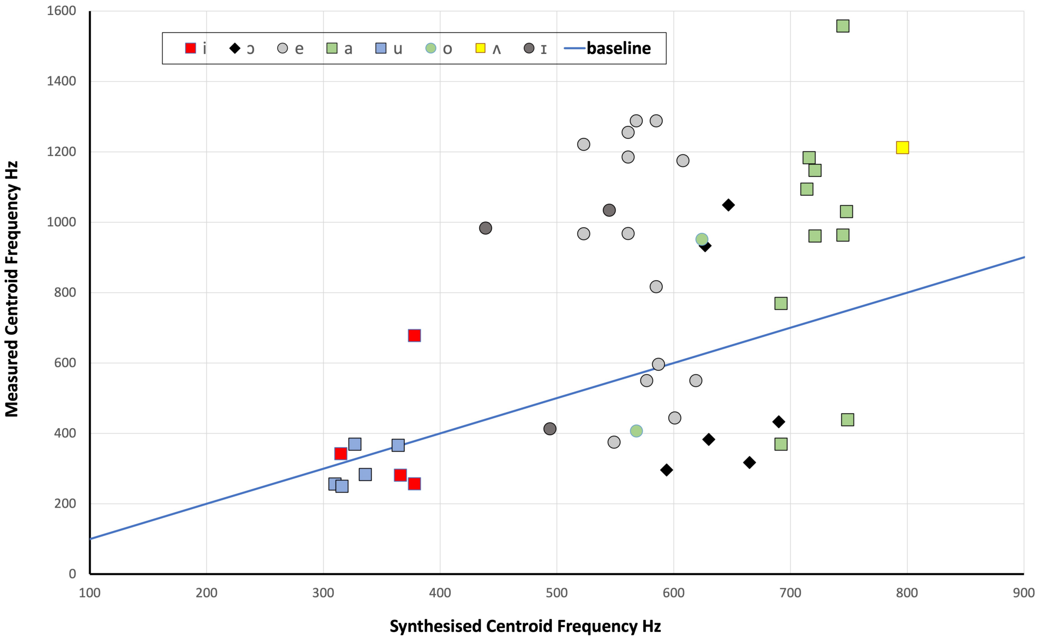
Figure 6.4: Example line spectrum and cumulative line spectrum for synthesised quartet model of archive sample 005

6.2 Results

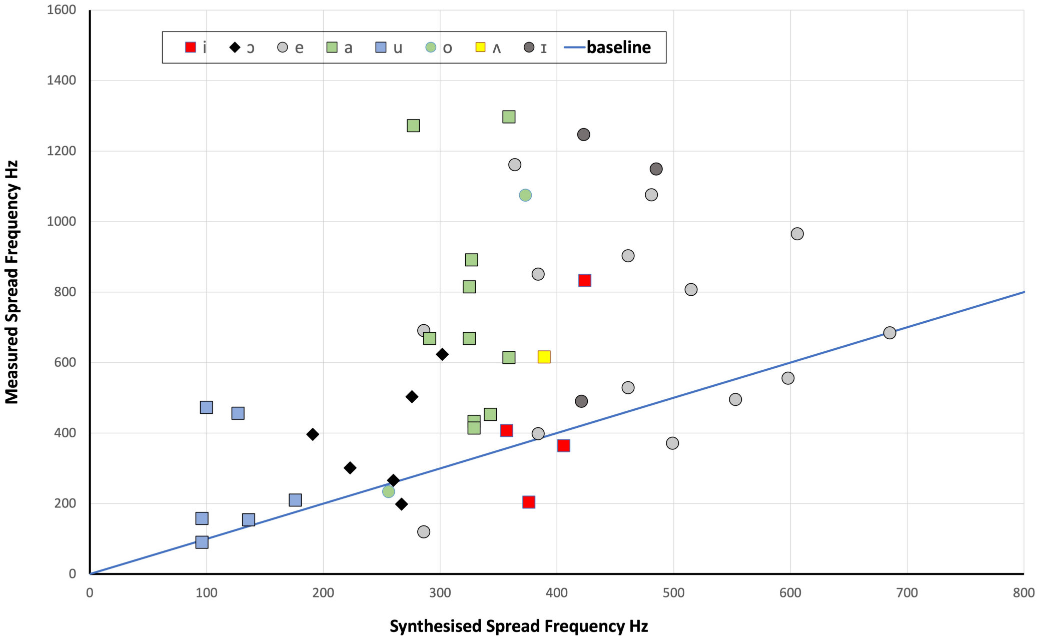
6.2.1 Comparison of computer model predictions with archive samples

The measured values of spectral centroid and spread of each archive sample are plotted in Figure 6.5 against the synthesised quartet values for the corresponding chord structure, chord root frequency and vowel. If the synthesised quartet predictions matched the archive sample measurements the results would lie on the diagonal blue “baselines”.

For the samples using vowels /u/ and /i/, which had relatively low centroids around 400 Hz, there is good correlation between the computer model and the archive samples. For the samples using vowel /a/, which had significantly higher centroids, there is considerable scatter and the computer model tends to underestimate the values found with the archive samples. The spectral spread for the /a/ vowel is generally greater than the /i/ and /u/ vowel, and again the higher values for the archive samples suggests the computer model underestimates this metric. Nevertheless, even with the scatter in the results, the computer model and archive samples highlight the considerable range in



(a) centroid



(b) spread

Figure 6.5: Synthesised quartet predictions plotted against archive sample measurements for the corresponding chords (a) centroid and (b) spread

acoustic metrics which can result from using different chord structures, root frequencies and vowels. The effect of changing each of these input variables was investigated further by running parametric studies using the computer model. The glottal flow for a voice part (and hence the glottal acoustic energy input) was kept the same for all the parametric iterations.

6.2.2 Impact of chord structure, chord root frequency and vowel on spectral centroid

The impact of changing the chord structure, chord root frequency and vowel on the spectral centroid was examined for four different chords by running the computer model across the range of singable chord root frequencies for a male quartet for the three corner vowels /i/, /a/, /u/. The four chords were a major triad with 10th voicing (frequency ratios 2:3:4:5), a Barbershop 7th with medium voicing (frequency ratios 4:6:7:10), a major triad with 8ve voicing (frequency ratios 4:5:6:8) and minor 6th with spread voicing (frequency ratios 6:9:10:14). The results are presented in Figure 6.6. The centroid for the /u/ vowel is consistently in the range 300 Hz to 400Hz across all chord root frequencies for all four chords. The centroid for the /i/ vowel is around 350Hz up to a chord root frequency of 140 Hz then it trends upwards and fluctuates with significant differences between the four chords. The centroid for the /a/ vowel is substantially higher with a minimum of 700 Hz and with pronounced peaks up to 900 Hz depending on chord structure and root frequency.

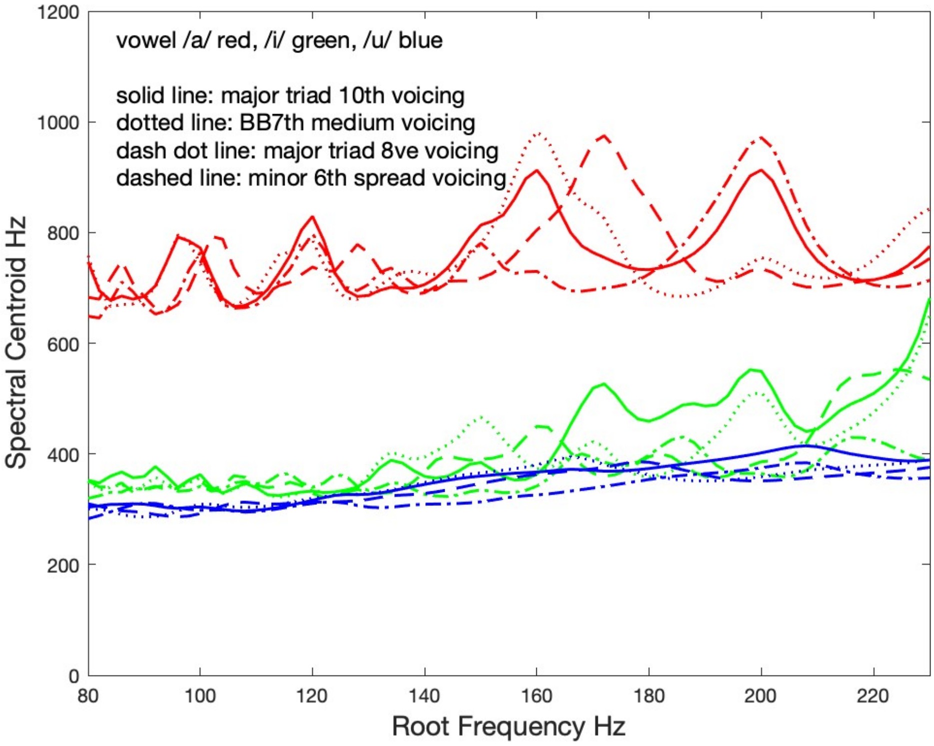


Figure 6.6: Plot of spectral centroid against chord root frequency for 4 different chords for the vowels /i/, /a/, /u/ calculated using the synthesised quartet model

6.2.3 Impact of chord structure, chord root frequency and vowel on spectral spread

The impact of changing the chord structure, chord root frequency and vowel on the spectral spread was examined for the same 4 chords across the range of singable chord root frequencies for the three corner vowels /i/, /a/, /u/. The results are presented in Figure 6.7.

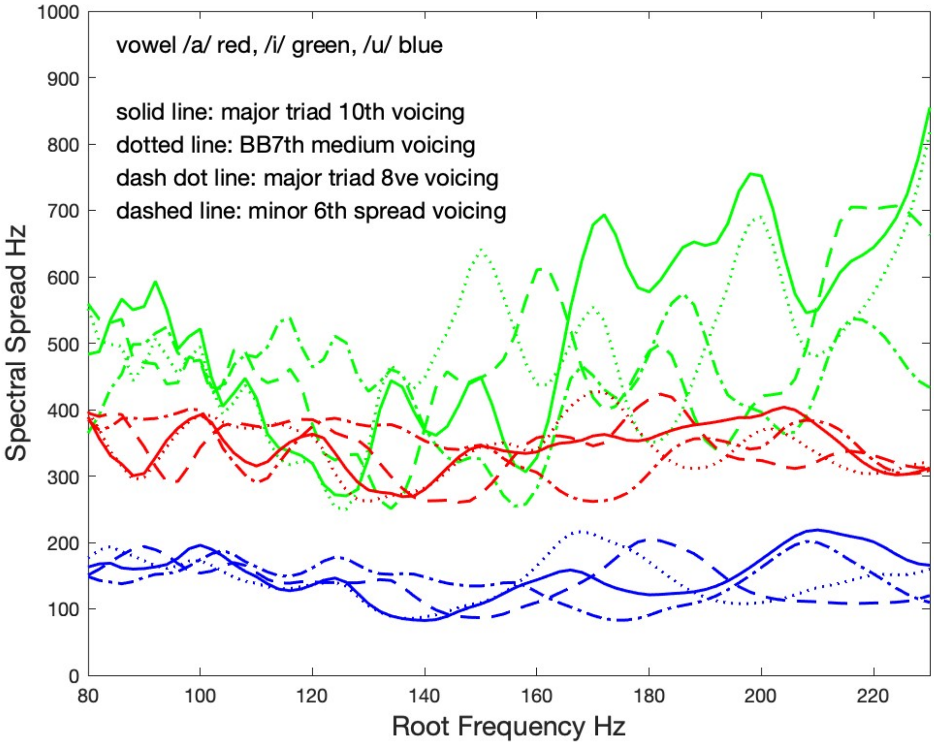


Figure 6.7: Plot of spectral spread against chord root frequency for 4 different chords for the vowels /i/, /a/, /u/ calculated using the synthesised quartet model

The spread for the /u/ vowel lies in the range 100 Hz to 200 Hz for all four chords but with differences in peaks depending on chord structure and chord root frequency. The spread for an /a/ vowel lies in the range 300 Hz to 400 Hz for all four chords and is similar to the /u/vowel with differences in peaks depending on chord structure and chord root frequency. In contrast the spread for the /i/ vowel ranges from 300 Hz to 800 Hz and is very sensitive to chord structure and chord root frequency.

6.2.4 Impact of chord structure, chord root frequency and vowel on sound power level

The impact of changing the chord structure, chord root frequency and vowel on the quartet's sound power output was examined for the same four chords across the range of singable chord root frequencies for the three corner vowels /i/, /a/, /u/. *The same glottal power input was used for each voice in each sample.* The results are presented in Figure 6.8. The sound power level output

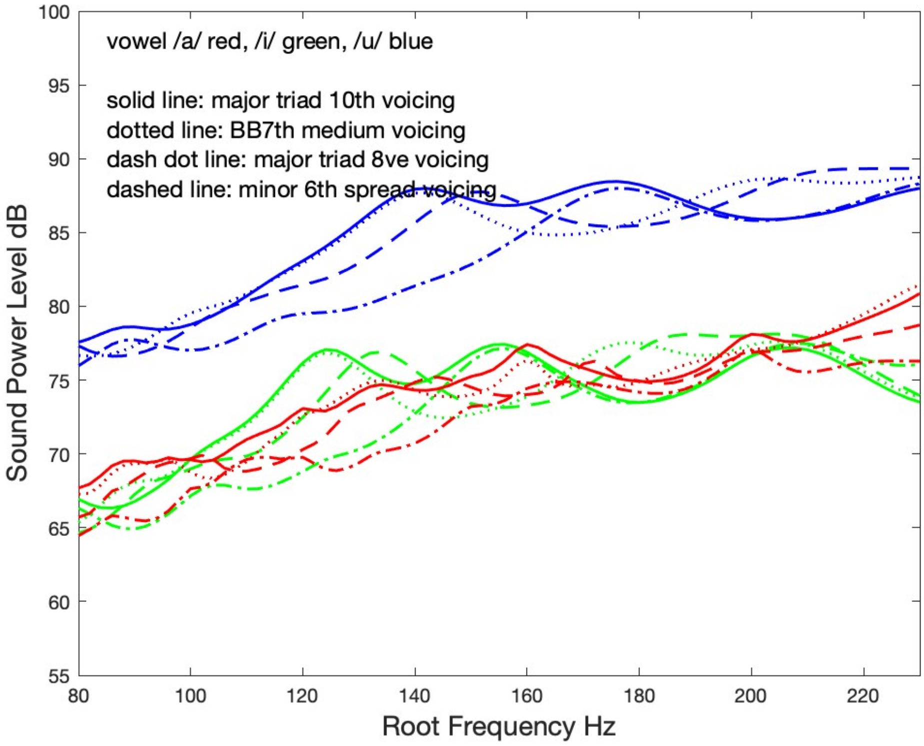


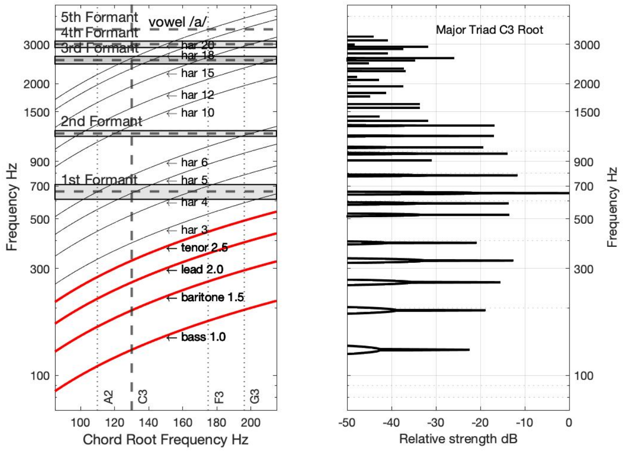
Figure 6.8: Plot of sound power level output L_w against chord root frequency for 4 different chords for the vowels /i/, /a/, /u/ calculated using the synthesised quartet model

L_w can be considered to be a measure of the vocal efficiency of the quartet as the glottal power input of the four voices was kept constant. The /u/ vowel produced greater total power output for all chord structures and root frequencies than either the /a/ or /i/ vowels, which had comparable levels. The larger values of power output for the /u/ vowel are a result of the vocal tract resonances (R1 and R2) being in the range of stronger lower harmonics; greater vocal efficiency is achieved with this vowel. Notably, for each vowel, the output power varies by as much as 6 dB depending on chord structure and root frequency. The peaks are a consequence of strong harmonics/ harmonic combinations aligning with vocal tract resonances. Resonance matching will lead to greater vocal efficiency and result in a boost to the total power output for the same power input. As explained by Titze (2018), most of the acoustic power generated at the glottis remains in the airway in the form of standing waves and is ultimately dissipated with only 0.1% being radiated at the lips. Consequently improved gain in the vocal tract can have a significant impact on the acoustic output. It is emphasised that when computing the acoustic output L_w no frequency weighting was applied to the power contributed by each harmonic. This should be taken into account when comparing the perception of the intensity of the sounds and the values of SPL obtained using a meter which generally have a frequency weighting applied.

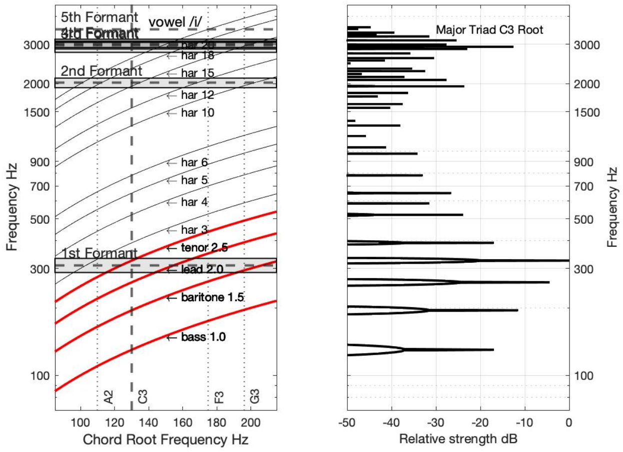
6.2.5 Impact of the singers vocal tract resonances on spectral centroid, spectral spread and sound power level

In the preceding sections the variation in centroid, spread and sound power level for different chord structures, chord root frequencies and vowels was attributed to the degree of alignment between the harmonics in a chord and the vocal tract resonances of the individual singers. This effect can be illustrated by referring to Figure 6.9. On the left hand side of the figure the phonation frequencies of the voices in a major triad chord with 10th voicing are shown in red, together with the harmonics, plotted against chord root frequency. The grey horizontal bands show the vocal tract resonances (R1, R2, R3...) for the corner vowels /a/, /i/ and /u/ in figures (a), (b) and (c) respectively. Examples of the resulting spectra for a chord on a root frequency of C3 are shown in the right hand side of Figure 6.9. As the root frequency of the chord increases across the singable range the harmonics in each voice are boosted when they align with the singers' vocal tract resonances. When the four voices are combined fluctuations in the centroid, spread and sound power level will result with changes in root frequency as evidenced in Figures 6.6, 6.7 and 6.8 .

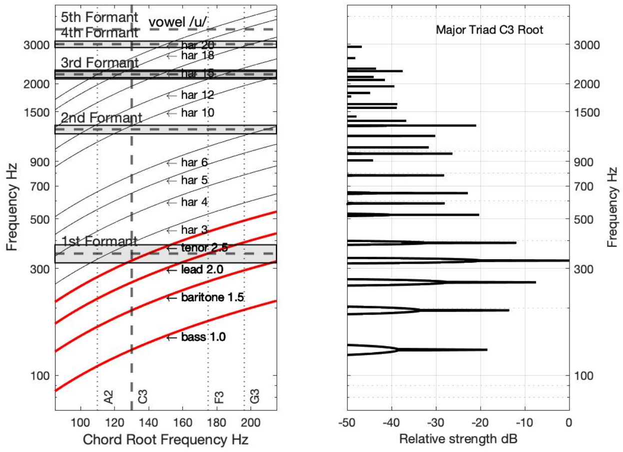
The values used for the vocal tract resonance frequencies in the computer model were derived from previous singing and speech research. In reality individual singers can have different vocal tract resonances without significantly altering the recognition of a vowel. If a singer adjusts a vocal tract resonance so it coincides with a harmonic in a chord it will increase the strength of that harmonic. The potential for an ensemble to implement "resonance tuning" can be illustrated by superimposing the computer model vocal tract gain envelope on to the measured frequency spectrum of an archive sample. Figure 6.10 shows the measured spectra of four of the archive samples with different chord structures, vowels and root frequencies. The spectra were scaled so the chord root frequency (the bass note) was around 70 dB for each sample. The vocal tract gain envelopes used in the computer model for the corresponding vowels are shown as the brown lines superimposed on the spectra. It can be seen that, although the vocal tract envelope generally matches the profile of the measured spectra, the strongest harmonics in each of the archive samples do not necessarily coincide with the tract resonance peaks of the computer model. The differences may arise because the computer model assumptions do not reflect the singers actual values and the harmonics have actually achieved their



(a) vowel /a/



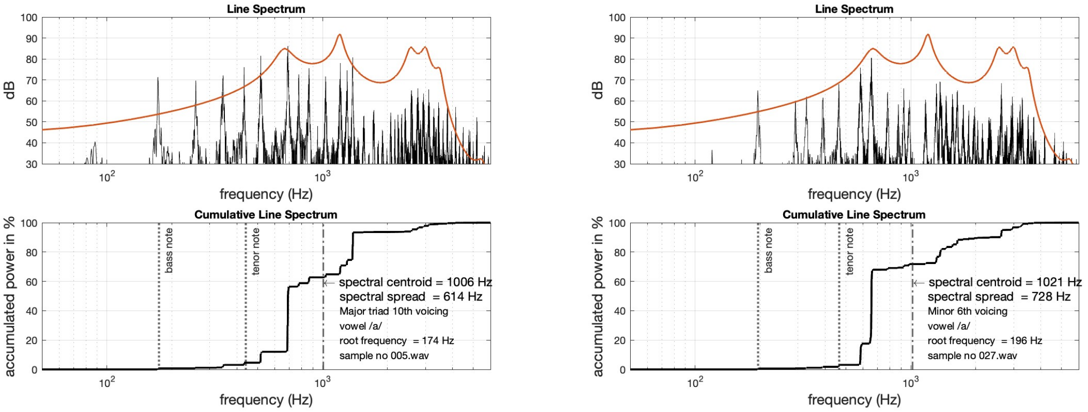
(b) vowel /i/



(c) vowel /u/

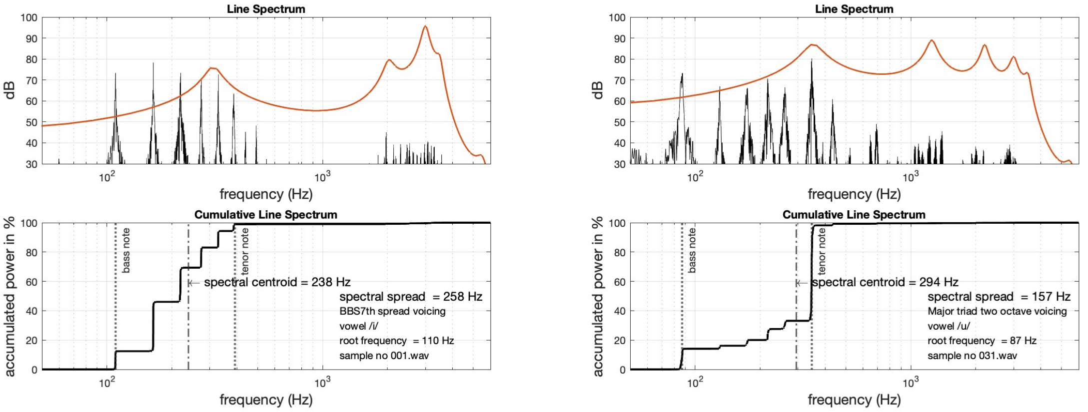
Figure 6.9: Left hand side: coincidence of chord harmonics in a major triad chord 10th voicing with vocal tract resonances at different root frequencies for different vowels. Right hand side: line spectrum for a chord with root C3

maximum values. Alternatively it may indicate there is opportunity to boost the harmonics further by adjusting the tract resonances so they coincide with the chord harmonics.



(a) Major triad 10th voicing /a/

(b) Minor 6th /a/



(c) BB7th spread voicing /i/

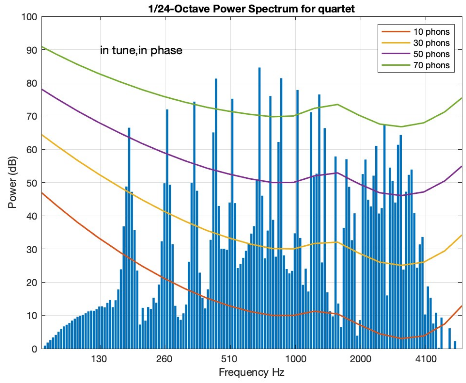
(d) Major triad 2 octave voicing /u/

Figure 6.10: Computer vocal tract gain envelope (brown lines) superimposed on archive sample spectra for different chord structures and vowels

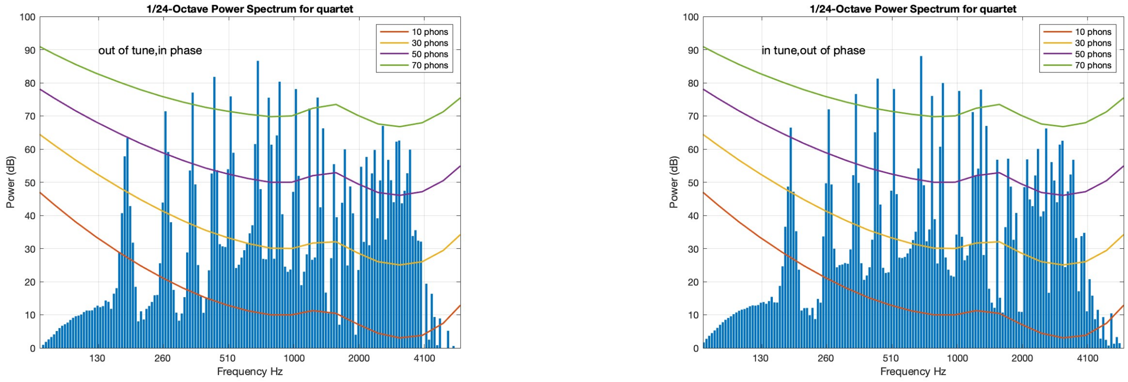
6.2.6 Impact of accuracy of intonation and phase between voices on sound power level

As the archive database had been extracted from diverse published sources it was not possible to measure the relative “loudness” of the samples so they were normalised to the amplitude of the maximum harmonic for comparison. This normalisation precluded evaluating the impact of chord structure, root frequency and vowel on the output sound power level. As described in the preceding sections the influence of these causal inputs was investigated by running the computer model using the same glottal power inputs for each sample, with the results presented in Figure 6.8. This analysis maintained the assumptions of accurate intonation and no phase differences between the voices.

Additional analyses were therefore undertaken to gauge the impact tuning inaccuracy and phase differences between voices could have on the peak strength of the harmonics in a sound. For this investigation a 1/24th octave filter bank was used to calculate the magnitude of the harmonic peaks so the distribution of acoustic power across frequency bands was more akin to the way the hearing mechanism filters sound.



(a) In tune in phase



(b) Out of tune, voices in phase

(c) In tune phase shift between voices

Figure 6.11: Spectral analysis of synthesised major triad chord 10th voicing using a 1/24 octave filter bank for three different scenarios (a) in tune and in phase (b) out of tune but in phase (c) in tune but out of phase

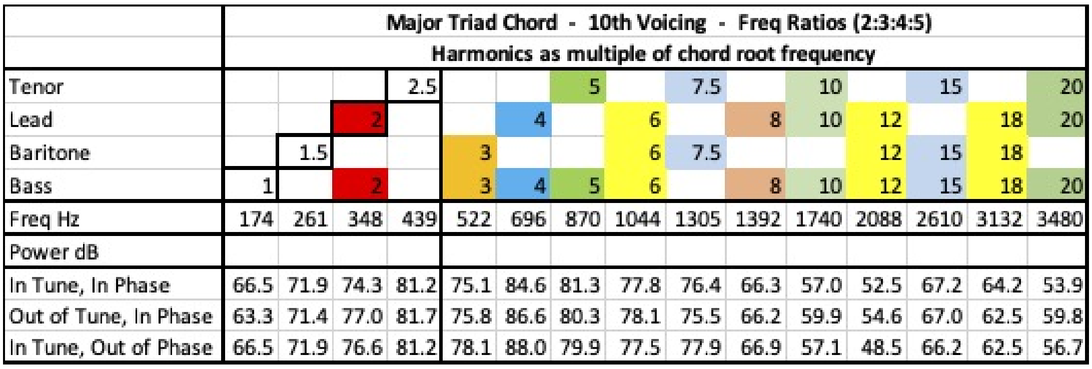
A major triad 10th voicing chord, with a root frequency of 174 Hz, was analysed using three different input assumptions:

(a) the four voices were in tune and in phase. The fundamental frequency ratios of the four voices were 2:3:4:5 and all singers starting at the same time (in phase)

(b) the four voices were out of tune but in phase. The frequency ratios were 1.98:3.03:4:5.05 which is 17 cents flat for the bass and 17 cents sharp for the baritone and tenor and the singers start in phase

(c) the four voices were in tune but out of phase. The fundamental frequency ratios of the four voices were 2:3:4:5 but relative to the lead note, which takes 2.87 msec/cycle, the bass, baritone and tenor start were offset by 0.7, 1.4 and 2.1 msec respectively

The resulting power spectra for the chord using each of these conditions are shown in Figure 6.11 overlaid with the equal loudness contours in phon. The phon contours were included to highlight that a change of a few dB in harmonic amplitude could have different perceptual impact depending on the frequency. The dB levels of significant harmonic combinations for the three scenarios are shown in Figure 6.12. Differences of 3 dB and more in the chord's harmonics occur between the three chords; for example the (bass + baritone) combination on the 3rd harmonic at 522 Hz and the (bass + baritone + lead) combination on the 12th harmonic at 2088 Hz. Changing the intonation accuracy or phase between voices causes a redistribution of acoustic energy between frequency bands which could produce perceptual differences in the sound.



	Major Triad Chord - 10th Voicing - Freq Ratios (2:3:4:5)														
	Harmonics as multiple of chord root frequency														
Tenor				2.5			5		7.5		10		15		20
Lead			2			4		6		8	10	12		18	20
Baritone		1.5			3			6	7.5			12	15	18	
Bass	1		2		3	4	5	6		8	10	12	15	18	20
Freq Hz	174	261	348	439	522	696	870	1044	1305	1392	1740	2088	2610	3132	3480
Power dB															
In Tune, In Phase	66.5	71.9	74.3	81.2	75.1	84.6	81.3	77.8	76.4	66.3	57.0	52.5	67.2	64.2	53.9
Out of Tune, In Phase	63.3	71.4	77.0	81.7	75.8	86.6	80.3	78.1	75.5	66.2	59.9	54.6	67.0	62.5	59.8
In Tune, Out of Phase	66.5	71.9	76.6	81.2	78.1	88.0	79.9	77.5	77.9	66.9	57.1	48.5	66.2	62.5	56.7

Figure 6.12: Impact on power in chord harmonics of inaccurate intonation and phase differences between voices

6.3 Discussion

The analyses of the archive database and computer model samples revealed a substantial range of the timbral metrics of spectral centroid, spectral spread and acoustic power output can be produced by a quartet, with chord structure, chord root frequency and vowel sung having significant influence. The effects were illustrated using the three "corner" vowels /a/, /i/ and /u/. The "corner" vowels define the limits of the vocal tract resonances in (R1, R2) space. Vowel /a/ has a high R1 (650 Hz) and low R2 (1200 Hz), vowel /i/ has a low R1 (350 Hz) and a high R2 (2100 Hz) and vowel /u/ vowel has a low R1 (350 Hz) and low R2 (1200 Hz). Other vowels have (R1, R2) combinations lying within these limits. Where the harmonics generated by a singer sit relative to their vocal tract resonances has a major influence on their vocal output and consequently on the aggregate sound of the ensemble, bearing in mind that the strength of the harmonics produced by the glottal source diminish in strength at a rate of around 6 dB/octave.

The phonation frequency ranges for untrained male Barbershop singers are broadly in the ranges given below, with experienced singers able to extend these ranges, especially when using falsetto:
Bass range from F2 (87 Hz) to C4 (261 Hz)
Baritone range from C3 (130 Hz) to E4 (330 Hz)
Lead range from D3 (147 Hz) to F4 (349 Hz)
Tenor range from B3 (247 Hz) to B4 (493 Hz)

The /a/ vowel has a high R1 around 650 Hz, which is above the phonation frequency of all four voice parts throughout their singable range, combined with a relatively low value of R2 around 1200 Hz. The greatest contribution to a chord's energy for this vowel arises when the harmonics in each voice part at 2, 3, 4, 5 or 6 x their phonation frequencies combine and align with R1. This can be seen in the spectrum for the chord with a root frequency of C3 on the right hand side of the Figure 6.9. It can also be seen in Figure 6.10 (a) where the strongest harmonic in the chord is at 690 Hz (4 x bass phonation frequency + 2 x lead phonation frequency), which is close to R1 for an /a/ vowel and in Figure 6.10 (b) where the strongest harmonic in the chord is at 660 Hz (2 x lead phonation frequency) and close to R1. The energy at R1 is augmented by the energy in the combined voice harmonics at 8, 9, 10 and 12 x the root frequency when they fall in the region of R2. As a consequence the *centroid for an /a/ vowel* is relatively constant around 700 Hz, largely determined by R1, but higher values arise when harmonic combinations match R2.

An /i/ vowel has a low R1 around 350Hz and a high R2 around 2100 Hz. At lower chord root frequencies, up to around 160 Hz, the singers' phonation frequencies and the lower harmonic combinations are in the region of R1 which dominates the behaviour. As the chord root frequency is increased the centroid rises as the harmonic combinations at 20, 18, 15, 12 and 10 x the root frequency make an increasing contributions, causing fluctuations in the values as they coincide with the high vocal tract resonances at R2 and R3. This can be seen in the spectrum for a chord with root frequency at C3 on the right hand side of the Figure 6.9 and in Figure 6.10 (c) where the bass' 3rd harmonic at 328 Hz, which aligns with R1, is comparable in strength to the bass, lead and baritone fundamental frequencies which are all below R1. The *centroid for an /i/ vowel* is relatively constant around 350 Hz, largely determined by R1, but higher values arise above a chord root frequency of 160 Hz when stronger harmonic combinations match R2.

An /u/ vowel has a low R1 around 350Hz and a relatively low R2 around 1200 Hz. At low root frequencies the lower harmonic combinations of the singers are in the region of R1. As the chord root frequency is increased, firstly the tenor, followed by the lead will have phonation frequencies coinciding with R1 which will become the predominant harmonic in the chord and will largely determine the value of the spectral centroid around 350Hz. Higher harmonic combinations make contributions when they coincide with the vocal tract resonances at R2 but the value of the centroid is dominated by R1. This can be seen in the spectrum for a chord with root frequency at C3 on the right hand side of the Figure 6.9 and in Figure 6.10 (d) where the strongest harmonic in the chord is at 347 Hz (the combination of the bass' 4th harmonic with the tenor's fundamental frequency). The *centroid for an /u/ vowel* is relatively constant around 350 Hz, largely determined by R1.

The behaviour of vowels with values of R1 and R2 lying between the corner vowels can be explained in the same way. The values of centroid generally lie within those of the corner vowels, see for example /e/ and /ɔ/ in Figure 6.5.

The *spectral spread for the /a/ and /u/ vowels* are relatively constant, with values around 350 Hz and 150 Hz respectively, across the chord root frequency range. Their behaviour is largely determined by the stronger lower harmonic combinations being in the region of R1 and R2. In contrast the *spectral spread for an /i/ vowel* fluctuates very significantly and is sensitive to the higher harmonic combinations aligning with R2 around 2100 Hz and R3 (Figure 6.7). The different behaviour between the corner vowels is also reflected in intermediate vowels with the factor which largely determines the spread being the difference between R2 and R1; the greater the value of (R2-R1) the greater the spread.

The measured spreads for the archive samples were generally higher than the synthesised values indicating the higher harmonics in the archive samples contributed a greater percentage of the acoustic energy than in the computer model. The differences can be attributed to one or more of the following: the singers have higher vocal tract resonances than assumed in the computer model; the open quotient of a singer is lower than the 0.7 assumed in the computer model (a less breathy sound), which will increase the strength of the higher harmonics in the glottal flow spectrum; there is greater skew in the glottal flow wave than in the computer model, which will increase the strength of higher harmonics; live singers tune their vocal tract resonances to the harmonics of the note being sung; the contribution of the singers' formant at R3, R4, R5 has been underestimated in the computer model.

The impact of changing the chord structure, chord root frequency and vowel on the quartet's combined *sound power level* was examined across the range of singable chord root frequencies. Using the same glottal power input in each voice, the combined output power varied by as much as 6 dB for each vowel depending on chord structure and root frequency (Figure 6.8). The peak values occur when strong harmonics/ harmonic combinations align with vocal tract resonances. The /u/ vowel produced greater total power output for all the chords and root frequencies than either the /a/ or /i/ vowels. The /u/ vowel is more efficient vocally, largely because the stronger lower harmonics are in the range of R1 and R2. The variation in total power output with chord structure root frequency and vowel indicates greater efficiency can be achieved in converting glottal input energy to combined radiated power through resonance matching. While the calculated sound power levels are related to the SPL measurements obtained using a meter, no frequency weighting was applied in the analysis.

Inspection of the results from four archive samples shown in Figure 6.10 reveals the peaks in the vocal tract resonance envelopes (the brown lines) used in the computer model were close to the strongest measured harmonic combinations in the archive samples but they did not completely align. While the resonant frequencies used in the computer model are assumed values, the comparison highlights the opportunity to increase acoustic power output if one or more of the singers modifies their vocal tracts to bring the resonances closer to the harmonics in the chord. There is increased opportunity to exploit resonance tuning with multiple voices compared to the solo voice as it may be possible for an ensemble to tune to R1 and R2 at the same time.

Inaccurate intonation and phase differences between the voices was shown to have an effect on the magnitude of the harmonic peaks in the spectrum. These factors could have a bearing on the perception of the sound.

- Vowels with a low R1 and a low R2 (/u/) produce low centroids and low spread
- Vowels with a low R1 combined with high R2 (/i/) produce low centroids but high spread
- Vowels with a high R1, above the phonation frequencies of all the singers, combined with a low R2 (/a/), produce high centroids but the reduced power in the lower harmonics limits the spread
- For a given chord structure, root frequency and vowel the total radiated power output can be boosted using resonance tuning on important harmonic combinations
- Inaccurate intonation and phase differences between participating voices can have a discernable impact on the strength of chord harmonics

On the basis of the parametric analysis with the computer model it is concluded the predominant factor affecting the acoustic metrics of a quartet is the proximity of each singers' harmonics to their vocal tract resonances. This proximity is greatly influenced by the choice of chord structure, chord root frequency and vowel sung but can be exploited further by the participating singers adjusting their vocal techniques to use resonance tuning.

6.4 Limitations

The synthesised quartet computer model is a useful tool for evaluating the effect changing the variables has on the acoustic characteristics of a vocal ensemble but it has limitations in the number of variables which can be changed at any one time:

- real singers adopt different vocal production methods which affect sound pressure level, glottal opening period and vocal tract resonances. The computer model uses the same parameters, apart from phonation frequency, for each singer
- the singers may employ resonance tuning by moving their vocal tract resonances towards the harmonics of the note they are singing. The computer model has fixed vocal tract resonances across all pitches for a given vowel
- the choice of resonant frequencies and bandwidths in the computer model will affect the prominence and magnitude of the gain provided by each vocal tract resonator
- the computer model assumes the voices combine as incoherent sources, but with accurate intonation some voice harmonics will have the same frequency and for a limited number of singers, such as a quartet, the phase differences between the contributing voices could have a significant influence
- the computer model may underestimate the contribution of the singers formant at R3, R4, R5

6.5 Summary

The analysis of the archive samples and the computer model show a significant range of the acoustic properties of spectral centroid, spread and total power output can be generated by a quartet, primarily influenced by chord structure, chord root frequency and vowel being sung. Importantly it was identified that singers can maximise these metrics by implementing resonance tuning. The computer model showed behaviours which are consistent with archive sample measurements but generally predicts lower values for the metrics for the corresponding chords, vowels and chord root frequencies. Closer correlation between the archive quartet samples and the computer model may be achievable by using more accurate voice parameters for each voice part in the computer model. To address these limitations and corroborate the conclusions from the archive analysis and computer modelling laboratory tests were conducted with a live quartet, which is the subject of the next chapter.

Chapter 7

Laboratory Experiments With a Live Quartet

The analysis of the archive samples in Chapter 5 revealed a wide range of the timbral features of *spectral centroid* and *spectral spread* results from changing the structure of a chord, its root frequency and the vowel sung. However, the impact these variables may have on the *acoustic power output* of the quartet could not be quantified because there was no comparable reference level for the recordings. The computer model in Chapter 6 was used to validate the findings of the archive sample analysis and was also used to assess the impact of the variables on the total output power. These two sets of investigations focused on the overall acoustic characteristics of a quartet without assessing each singer's contribution to the sound. The individual singers can affect the overall ensemble output through their intonation accuracy and resonance tuning, and possibly making adjustments in real time while sustaining longer duration chords.

To validate the earlier conclusions about the variables affecting the timbral metrics and to investigate the influence the individual singers can have on the overall characteristics, tests were carried out with an award-winning Barbershop quartet under laboratory conditions. The tests gave the opportunity to monitor the individual singers' vocal characteristics and see if they modified them while singing.

Singing exercises were created to investigate:

- whether chord structure, chord root frequency, and vowel sung had an impact on spectral centroid and spread as evidenced by the archive sample analysis and the computer model, and assess the impact these variables had on the intensity of the quartet's sound
- the influence of vowel and note sung on the spectral centroid, spread and intensity of the individual singers
- how intonation accuracy, glottal open quotient and vocal tract resonances compared between singers and with the inputs for the synthesised quartet computer model
- the effect of adjustments made by the singers in real time had on the acoustic output of the quartet when singing sustained chords

Figure 7.1: Sound Hypothesis in the AudioLab, University of York

Sound Hypothesis, the quartet which undertook the tests, perform Barbershop singing to a very high standard. They won the UK Barbershop quartet gold medal in 2019. The testing took place in the AudioLab at the University of York (Figure 7.1). The testing protocol was approved by the Ethics Committee at the University of York; approval reference Cookson180222. Sound Hypothesis gave consent to be identified and to use the photographs and recordings taken during the laboratory test programme.

7.1 Methodology

7.1.1 Series 1 tests

The tests in Series 1 were carried out in a studio which had been acoustically treated to minimise reverberation and external noise. The quartet sang a defined 15-bar chord sequence with the acoustic output of the individual singers and the full quartet being recorded. Electroglottographic (EGG) measurements were taken concurrently for each singer to determine their fundamental frequency and glottal open quotient. On completion of the 15-bar exercise the quartet sang five of their repertoire songs which contained words using the same vowels as in the 15-bar exercise with the same acoustic measurements being taken.

Each singer was fitted with a face-mounted microphone (DPA4066-F omnidirectional) approximately 4 cm from the lips to record their individual sound production. The microphones were placed on the side of the face furthest away from the greater number of singers to minimise crosstalk (Figure 7.1). To measure the fundamental frequency and glottal open quotient two electrodes were placed on either side of the larynx of each singer's neck to record an electroglottographic (EGG) signal (Henrich, d'Alessandro, et al. 2004). The output was fed into a digital field Laryngograph microprocessor.

The singers stood in an arc in their conventional quartet contest formation; baritone, lead, bass, tenor from left to right when viewed from the front. An XY microphone pair (Rode NT4 stereo) was placed 1100 mm from the singers at mean chest height to record the quartet output. The angle between the microphone pair was 100 degrees. The recordings from the two XY microphones were added and averaged. The acoustic environment, singer positioning and microphone configuration were chosen so the combined output of the singers recorded at the XY pair was a reasonable approximation to superimposed planar waves without reverberations. The DPA microphone and EGG signals for each singer and the XY pair were all fed concurrently through a Focusrite Mk2 Dynamic audio interface and captured on a Reaper digital audio workstation.

To calibrate sound levels each quartet member sang at a medium volume level in the median range for their voice part. Calibration SPL levels were recorded using a Tenma 1350A digital sound level meter held close to the DPA microphone. Each singer sang an ascending scale through their full vocal range in their normal singing voice at a medium level with one breath for each note of approximately 4 seconds duration. A pitch note was played on an electric keyboard for the start note. The process

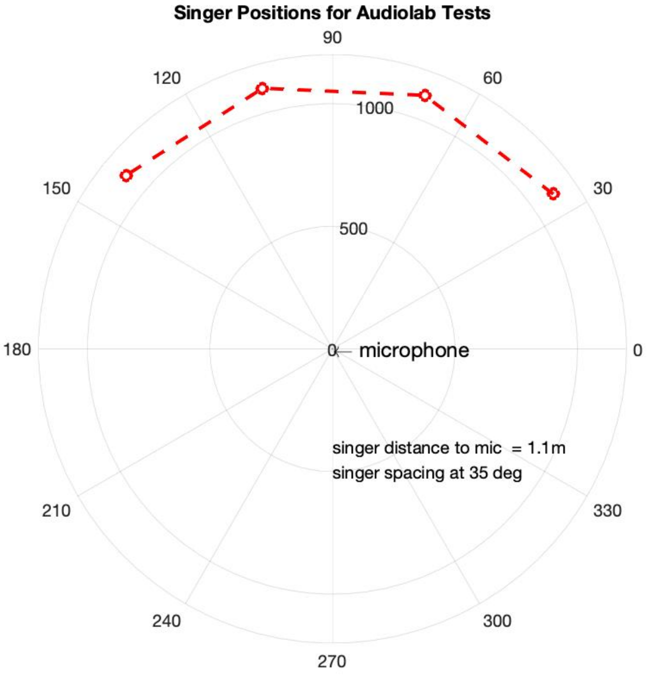


Figure 7.2: Singer Positions: Left to Right Baritone:Lead:Bass:Tenor when viewed from the front

was carried out four times using four different vowel sounds /ɪ/, /ʊ/, /ɑ/, /a/ (Hid, Hood, Hard, Had).

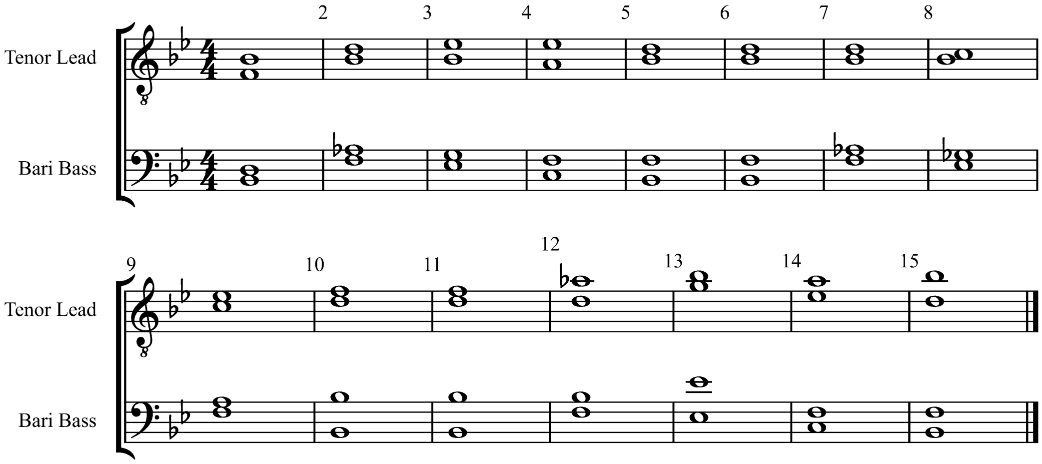


Figure 7.3: AudioLab Exercise. 15-bar chord sequence

The 15-bar chord exercise used in the tests is shown in Figure 7.3. It comprised a series of different chord structures and chord root frequencies as set out in Table 7.1. A pitch note giving the root note of the first chord (B♭2) was played on an electronic keyboard. Each voice part sang the exercise solo before singing it as a quartet. They only took a breath between bars 5 and 6 and then between bars 10 and 11. One run through of the exercise took approximately 35 seconds. The exercise was

Bar	Chord Structure	Structure	Freq Ratio	Root Note	Chord Notes
01	Triad octave voicing	1-3-5-8	4:5:6:8	B♭2	B♭2, D3, F3, B♭3
02	BB7th close voicing. Bass 5th	5-7♭-8-10	6:7:8:10	B♭2	F3, A♭3, B♭3, D4
03	Triad octave voicing	1-3-5-8	4:5:6:8	E♭3	E♭3, G3, B♭3, E♭4
04	BB7th medium voicing. Bass 5th	5-8-10-15♭	3:4:5:7	F2	C3, F3, A3, E♭4
05	Triad 10th voicing	1-5-8-10	2:3:4:5	B♭2	B♭2, F3, B♭3, D4
06	Triad 10th voicing	1-5-8-10	2:3:4:5	B♭2	B♭2, F3, B♭3, D4
07	BB7th close voicing. Bass 5th	5-7♭-8-10	6:7:8:10	B♭2	F3, A♭3, B♭3, D4
08	Minor 6th close voicing	1-3♭-5-6	30:36:45:40	E♭3	E♭3, G♭3, B♭3, C4
09	BB7th close voicing	1-3-5-7♭	4:5:6:7	F3	F3, A3, C4, E♭4
10	Triad 12th voicing	1-8-10-12	2:4:5:6	B♭2	B♭2, B♭3, D4, F4
11	Triad 12th voicing	1-8-10-12	2:4:5:6	B♭2	B♭2, B♭3, D4, F4
12	BB7th medium voicing. Bass 5th	5-8-10-15♭	3:4:5:7	B♭2	F3, B♭3, D4, A♭4
13	Triad 12th voicing	1-8-10-12	2:4:5:6	E♭3	E♭3, E♭4, G4, B♭4
14	BB7th spread voicing. Bass 5th	5-8-15♭-18	3:4:7:10	F2	C3, F3, E♭4, A4
15	Triad 2 x octave voicing	1-5-10-16	2:3:5:8	B♭2	B♭2, F3, D4, B♭4

Table 7.1: 15-bar exercise

conducted with two repetitions of seven vowels which were recorded back to back in one studio session without any adjustments being made to the testing arrangements.

The quartet then performed five songs from their repertoire which contained chords using the same vowels as the exercise. The EGG and microphone measurements used in the 15 chord exercise were also used for recording the full repertoire songs.

7.1.2 Series 2 tests

The tests in Series 2 were to estimate the vocal tract resonances of each singer across their vocal range when singing different vowels. The test procedure used a resonance measurement device (RMD) previously used by Vos et al. (2017) at the University of York based on a methodology initially developed by Epps et al. (1997). The RMD excites the vocal tract by playing a sweeping synthesised broadband noise signal just outside a singer's mouth while they sing a given vowel on a fixed note. The combined signal (singer and noise source) is measured with a microphone also located at the lips. The broad band noise source consisted of 734 harmonics spaced at 5.38 Hz intervals across the frequency range from 50 Hz to 4000 Hz. The sample rate was set at 44,100/s with 8192 samples per window used for the spectral analysis. The microphone signal was fed through a Focusrite Scarlett Solo audio interface.

Practice runs were carried out for each singer to familiarise themselves with the apparatus. The singer was asked to use their normal singing voice at a medium level and maintain a constant mouth shape during the exercise. The acoustic impedance of the singer with the lips closed was first calibrated using the broad band source held close to the lips. Each singer then sang through their full vocal range with each note held for approximately 5 seconds. The exercise was repeated for seven vowels /ɪ/, /ʊ/, /ɑ/, /a/, /i/, /ɛ/, /ɒ/. The output was monitored in real time and the singer repeated the exercise if the output had inconsistencies in performance.

7.1.3 Methods of analysis

The spectral centroid, spread and SPL for both the solo singers and the quartet were derived from the XY microphone pair measurements. The original plan was to use the DPA face microphone recordings to determine the contribution of each voice to the quartet, after adjusting for cross talk between singers, but it was found the DPA spectra differed from the XY spectra at the higher harmonics as a result of the non-uniform acoustic radiation pattern from the mouth. Additionally the cross talk between the singers' DPA microphones was evident. Instead, each singer's SPL, spectral centroid and spectral spread when singing the 15-bar exercise *solo* was derived from the XY microphone recordings to give an indication of how the characteristics varied, recognising they may adapt their vocal technique when singing together. To analyse the spectral centroid and spread through the duration of the 15-bar exercise incremental segments of 8192 samples with 50% overlap equating to 93 ms steps with a 5.4 Hz resolution were analysed. The SPL levels were determined using the Matlab function "splMeter" with a fast time weighting and A frequency weighting. To visualise the changes in frequency spectrum with time the Matlab programme "pspectrum" was used with segments of 50 msec duration giving a 20 Hz resolution with a 25% overlap.

Throughout all the exercises the fundamental frequencies (F_0) and open quotients (O_q) of each singer were computed from the EGG recordings using the open source Matlab programme "oq_egg" documented by Henrich N (2017). F_0 is calculated from the fundamental period which is taken to be the duration between two consecutive glottal closing instants which can be accurately determined. O_q is the ratio between the open time and the fundamental period. The open time is the duration between the glottal opening instant and the consecutive closing instant. There are alternative ways for defining the open and closing instants (Henrich, d'Alessandro, et al. 2004) which means it is not an absolute measure but it is suitable for comparative purposes provided the method of calculation is consistent. In the analysis used here the open and closing instants are taken to occur at 50% of the difference between the maximum and minimum values of the EGG signal.

The RMD analysis was carried out using the open source software (Jeanneteau M 2019) originally written by J Wolfe and J Smith at the Acoustics Laboratory at the University of New South Wales.

7.2 Results

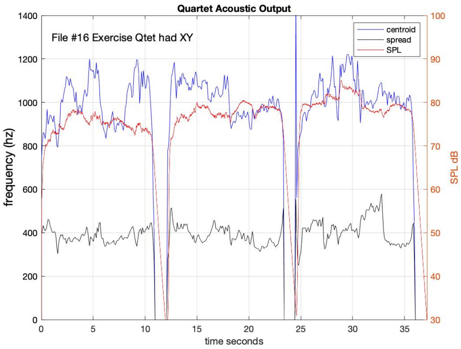
7.2.1 Influence of compositional choices on acoustic output

To assess the influence of the compositional choices of chord structure, chord root frequency and vowel on the quartet's acoustic output the 15-bar chord exercise was analysed for three "corner" vowels /a/, /i/, /ʊ/. The spectral centroid, spread and SPL for the quartet, plotted against time, through the performance of the exercise for each vowel are shown in Figure 7.4. The figures show a marked difference in behaviour between the vowels for the full range of chord structures and chord root frequencies. The centroids for the /a/ vowel and /ʊ/ vowel are around 1000 Hz and 500 Hz respectively throughout the 15 bars, while the centroid for the /i/ vowel fluctuates considerably between 400 Hz and 1000 Hz. Similarly the spreads for the /a/ and /ʊ/ vowels are steady at 400 Hz and 300 Hz while the /i/ vowel varies from 400 Hz to 900 Hz. The SPL for the /a/ vowel, which ranges from 70 to 80 dB, is noticeably higher than the /i/ and /ʊ/ vowels, which are in the range 60 to 70 dB.

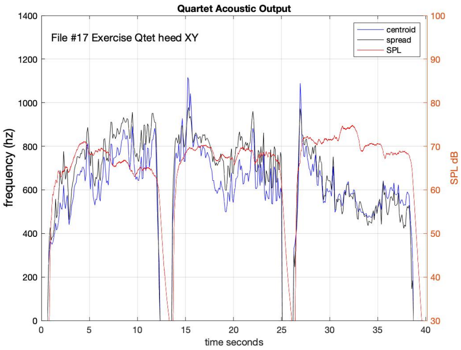
7.2.2 Influence of singer's vocal characteristics on acoustic output

When analysing the archive samples it was not possible to isolate the characteristics of the individual voices whereas for the synthesised quartet model the vocal characteristics were known inputs but were assumed to be the same for all four voices. In the Audiolab tests it was possible to monitor each singer while performing to see if their vocal characteristics differed and how they compared with the parameters assumed in the computer model. When recording the acoustic output of the quartet each singer's EGG measurements were taken concurrently. Separate tests using the RMD device were conducted to assess the vocal tract resonances of each singer when performing different vowels.

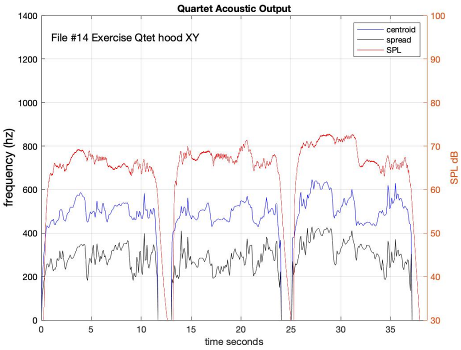
Examples of the Electrolaryngograph (ELG) recordings for each voice part singing an /a/ vowel are shown in Figure 7.5. The amplitude is a measure of the extent of opening of the vocal folds.



(a) /a/



(b) /i/



(c) /ʊ/

Figure 7.4: Quartet centroid, spread and SPL for different vowels (a) vowel /a/ in had (b) vowel /i/ in heed (c) vowel /ʊ/ in hood

Although there is no defined relationship between EGG measurements and the flow rate of the sound wave above the glottis a qualitative comparison of the results is illuminating.

Firstly the skew in the extent of glottal opening shows a marked difference between the opening and closing phases, which is expected when singing in the modal voice register (Henrich, d'Alessandro, et al. 2004). The rate of closing is faster than the rate of opening which has an impact on the spectrum for the glottal flow. Secondly there is an apparent difference in the amplitude of the signal between the voices. Although no absolute comparison can be made between the singers because

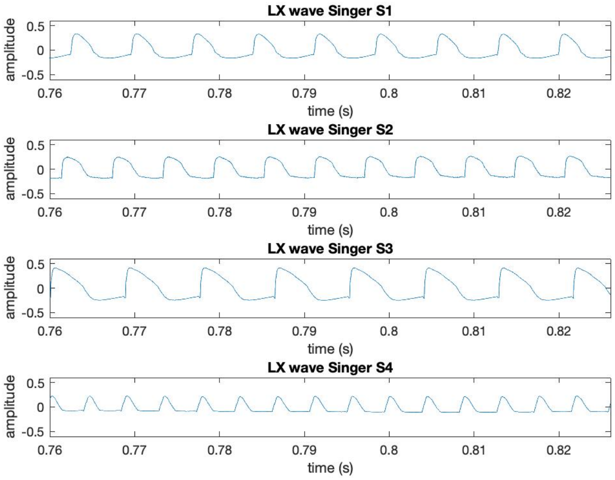


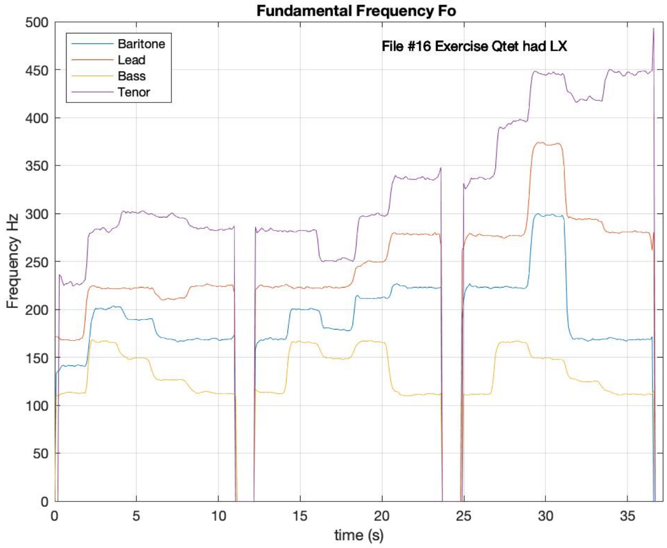
Figure 7.5: Electrolaryngograph waves for singers S1 (baritone), S2 (lead), S3 (bass), S4 (tenor)

there was no calibration method there does appear to be some relationship between the phonation frequency and the amplitude. To generate the same average glottal power a smaller vocal fold opening would require higher peak flow rates.

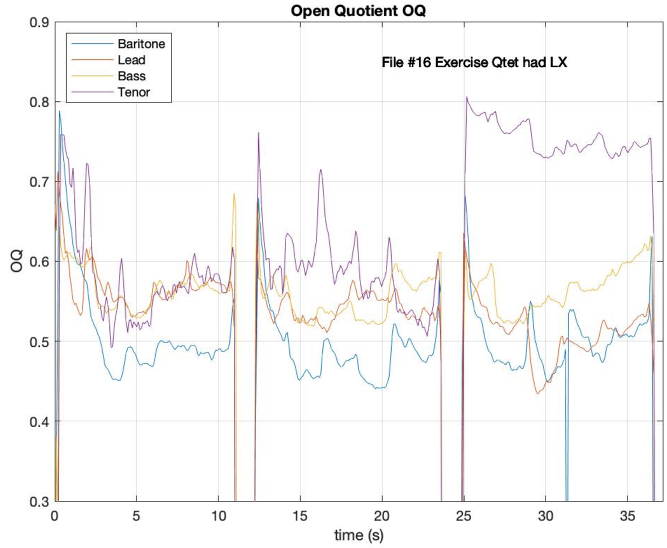
The fundamental frequency F_0 and the glottal open quotient O_q for each voice part, computed from the EGG recordings, when singing the 15-bar exercise as a quartet on an /a/ vowel are shown in Figure 7.6. All voice parts reached their respective notes in the exercise quickly, accurately and maintained them. The O_q values for the lead and bass singers were between 0.5 and 0.6 throughout. The tenor O_q was around 0.6 for the first 10 bars but stepped up to 0.75 and above in the final 5 bars when the mode of singing changed to falsetto at the higher phonation frequencies. The baritone maintained the lowest values for O_q in the range 0.45 to 0.5 throughout. These values contrast with the value of 0.7 for O_q used for all singers in the computer model.

The vocal tract resonances for each singer on each vowel were estimated from the RMD recordings obtained from the series 2 tests. An illustrative output is shown in Figure 7.7 where the results of the baritone singing the vowel /a/ at three different phonation frequencies; measured as 171 Hz (F3), 191 Hz (G3) and 216 Hz (A3) are overlaid. The harmonics of the notes being sung are the large peaks in the figure. The strongest harmonic is at 651 Hz, which is the 3rd harmonic of the A3 being sung and is close to R1. The peaks in the part of the spectrum responding to the broadband noise signal correspond to the vocal tract resonances R1, R2 and R3 at 640 Hz, 1240 Hz and 2600 Hz.

Figure 7.8 shows the RMD results for each voice part singing an /a/ vowel through a scale of 10 notes with the results overlaid. The tenor also sang through to C5, changing the mode of singing



(a) fundamental frequency



(b) open quotient

Figure 7.6: Singers characteristics during 15-bar exercise (a) fundamental frequency (b) open quotient

to falsetto at G4 (the results are in Appendix C). The spectral response to the broadband noise signal would be the same for each voice part on each note if the tract resonances were constant for a given vowel. However the resonances (R1, R2, R3) differed between voice parts and altered with pitch, resulting in the thickened lines in the spectra. Even though the vowel being sung was intentionally and recognisably the same, the resonant frequencies of the vocal tract changed with pitch. The corresponding figures for the vowels /i/ and /ʊ/ are given in Appendix C and display similar behaviours. Based on the RMD results the general values of the resonances R1, R2, R3 can be estimated. The values for an /a/ vowel for each voice part are shown in Table 7.2. While these

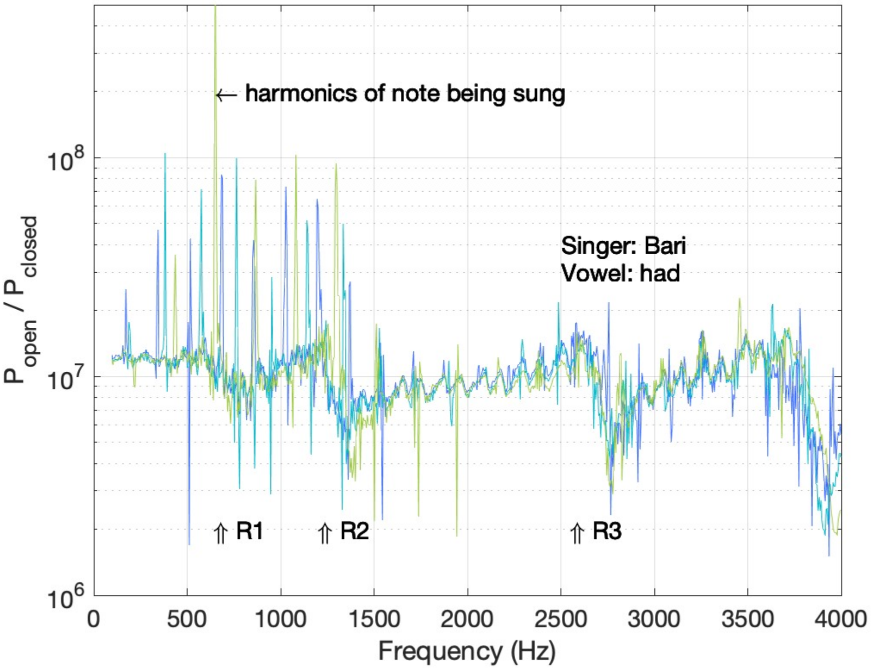


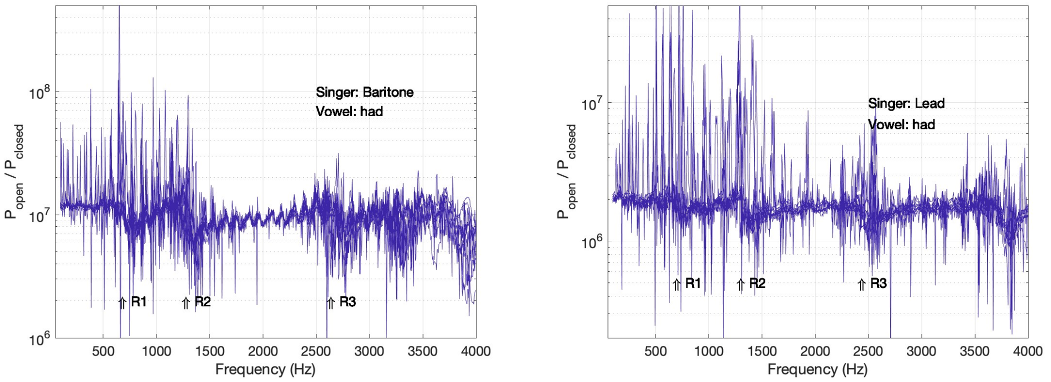
Figure 7.7: RMD results overlaid for the baritone singing F3, G3, A3 on /a/ vowel showing vocal tract resonances at R1,R2,R3

estimated resonances lie within the frequency ranges expected for an /a/ vowel there will be some variation depending on the pitch of the note being sung. With untrained male singers the larynx tends to rise with increasing phonation frequency causing a reduction in pharynx length and a change in the resonant frequencies (Sundberg 1987). In addition there are differences between the voice parts because of the different morphologies of the individuals. The first three resonances are comparable to the values assumed in the computer model but there is also evidence of a "singers formant" in all four voices around 3500 Hz which is more prominent than assumed in the computer model. It is also worth noting the dips (anti-resonances) in the frequency responses just above the resonances. These anti-resonances will have the effect of damping harmonics falling within those frequency ranges.

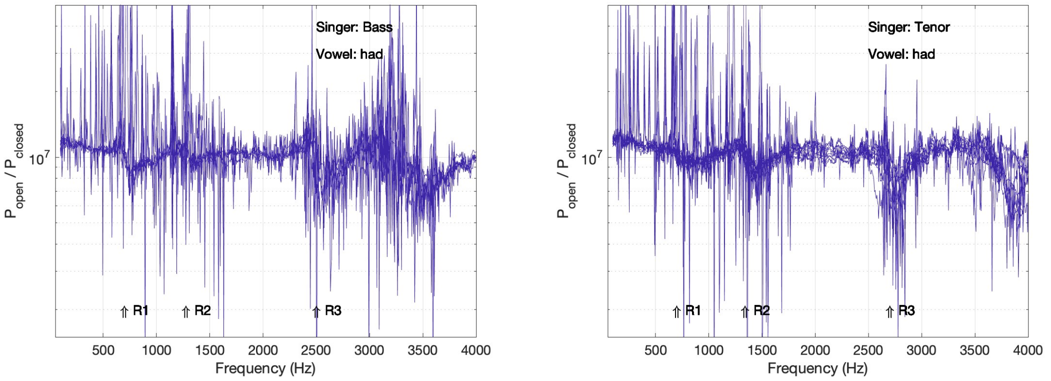
Voice Part	R1	R2	R3	Note Range
Baritone	640	1240	2600	F2 to A3
Lead	660	1260	2400	C3 to E4
Bass	660	1240	2460	G2 to E3
Tenor	760	1300	2660	C3 to D4
Synth quartet	660	1200	2550	

Table 7.2: Vocal tract resonances in Hz for /a/ vowel

The acoustic metrics for the output of each singer when singing the 15-bar exercise were derived from the XY microphone recordings when they sang their part solo. Each singer's SPL (red line),



(a) baritone singing F2 to A3 (b) lead singing C3 to E4



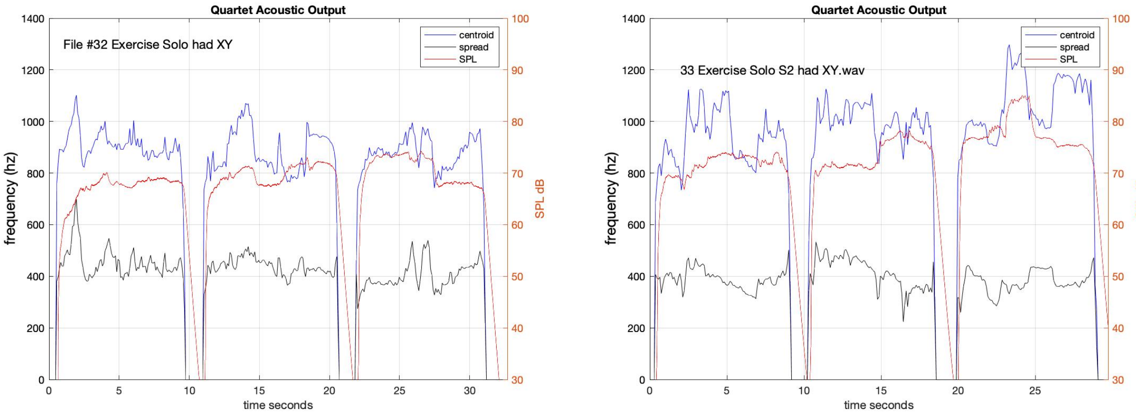
(c) bass singing from F2 to C4 (d) tenor singing from C3 to D4

Figure 7.8: Overlaid results from RMD tests for each voice part as they sang through a vocal range on /a/ vowel. (a) baritone (b) lead (c) bass (d) tenor. The thickening of the lines results from a shift in the resonances with pitch

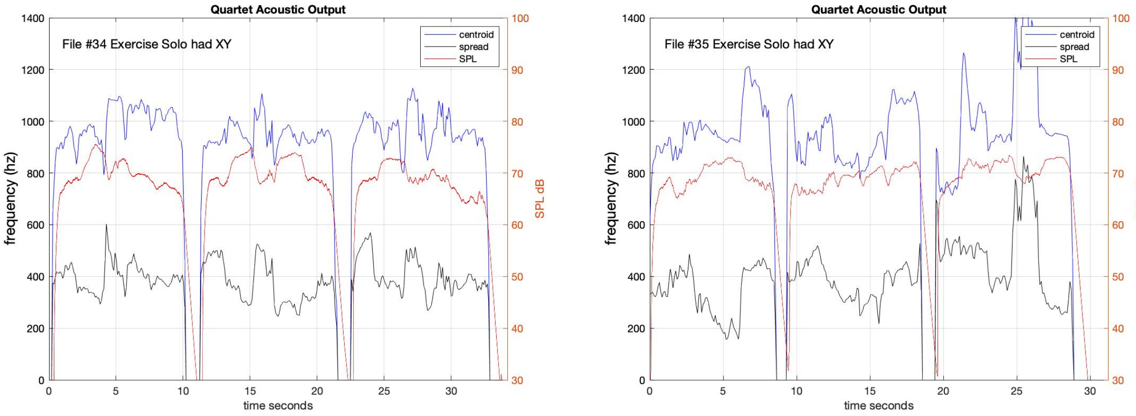
spectral centroid (blue line) and spectral spread (black line) for an /a/ vowel are shown in Figure 7.9. While each singer generally maintained a steady loudness throughout the exercise, as measured by the SPL, there are some noticeable changes in the centroid and spread on specific notes. From the RMD results the /a/ vowel has R1 at 660 Hz and R2 at 1240 Hz for the lead, baritone and bass, and R1 at 760 Hz and R2 at 1300 Hz for the tenor. The large increases in the lead centroid at 24 seconds and in the tenor centroid at 25 seconds occur when they are singing G4 (392 Hz) and A4 (440 Hz) respectively and their 3rd harmonics are close to R2. The spectral spread does not alter significantly throughout the exercise except for the peak in the tenor voice when the third harmonic aligns with R2.

Figure 7.10 shows the results for the /ʊ/ vowel which has R1 at 450 Hz and R2 at 1100 Hz. The large increases in the lead centroid at 5 seconds and in the baritone centroid at 17 seconds are when they are singing A3 at 220 Hz and their 2nd harmonic coincides with R1.

Figure 7.11 shows the results for the /i/ vowel which has R1 at 310 Hz and R2 at 2020 Hz. The



(a) baritone (b) lead

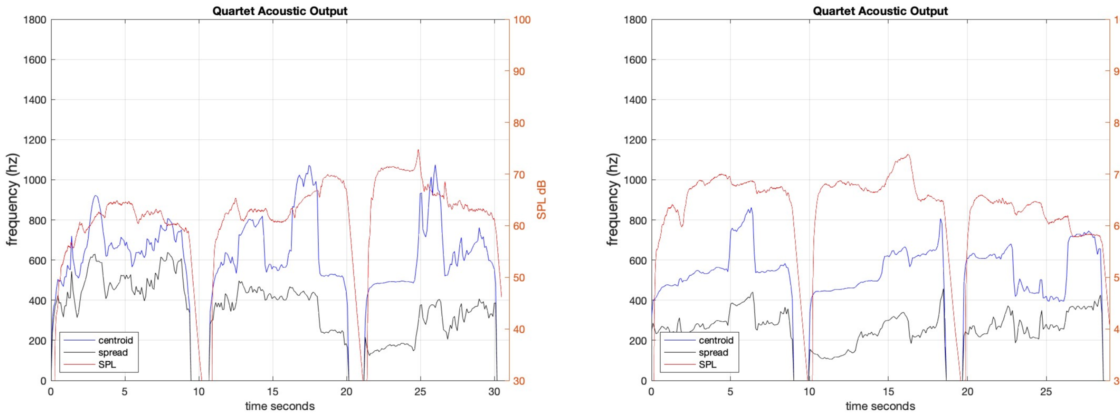


(c) bass (d) tenor

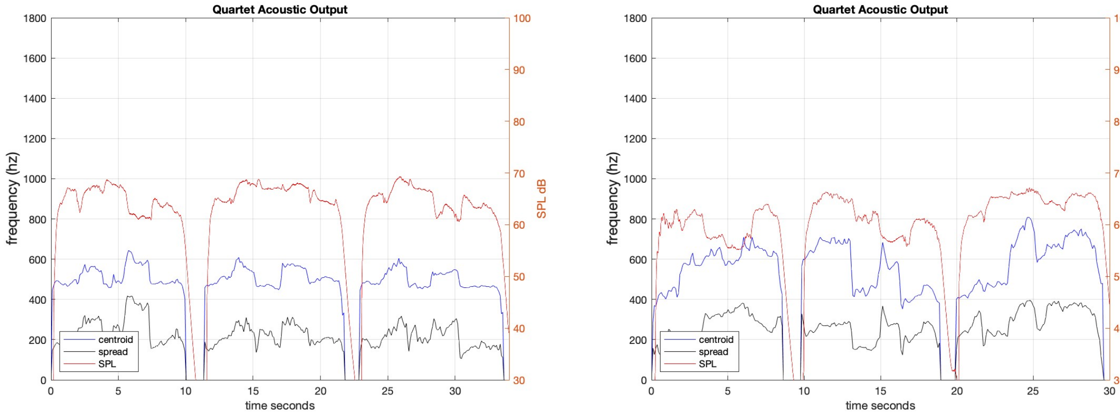
Figure 7.9: Acoustic characteristics of solo voices singing the 15-bar exercise on /a/ vowel (a) baritone (b) lead (c) bass (d) tenor

step in the baritone at 25 seconds and the step in the lead at 18, 21 and 28 seconds are when they are singing D4 at 290 Hz which is in the region of R1.

The impact of a singers harmonics aligning with their vocal tract resonances can also be illustrated by plotting the power spectra for each solo voice against increments of time. The results when singing the exercise on an /a/ vowel are shown in Figure 7.12. The peaks in the power occur when the notes being sung match the tract resonances in the region of R1 (660 Hz) and R2 (1240 Hz). A different power scale was used for the lead voice compared to the other voice parts because of the strength of the lead's harmonics. The incremental spectra when singing *as a quartet* is shown in Figure 7.12 (e). The peaks occur when harmonics in the different voices combine and align with R1 and R2. The comparable power spectra when singing an /ʊ/ vowel and /i/ vowel are plotted against time for the duration of the exercise in Figures 7.13 and 7.14 respectively.



(a) baritone (b) lead



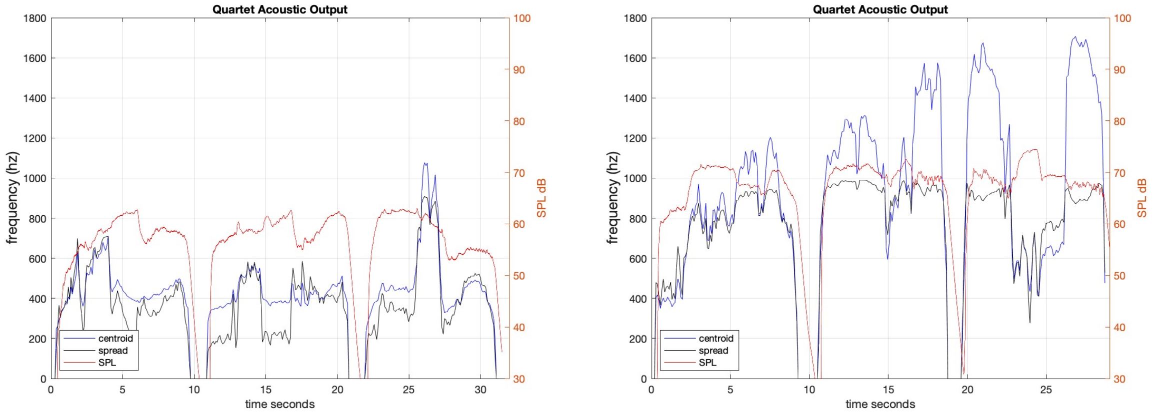
(c) bass (d) tenor

Figure 7.10: Acoustic characteristics of solo singers on /ʊ/ vowel

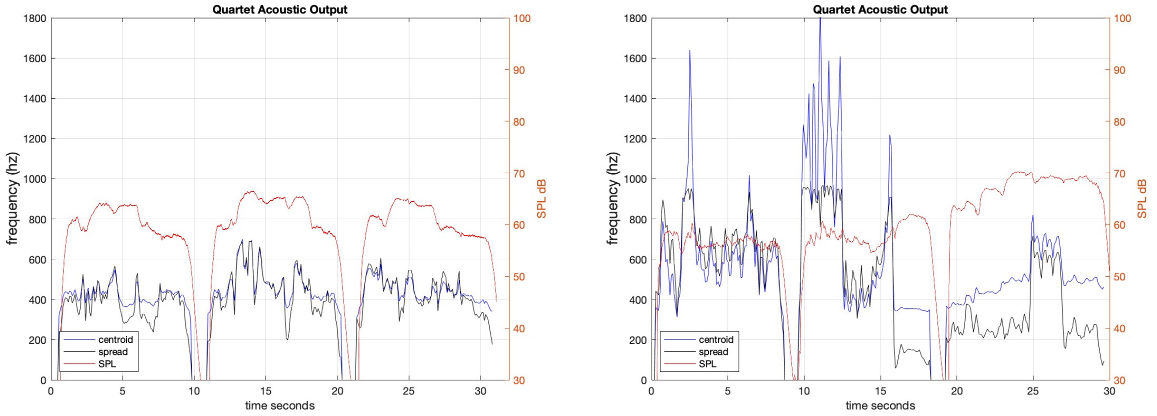
7.2.3 Influence of singer adjustments in real time on acoustic output

Each of the chords in the 15-bar exercise was sustained in a steady state for a relatively short duration (approximately 2 seconds). To assess the impact of singers modifying their vocal production while sustaining chords, samples of the quartet performing repertoire songs were analysed. One example is the last chord in bar 109 at the end of the song “This Could Be the Start of Something Big” as shown in Figure 7.15.

The word “grand” was sustained for 5 seconds. The audio sample extracted for analysis included the end of the chord in bar 108 with the word “grand” starting at 1 second into the sample. For this chord the notes sung by the quartet were not in the conventional bass:baritone:lead:tenor order. The singers revoiced to suit their individual voices with the lead singing above the tenor. Figure 7.16 shows (a) the fundamental frequency of each voice part calculated from the EGG recordings, (b) the intonation accuracy of each voice in cents relative to a target note in just intonation with the lead



(a) baritone (b) lead



(c) bass (d) tenor

Figure 7.11: Acoustic characteristics of solo singers on /i/ vowel

note, (c) the open quotients of the individual singers and (d) the centroid, spread and SPL of the full quartet. From Figure 7.16 (a) it can be seen the lead is singing the highest note at 591 Hz in full voice, above the tenor at 470 Hz singing falsetto. The harmony parts quickly tuned accurately to within ± 20 cents of their target note (Figure 7.16 (b)). The fundamental frequency, open quotient and intonation accuracy of all voice parts were very stable from 2 to 5 seconds while the vowel /æ/ was sung. However the centroid after reaching 1300 Hz (E6) at 2 seconds then trended down to 1100 Hz (C6♯) through this period. The spread reached 530 Hz at 2 seconds then trended down to 470 Hz through the period. The SPL reached 96 dB at 2.5 seconds then reduced slightly to 93 dB thereafter (Figure 7.16 (d)). From 5 seconds onwards the intonation accuracy reduced and the open quotients and acoustic metrics changed as the singers transitioned from the /æ/ vowel to “nd” at the end of the word, even though the acoustic intensity was maintained.

The time averaged fine resolution line spectrum for the central period of the chord is shown in Figure 7.17. For the chord being sung (slightly lower than the chord as scored B3-B4-F♯4-D♯5) the distribution of acoustic power can be seen in the cumulative line spectrum in Figure 7.17). The overlaid blue line is the gain envelope for the vocal tract resonances used in the computer model which shows the indicative values of R1 at 620 Hz and R2 at 1660 Hz for an /æ/ vowel. The strongest partials in the chord are the 4th, which is the lead fundamental on 590 Hz with 16% of the energy, the 10th which is the bass' 7th harmonic on 1657 Hz with 14% of the energy and the 11th which is a combination of the lead's 3rd harmonic with the baritone's 5th harmonic on 1773 Hz with 23% of the energy. The 4th partial is close to R1 and the 10th and 11th partials are close R2. While the absolute peak of the 10th partial is greater than the 11th partial there is more energy concentrated around the 11th partial as is evident in the cumulative line spectrum. The peak in the 11th partial is lower because the lead and baritone are not precisely tuned.

An incremental power spectrum for the sample as it progresses in 50 msec increments is plotted against time in Figure 7.18. The power concentrated around R2 (1660 Hz) is very apparent at 2 seconds. From 2 seconds onwards an increase in power is evident around R1 (620 Hz).

For the relatively short duration of this chord sequence each singer would be able to sustain their total acoustic power production but a redistribution of energy from R2 to R1 would cause the spectral centroid and spread to reduce. This shift, which could result from an adjustment of any of the singer's vocal tract resonances (indicated by a movement in the peaks of the blue line in Figure 7.17), or slight divergence from accurate tuning in combined harmonics, would result in a change in the perception of the timbral qualities and the loudness of the chord with time. It would therefore seem feasible that the converse of raising the centroid could be realised by redistributing energy the other way, from the lower to the higher resonance. This detailed analysis of one repertoire chord illustrates how small changes in intonation accuracy and resonance tuning can change the overall acoustics of a quartet sound in real time as the singers make adjustments when performing together and could be a plausible explanation for the perception of Expanded Sound.

7.3 Discussion

7.3.1 Effect of compositional choices on the ensemble sound

The results from the quartet singing the 15-bar exercise showed a very clear difference in the timbral metrics of spectral centroid and spread between the /a/, /i/ and /ʊ/ vowels over the full range of chord structures and chord root frequencies. The /a/ vowel had a centroid around 1000 Hz and the /ʊ/ vowel had a centroid around 500 Hz throughout the 15-bars, while the centroid for the /i/ vowel varied over a range of 400 Hz to 1000 Hz. The spread for the /a/ and /ʊ/ vowels was steady at 400 Hz and 300 Hz respectively while the /i/ vowel spread ranged between 400 Hz and 900 Hz. These behaviours are consistent with the results from the archive samples and the computer model predictions and support the conclusion that the overall timbral metrics of the quartet are strongly influenced by the degree of alignment between the harmonics and the vocal tract resonant frequencies of the individual singers.

The measured centroid and spread values realised in the tests were numerically higher than the predictions from the computer model. The divergences may be explained by (a) the differences between each singer's voice characteristics and the corresponding input parameters used in the computer model and (b) an underestimation of the contribution made by the higher harmonics which are reinforced by the singers' formant.

The sound intensity, as measured by the SPL (A-weighted) level, of the /a/ vowel was consistently higher than the level of the /ʊ/ and /i/ vowels throughout the exercise. The A-weighting reduces the strength of harmonics below 1000 Hz, following the inverted shape of the 40 dB equal-loudness contour. The greater SPL level for the /a/ vowel is because the strong harmonic combinations in the voice parts were in the region of R1 (660 Hz) and R2 (1200). In contrast the lower harmonics in the region of R1 (450Hz) for an /ʊ/ vowel and R1 (310Hz) for an /i/ vowel are diminished by the weighting. In the computer model, which did not apply frequency weighting, the absolute power of the /u/ vowel was shown to be greater than the /a/ vowel and the /i/ vowel.

The results reinforce the conclusion that the compositional selections of chord structure, root frequency and vowel which align voice harmonics with the vocal tract resonances will create the best conditions for singers to maximise the acoustic features of high spectral centroid, spectral spread and acoustic power.

7.3.2 Effect of individual singer's vocal characteristics on the ensemble sound

The results of the 15-bar exercise when sung solo by each voice part illustrates how the vocal characteristics of each singer can affect the quartet sound on each of the chords. For example, for an /a/ vowel the vocal tract resonances R1 and R2 are around 660 Hz and 1200 Hz respectively, so the phonation frequencies of all four voices parts would lie below R1 throughout the full 15 bars. However when a note being sung by one of the voice parts has higher harmonics matching the tract resonances, that singer's contribution has most impact. While the overall SPL is maintained the net result is there will be step changes in the centroid and spread, with a corresponding change in the overall sound.

These results for the individual voices demonstrate the impact matching the tract resonances with the harmonics has on a singer's vocal characteristics and thereby their contribution to the ensemble sound. The effects will be even more pronounced when the higher harmonics in more than one voice match.

7.3.3 Effect of singer's adjustments on the ensemble sound

Each of the chords in the 15-bar exercise had a relatively short duration of around two seconds. In live performances chords are frequently sustained for longer durations which gives the singers the opportunity to make vocal adjustments in real time as they react to the sound being produced. The analysis of the repertoire chord showed this effect. A redistribution of the energy in the region of the tract resonances caused the spectral centroid, spread and SPL to alter. This shift would result in a change in the perception of the timbral qualities and the loudness of the chord with time. This analysis of one repertoire chord illustrates how small changes in intonation accuracy and resonance tuning can change the overall acoustics of a quartet sound in real time when singers are performing together and could be a plausible explanation for the perception of Expanded Sound.

7.4 Limitations

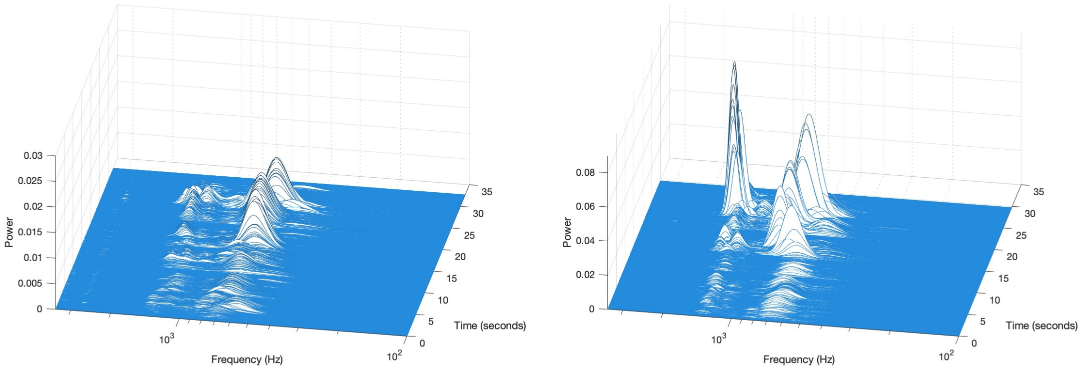
When evaluating the contribution the individual singers made to the quartet sound the recordings from the face-mounted DPA microphones were not used as they were found to underestimate the higher frequency components in a singer's sound as a result of directivity. In their place recordings were taken using the XY microphones with the singers performing the exercise solo. Consequently any modifications the singers made when performing as a quartet would mean the solo recordings may not be true representations of the ensemble singing.

7.5 Summary

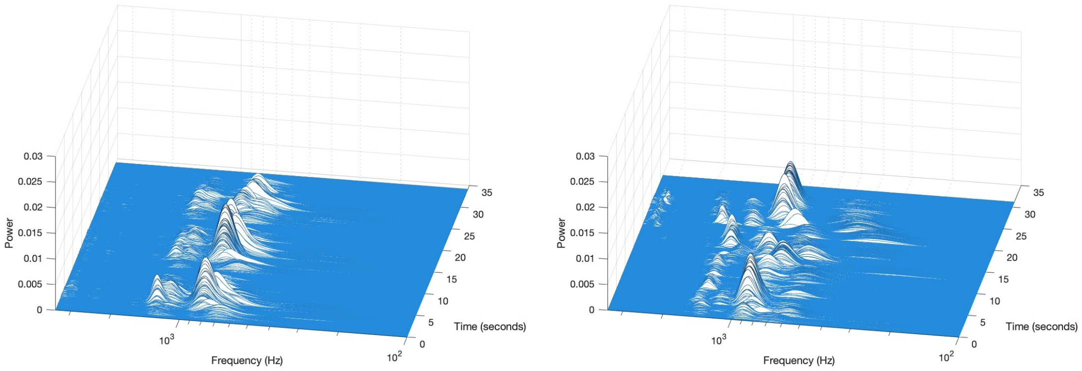
The analysis of the archive database samples demonstrated that a wide range of the timbral characteristics of spectral centroid and spread can be produced by a quartet, with the chord structure, chord root frequency and vowel being determining factors. The parametric studies using the synthesised quartet model showed similar results to the archive samples but generally predicted lower values of the metrics for the corresponding chords, vowels and pitches. The results from the test programme with a quartet under controlled laboratory conditions validated these conclusions for the centroid and spread and also demonstrated the acoustic intensity is influenced by the same variables. Differences in the absolute values of the metrics between the computer model and the archive and laboratory tests can be attributed to singers possessing different vocal characteristics (glottal power, glottal opening periods, vocal tract resonances) which the laboratory tests demonstrated differed between singers and the computer model. The analysis of sustained chords demonstrated singers do make vocal adjustments when performing (intentionally or not) which affect the intensity and spectral properties of the ensemble's acoustic output.

Overall it is concluded that for a given vocal power input by each singer in an ensemble, a primary factor determining an ensemble's acoustic character, as measured by intensity, centroid and spread, is the extent of alignment of vocal tract resonances with the harmonics in the individual voices. Being able to maximise these characteristics will be affected by the choice of chord structure, root frequency and vowel sung. Combinations of the matching higher harmonics in multiple voices can contribute a significant percentage of the total energy. Accurate tuning between the voice parts in just intonation is a necessary condition to optimise the effects.

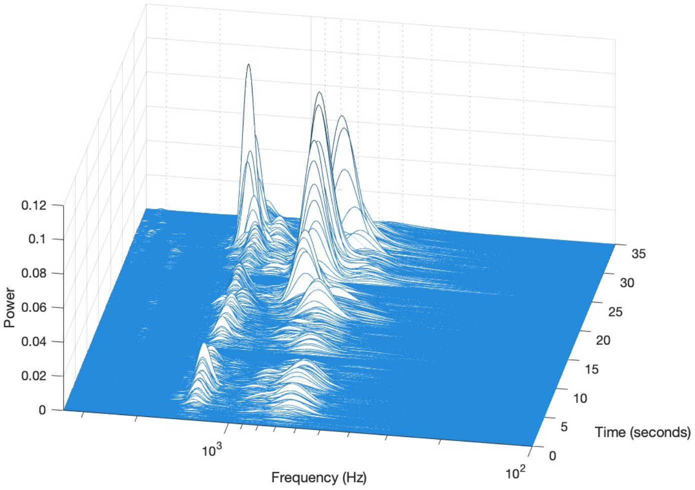
Having identified a set of objective acoustic measures to represent Barbershop quartet singing and determined the primary factors affecting these metrics the question remains whether these absolute acoustic properties correlate with the listener's perception which is the subject of the next chapter.



(a) baritone (b) lead

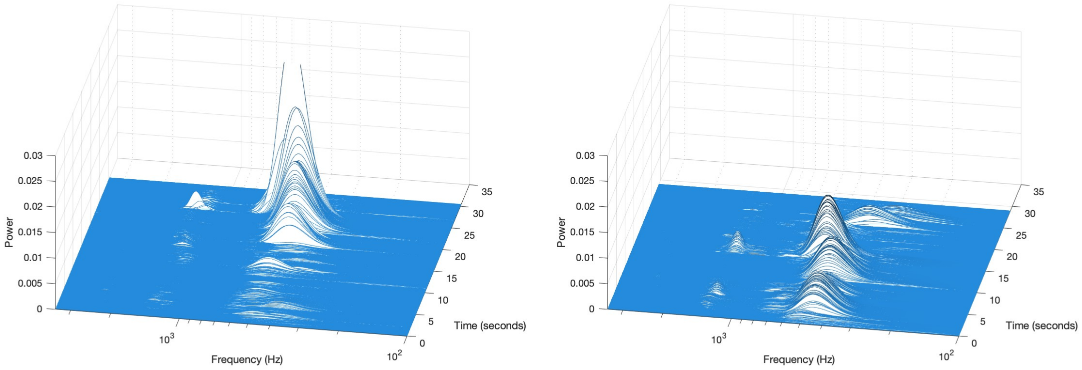


(c) bass (d) tenor

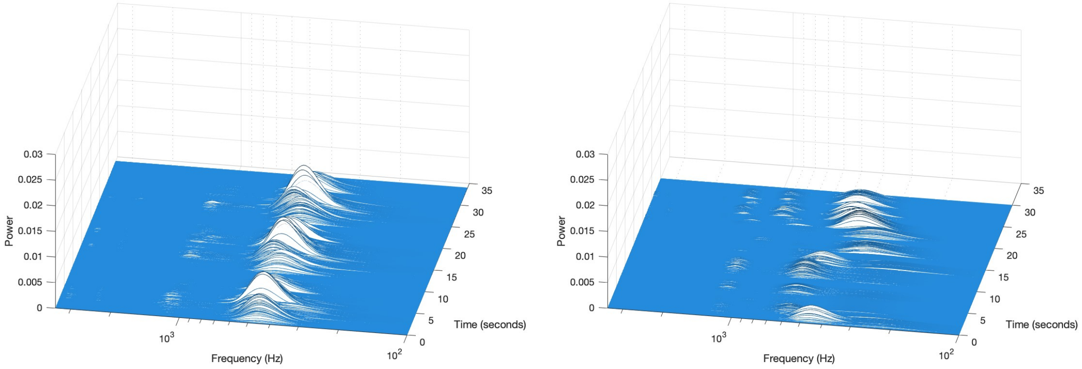


(e) full quartet

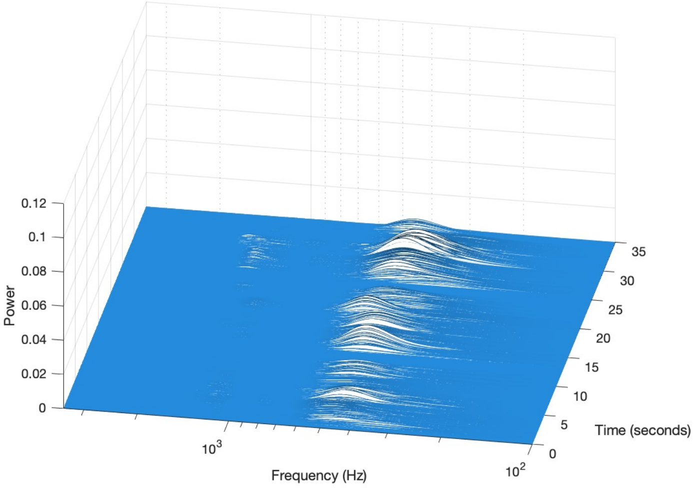
Figure 7.12: 15-bar exercise on /a/vowel: power spectrum against time for the solo singers (a) baritone (b) lead (c) bass (d) tenor and (e) full quartet



(a) baritone (b) lead

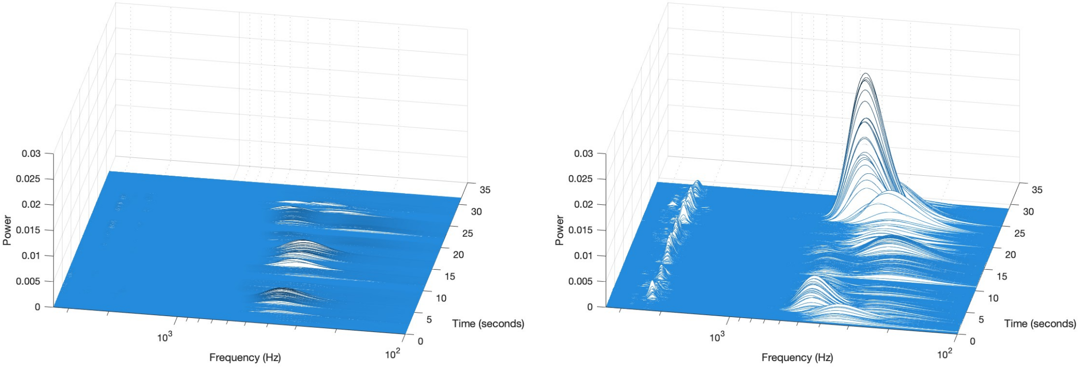


(c) bass (d) tenor

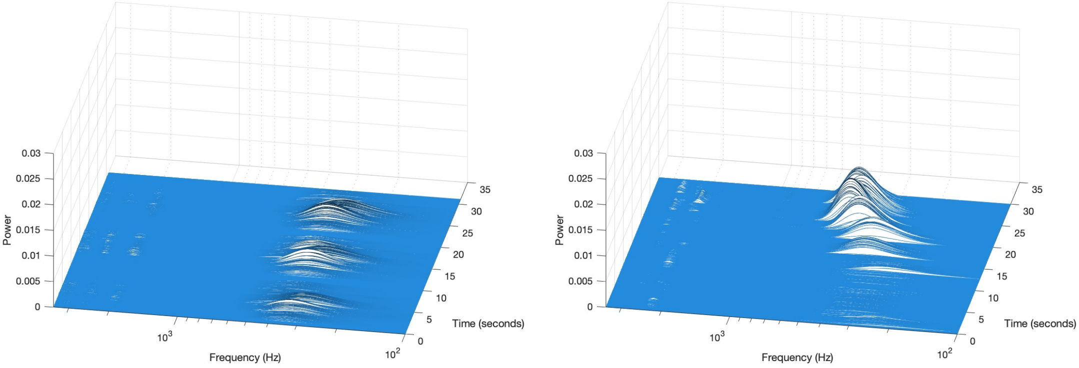


(e) quartet

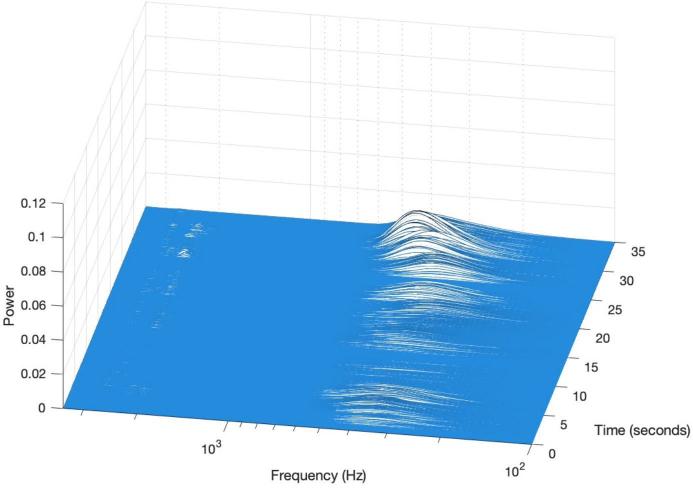
Figure 7.13: 15-bar exercise on /ʊ/vowel: power spectrum of solo singers (a) baritone (b) lead (c) bass (d) tenor and (e) full quartet



(a) baritone (b) lead



(c) bass (d) tenor



(e) quartet

Figure 7.14: 15-bar exercise on /i/vowel: power spectrum of solo singers (a) baritone (b) lead (c) bass (d) tenor and (e) full quartet

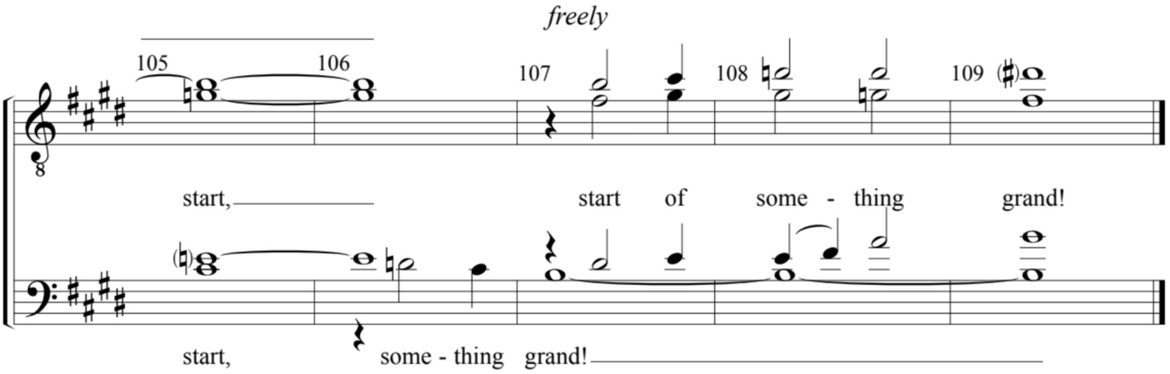
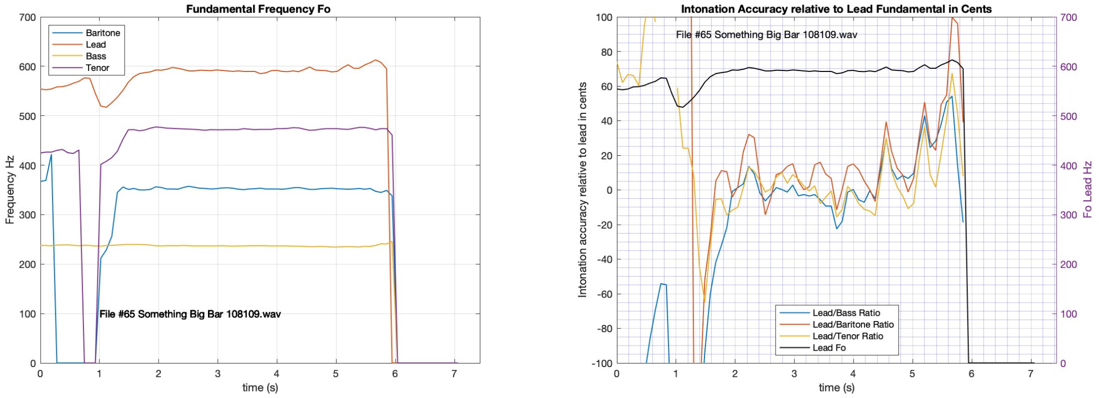
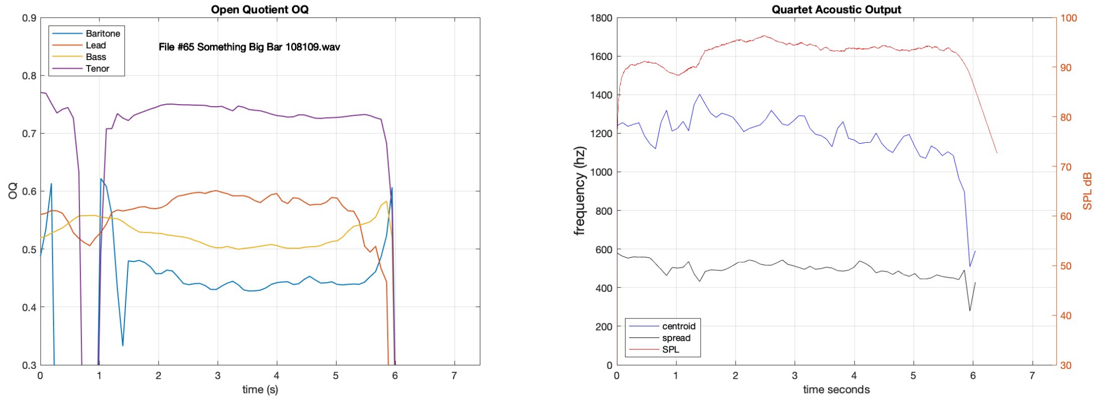


Figure 7.15: Chord sequence at the end “Start of Something Big”



(a) Fundamental Frequency

(b) Intonation Accuracy



(c) Open Quotient

(d) Acoustic Metrics

Figure 7.16: Acoustic measures for bars 108/109 of song “Something Big”

(h)

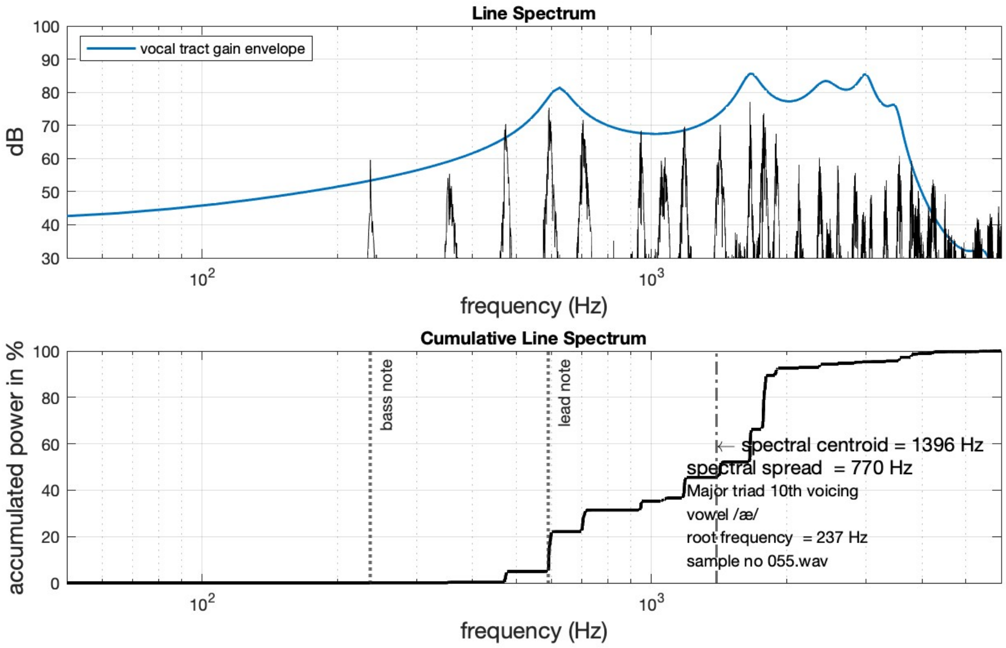


Figure 7.17: Line and cumulative line spectra for "Something Big" Bar 108/109 overlaid with the vocal tract envelope for an /æ/ vowel from the computer model (blue line)

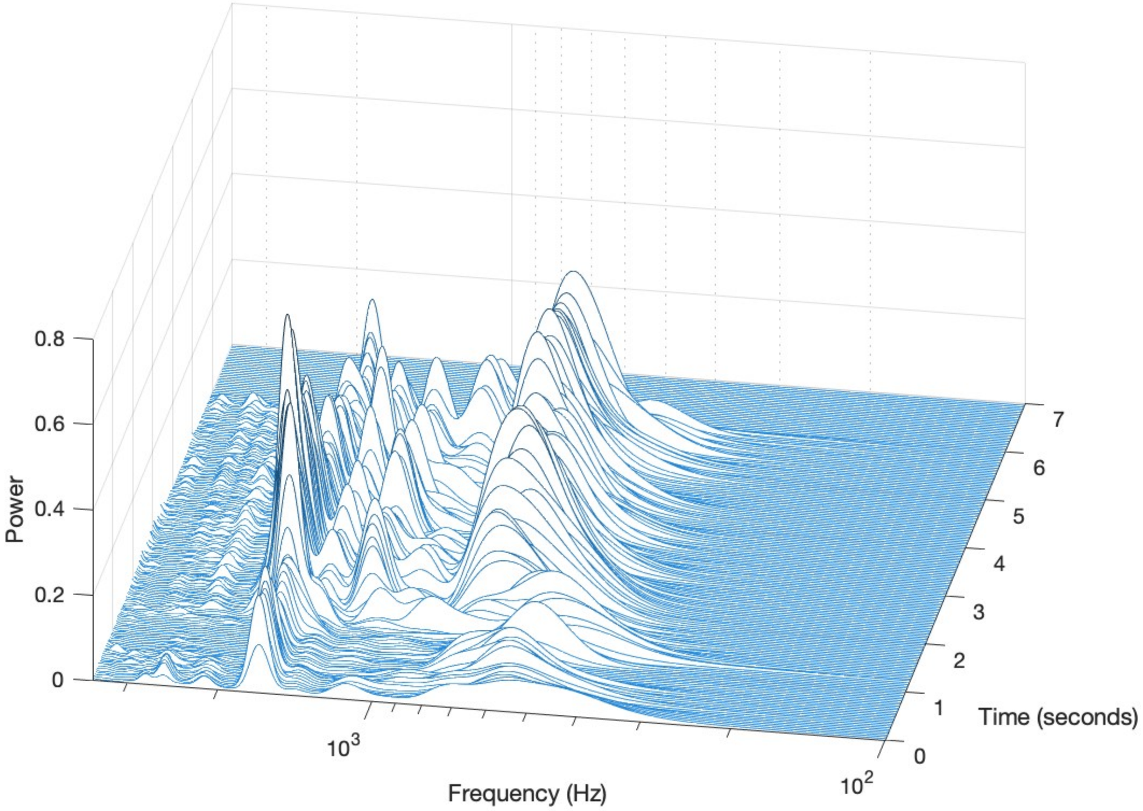


Figure 7.18: Power spectrum vs time for "Something Big" Bar 108/109

Chapter 8

Perception Tests

The previous chapters focused on *absolute measurements* of the acoustic characteristics of Barbershop singing and explored the factors which affect those measures. The question remains whether these metrics are the physical correlation of the *perceptual descriptions* commonly used for the sound produced. Some inconsistencies are to be expected because the acoustic metrics are calculated relative to a fixed reference datum across the frequency range whereas a listener's hearing and perception of loudness varies with frequency. Furthermore, the computed spectral metrics are time averaged values whereas the sound arriving at a listener's ears varies continuously with time. The temporal dimension is an important factor in listener perception because (a) phase differences between the contributing voices could be influential and (b) singers may react to what they hear and adjust their vocal techniques in real time so chords are never perceived in a truly steady state.

To investigate the relationship between absolute measures and subjective descriptions perception tests were conducted with trained Barbershop judges as the participants. The survey specifically sought to compare the objective metrics derived from the chord analysis described in the previous chapters with the idiomatic expressions of "Lock", "Ring" and "Expanded Sound" commonly used when describing Barbershop singing.

The testing protocol described below was approved by the Ethics Committee at the University of York. Approval reference Cookson20240322.

8.1 Methodology

8.1.1 Audio sample database

For the perception tests a set of 50 audio samples was assembled, drawn from the archive database of real quartets, the recordings of Sound Hypothesis performing under laboratory conditions and from synthesised computer chords. The duration of each test sample was a minimum of 2 seconds. The database included samples which were out of tune, samples with an audible lead-in chord sequence and samples of the same chord sung by different quartets. The synthesised samples were constructed using the source/filter model described in Chapter 6 for each voice part and then adding the amplitude of each voice at each sample interval in the time domain. As there was no reference for calibrating the SPL levels of the samples they were normalised using DAW "GarageBand" to give equal peak volumes then analysed for the value of spectral centroid, spectral spread and intonation accuracy. The spectral centroid and spread of each sample were computed using the methods described in Chapter 5. Quantifying the intonation accuracy of the samples used in the perception tests was an adaptation of the method used for the solo voice. The deviation in tuning 'c' for each voice part was computed in cents as:

$$c = 1200 * log_2 \left[\frac{F_h}{r * F_l} \right] \tag{8.1}$$

where F$_h$ is the fundamental frequency being sung by a harmony part, determined from measuring the peaks in the chord's spectrum and 'r' is the target frequency ratio relative to the lead's fundamental frequency F$_l$ in just intonation. The computed value of 'c' for each voice in each chord sample is shown in Figure 8.1. The actual frequency sung can be above or below the ideal frequency. A simple measure of the overall intonation accuracy of a chord was taken to be the intonation spread 'R' in cents computed as the maximum deviation by a voice part above its ideal frequency plus the maximum deviation by a voice part below its ideal frequency. The details of each sample in the perception test database and their values of centroid, spread and R are given in Table 8.1. Samples with a value of R greater than 40 cents were classified as being "out of tune". The cut off at 40 cents was selected as the midway point between perfect intonation and the "most dissonant" interval between two frequencies which occurs when the frequency difference is around a quarter of the critical bandwidth (Rasch and Plomp 1999). The most dissonant frequency difference is about 20 Hz in low frequency regions and a little less than a semitone at frequencies above 500 Hz. The "out of tune" samples (nos 5, 17, 25, 31, 33, 39, 49) using this criterion are highlighted in magenta in Figure 8.1 and Table 8.1

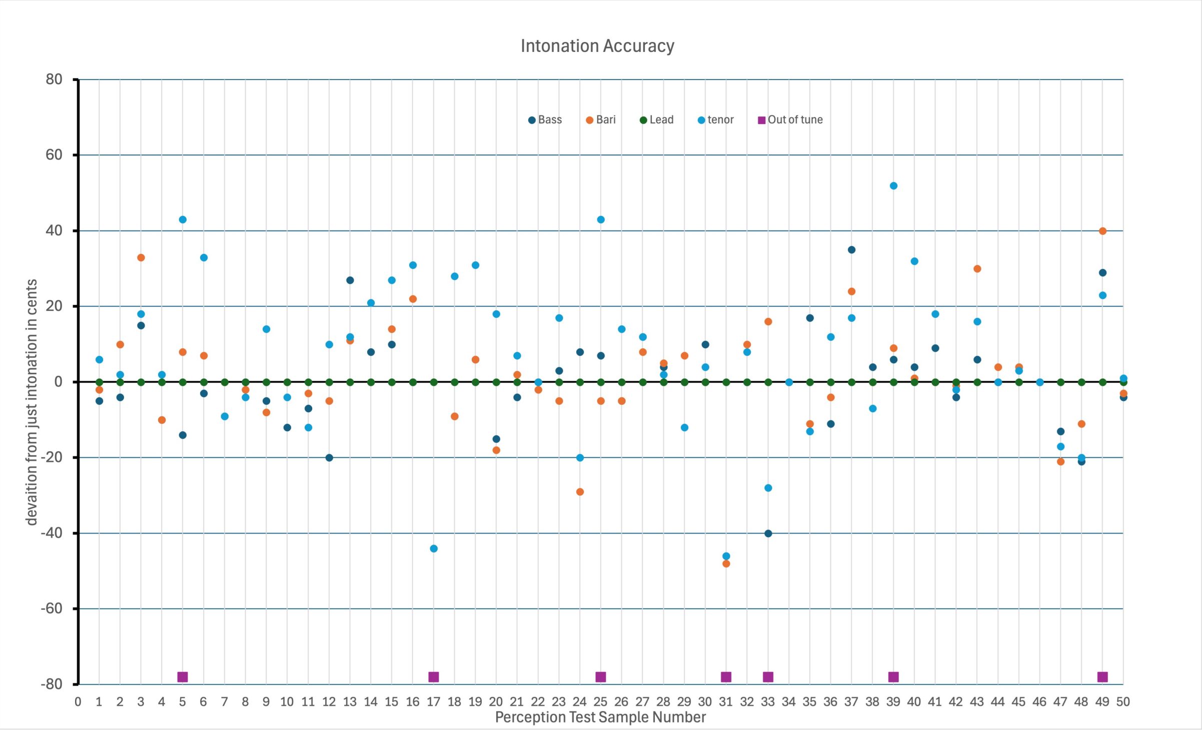


Figure 8.1: Intonation accuracy of each voice part in 50 test samples measured in cents relative to the target frequency for a note tuned to the lead in just intonation. Samples with an intonation spread > 40 cents, classed as "out of tune", are identified by magenta squares

Prior to interpreting the perception test responses the relationship between the spectral centroid and the spectral spread was checked to see if they act as independent outputs. The spectral spread is plotted against the spectral centroid for each sample in the database in Figure 8.2 with the least squares fit plotted as the blue line. The figure shows that while there is scatter, in general an increase in spectral spread will occur as the centroid value increases.

8.1.2 Test participants

Ten participants took part in the survey. All the participants are singing and musicality judges from the British Association of Barbershop Singers and the Ladies Association of British Barbershop Singers judging panel. The demographics of the participants are: 9 certified judges and 1 trainee judge, 8 male and 2 female, 4 aged 20-39 years, 1 aged 40-59 years, 5 60+years. All had normal hearing and all used headphones when completing the survey. The participants answers to the survey were presented anonymously.

Sample No	Voicing	Freq Ratios	Centroid Hz	Spread Hz	R cents	Comments
PT1	BBS7th spread voicing	2:5:3:7	242	299	11	
PT2	major triad 12th voicing	2:5:4:6	1402	1203	14	
PT3	BBS7th medium voicing	4:7:6:10	768	948	33	
PT4	Major triad 10th voicing	2:3:4:5	1023	652	12	
PT5	BBS7th med voicing Bass 5th	3:5:4:7	397	253	57	
PT6	BBS7th close voicing	4:5:6:7	408	511	36	
PT7	Major triad 10th voicing	2:3:4:5	1174	583	9	
PT8	Major triad 10th voicing	2:3:4:5	1372	804	4	Sound Hypothesis “Grand”
PT9	Minor 6th (alt)	6:9:10:14	1270	956	22	
PT10	Major triad 12th voicing	2:4:5:6	1239	923	12	
PT11	Major triad 2 octave voicing	2:5:6:8	243	87	12	
PT12	Minor 6th (alt)	6:10:9:14	342	320	30	
PT13	BBS7th close voicing	4:5:6:7	823	454	27	
PT14	BBS7th med voicing Bass 5th	3:5:4:7	1301	1306	21	
PT15	BBS7th medium voicing	4:7:6:10	470	604	27	
PT16	BBS7th close voicing	4:5:6:7	317	173	31	lead in chord sequence
PT17	Major triad 10th voicing	2:3:4:5	345	154	44	synth out of tune
PT18	Minor 6th (alt)	6:9:10:14	1369	923	37	
PT19	BBS7th med voicing Bass 5th	3:5:4:7	1045	591	31	
PT20	Minor 6th (alt)	6:9:10:14	1191	856	36	
PT21	Major triad 10th voicing	2:3:4:5	1081	536	7	
PT22	Major triad 12th voicing	2:5:4:6	1222	732	2	
PT23	BBS7th medium voicing	3:4:5:7	150	1272	22	
PT24	BBS7th medium voicing	4:6:7:10	886	583	37	
PT25	BBS7th medium voicing	4:7:6:10	738	890	43	
PT26	Major triad 10th voicing	2:3:4:5	1006	668	19	
PT27	Major triad octave voicing	4:6:5:8	504	580	12	
PT28	BBS7th medium voicing	4:7:6:10	548	766	5	
PT29	Major triad 10th voicing	2:3:4:5	302	203	12	
PT30	Major triad 10th voicing	2:3:4:5	1829	1432	10	
PT31	Major triad 12th voicing	2:5:4.097:6	1210	727	48	synth out of tune
PT32	Major triad 10th voicing	2:3:4:5	281	118	10	
PT33	BBS7th spread voicing	2:3:5:7	289	204	56	
PT34	Major triad 10th voicing	2:3:4:5	345	149	0	synth perfect tuned
PT35	Major triad 10th voicing	2:3:4:5	335	339	30	
PT36	Major triad spread voicing	2:4:5:6	328	211	23	
PT37	Major triad octave voicing	4:5:6:8	328	314	35	
PT38	Major triad 12th voicing	2:5:4:6	868	490	11	
PT39	minor triad 10th voicing	6:9:12:14	1045	703	52	
PT40	minor triad 10th voicing	6:9:12:14	1317	865	32	
PT41	Major triad 10th voicing	2:3:4:5	1207	689	18	
PT42	Major triad 10th voicing	2:3:4:5	1143	797	4	
PT43	Major triad 2 octave voicing	2:5:6:8	294	163	30	
PT44	minor triad 10th voicing	6:9:12:14	810	703	4	
PT45	Major triad 10th voicing	2:4:3:5	498	317	4	
PT46	BBS7th close voicing	4:5:6:7	353	135	0	synth perfect tuned
PT47	BBS7th close voicing	4:6:5:7	404	426	21	
PT48	minor 6th	5:6:7:9	1284	711	21	
PT49	BBS7th medium voicing	4:7:6:10	438	561	40	
PT50	Major triad 12th voicing	2:5:4:6	1302	806	5	same chord as PT2

Table 8.1: Database of perception test samples. Out of tune samples highlighted in magenta

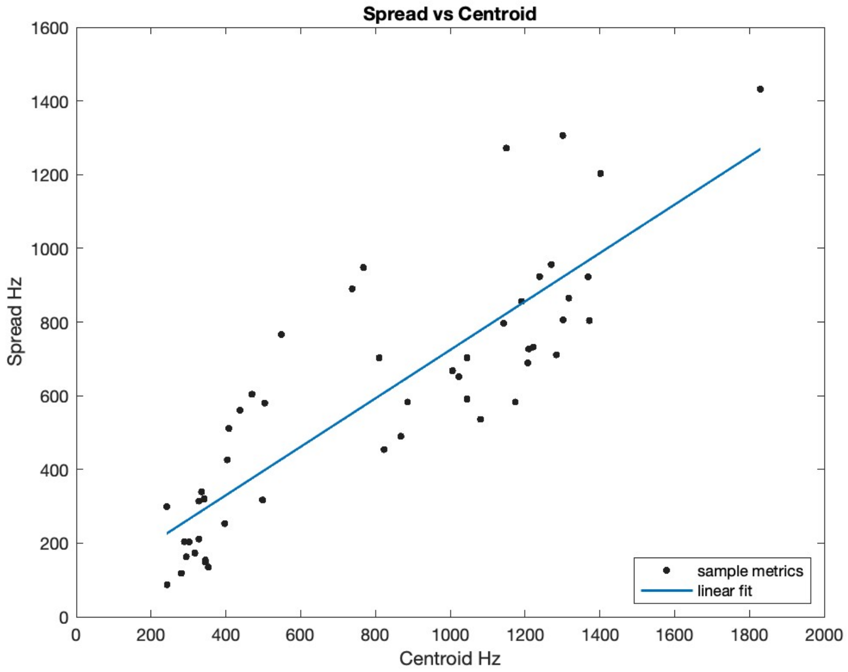
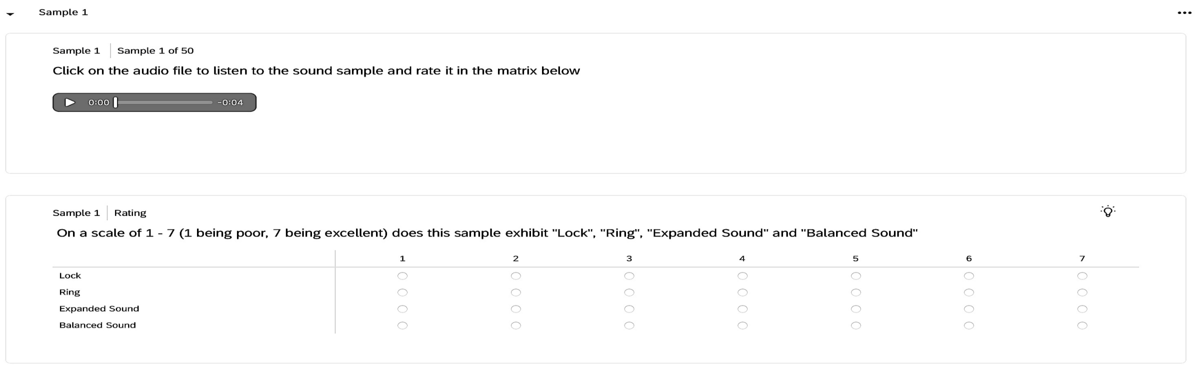


Figure 8.2: Spread vs Centroid

8.1.3 Test procedure

The survey was carried out using Qualtrics, a commercial cloud-based platform, which offers applications for collecting and analysing data. Access to the software was provided by the University of York. The 50 audio test samples were uploaded to a questionnaire on the survey platform and access to the survey was provided to participants through an internet link. The participants were firstly asked to define their understanding of the concepts of Lock, Ring and Expanded Sound. The participants were then asked to listen to an audio sample and set the volume of their headphones or speakers to a level which was comfortable and not to adjust the setting once they had started the survey. They were then asked to listen to each audio sample in turn and rate it for each category of Lock, Ring and Expanded Sound on a scale of 1 (poor) to 7 (excellent). A response panel for one sample is shown in Figure 8.3. They were able to return to a sample and edit their response until such time as they submitted the survey as complete. The results of the survey were processed and reported using Matlab functions to calculate the median, 25th and 75th percentiles and mean values of the responses for each of the 50 samples.



Sample 1

Sample 1 | Sample 1 of 50

Click on the audio file to listen to the sound sample and rate it in the matrix below

Sample 1 | Rating

On a scale of 1 - 7 (1 being poor, 7 being excellent) does this sample exhibit "Lock", "Ring", "Expanded Sound" and "Balanced Sound"

	1	2	3	4	5	6	7
Lock	○	○	○	○	○	○	○
Ring	○	○	○	○	○	○	○
Expanded Sound	○	○	○	○	○	○	○
Balanced Sound	○	○	○	○	○	○	○

Figure 8.3: Response form for online perception test sample survey

8.2 Results

8.2.1 Participants' understanding of idiomatic descriptions

Each participant's interpretation of the descriptors Lock, Ring and Expanded Sound are reproduced verbatim in Table 8.2. The participants are identified by number 1 through to 10. The participants primarily associated Lock with accurate vertical tuning in just intonation but with an additional perceptual sense of a unified/integrated/fused sound. Ring was strongly associated with the creation of overtones, leading to a sense of "brilliance" in the sound. Expanded Sound was associated with a sense that the combined sound is greater than the four component voices, which some of the respondents went on to attribute to the creation of overtones and undertones.

8.2.2 Participants' Rating of Chord Samples

The ratings, on a scale of 1 (poor) to 7 (excellent), given by the participants for each of Lock, Ring and Expanded Sound for the 50 samples are shown in Figure 8.4. The *median* value for each sample is shown by the black dot and the lower and upper quartiles are shown by the blue bar. Samples classified as "out of tune" are identified by a magenta star. Major triad chords are identified by a red circle, Barbershop 7th chords by a blue square and minor 6th chords by a green diamond. For this database using different chord structures did not appear to have a strong influence on the ratings.

Figure 8.5 shows the *mean* Lock rating given by the test participants plotted against R for each test sample. The blue line is a least squares fit for the data and indicates a decreasing rating of Lock as the intonation deviation range R increases. Seven of the samples (numbers 5,17,25,31,33,39,49) have a value of R greater than 40 cents and in the following discussion these are classed as having "poor

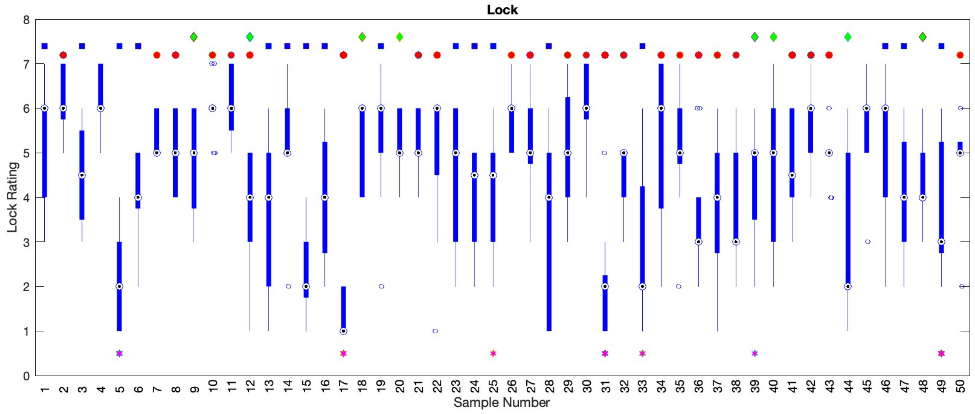
Lock
1. Justly tuned and well balanced chords, with matched sound
2. Highly accurate intonation (in line with the harmonic series) and vowel/resonance matching such that the chord feels much more like one sonic object rather than four constituent notes.
3. Vertical alignment of tuning chords
4. Stable ensemble, principal result of just intonation
5. Perception of a chord where the constituent notes are almost not discernable
6. The buzzing sound quality that arises when voices sing justly tuned consonant intervals (with a closely aligned harmonic series)
7. Yes
8. When the notes of a chord are accurately sung in such a way as it then becomes challenging to deviate from the note, intonation, and vowel choice
9. Chords with accurate harmonic intonation and matched vowel sounds such that the chord sounds unified
10. Similar to ring

Ring
1. The reinforcement of overtones from multiple voices singing locked chords
2. A pleasing resonance in the voice that allows it to carry
3. Creation of overtones
4. Low-order harmonics reinforcing each other to generate prominent overtones
5. Perception of brilliance
6. The ringing sound quality that arises when voices sing together with well-matched vocal timbres (closely aligned formant structures)
7. yes
8. The presence of reinforced overtones
9. Creation of overtones
10. Similar to Lock

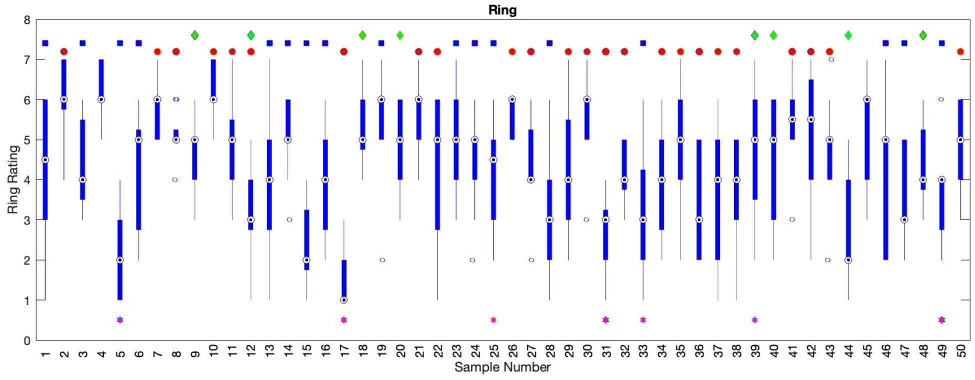
Expanded Sound
1. The perception that there are more than four voices contributing to the sound of a quartet, due to reinforced overtones.
2. The consistent presence of audible overtones in an ensemble sound, often due to mutually-reinforced harmonics from more than one sung note.
3. Reinforcement and sum of overtones
4. The interaction of harmonics producing a lustrous, spatial quality to the sound.
5. Perception that a greater number of voices are contributing to the sound
6. The chords produced by the ensemble are rich in overtones and undertones
7. yes
8. When notes are sung perfectly in tune and the upper harmonics are reinforced, creating a sounds that is perceived as greater than the number of singers
9. The effect in which the sound is greater than the component voices

10 Created overtones

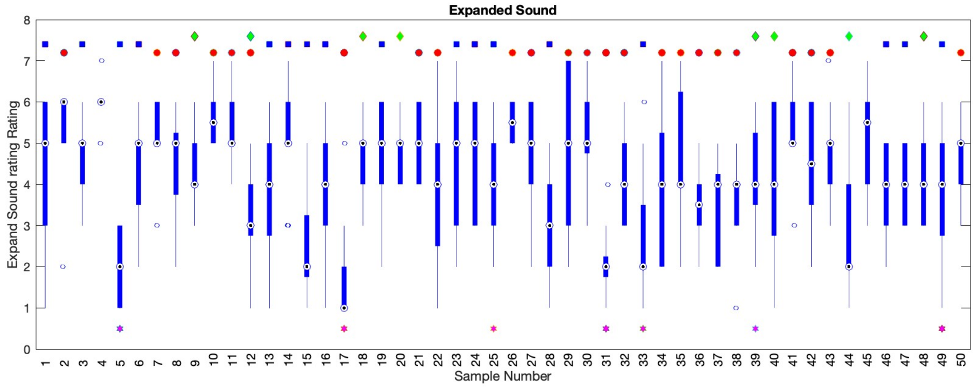
Table 8.2: Verbatim response of each survey participants' interpretation of the idiomatic descriptors



(a) Lock



(b) Ring



(c) Expanded Sound

Figure 8.4: Rating of chord samples on a scale of 1 - 7. "out of tune" samples identified by magenta star. Major triad chords identified by red circle, Barbershop 7th chords by blue square and minor 6th chords by green diamond

intonation". These samples are highlighted in magenta in Table 8.1 and the figures below. Notably samples 17 and 31 were synthesised "out of tune" chords and received ratings in the region 1 - 2. Samples 34 and 46 were synthesised perfectly in tune chords and received high ratings for Lock (6). Sample 8 was the Sound Hypothesis chord from the end of "Something Big" analysed in detail in Chapter 7 and received high ratings (5) for Lock, Ring and Expanded Sound.

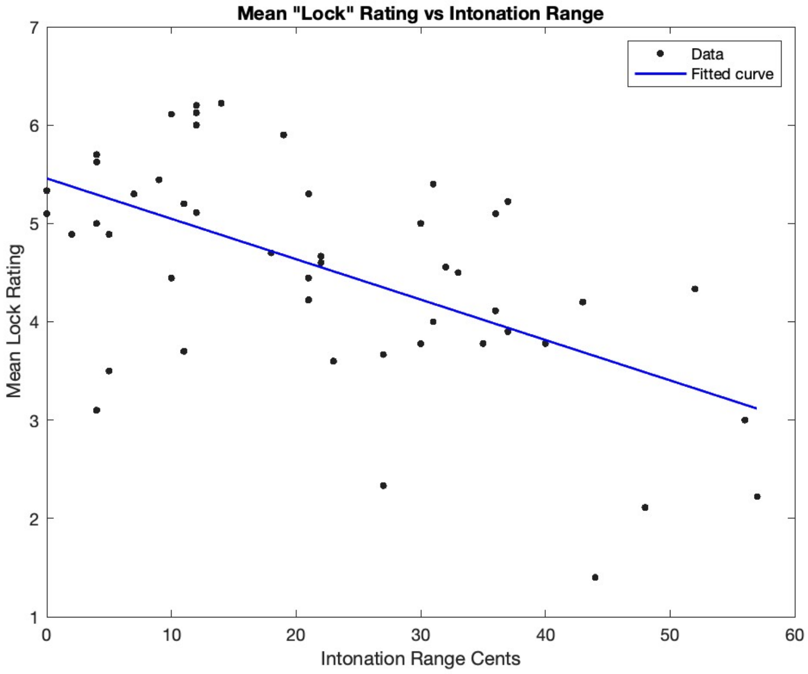
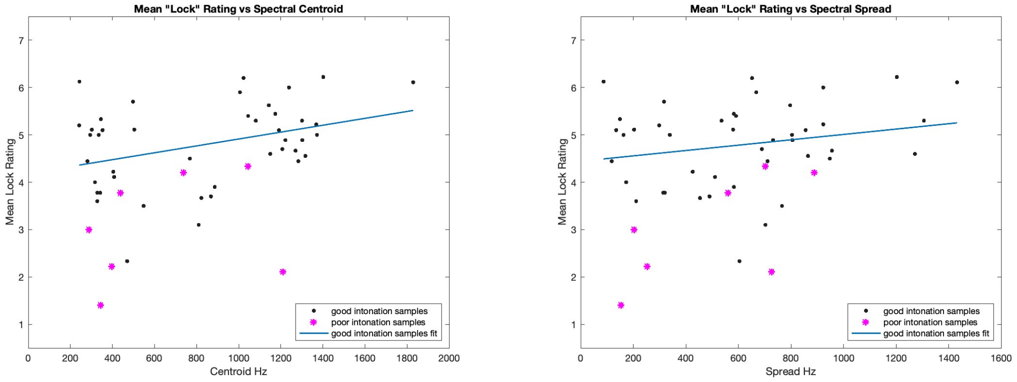


Figure 8.5: Lock vs Frequency Divergence Range (R)

The *mean* Lock rating for each sample is plotted against its computed spectral centroid and spread in Figure 8.6. The seven samples identified as having poor intonation were generally given low Lock ratings, although one of these samples (25) was rated quite highly. This may be because only one of the harmony voice parts is significantly out of tune, the other two are well tuned. A least squares fit line, excluding the poor intonation samples, is shown in blue. The least squares fit values range from 4.36 to 5.51 as the centroid increases and the spread ranges from 4.49 to 5.25. These values suggest if the poor intonation samples are excluded the Lock rating is not greatly influenced by the value of centroid or spread.

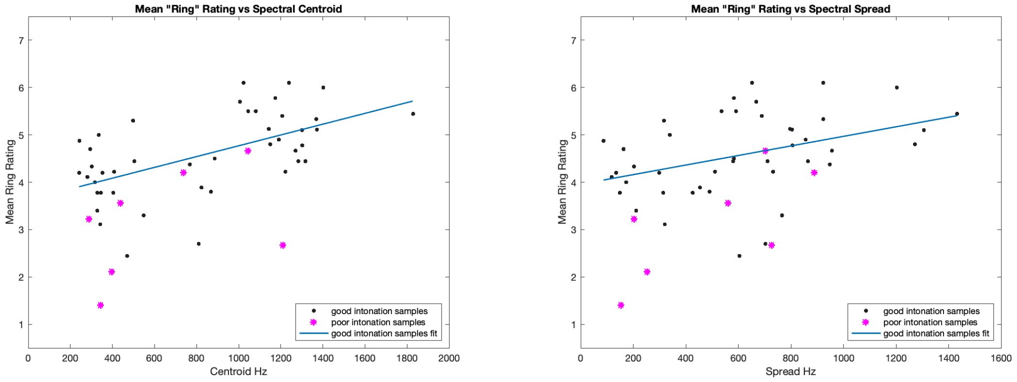
The *mean* Ring rating for each sample is plotted against the sample's computed spectral centroid and spread in Figure 8.7. The seven samples identified as having poor intonation were generally given low Ring ratings. A least squares fit line, excluding the poor intonation samples, is shown in blue. The least squares fit value ranges from 3.9 to 5.7 as the centroid increases and the spread ranges



(a) Lock vs Centroid (b) Lock vs Spread

Figure 8.6: Mean values of Lock vs Centroid and Spread

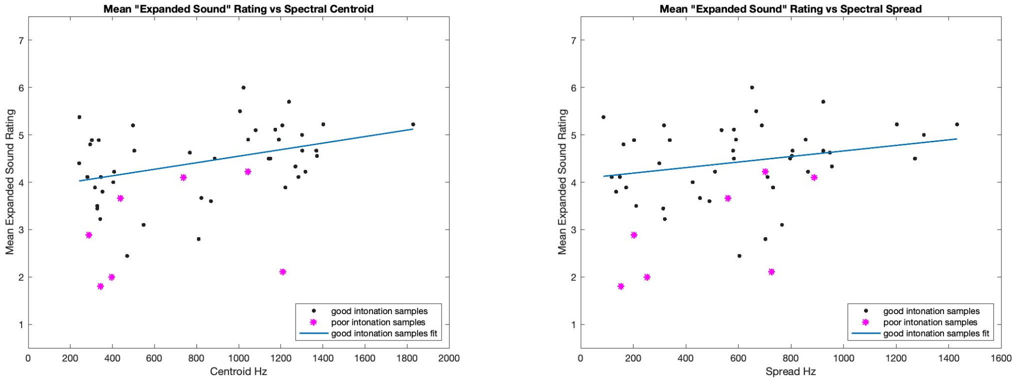
from 4.04 to 5.39. These values suggest if the poor intonation samples are excluded the Ring rating does trend up with increasing values of centroid and with a concomitant increase in spread.



(a) Ring vs Centroid (b) Ring vs Spread

Figure 8.7: Mean values of Ring vs Centroid and Spread

The *mean* Expanded Sound rating for each sample is plotted against the computed centroid and spread in Figure 8.8 with the poor intonation samples again identified in magenta. The poor intonation samples were generally given low Expanded Sound ratings. A least squares fit line, excluding the poor intonation samples, is shown in blue. The least squares fit values range from 4.02 to 5.12 against centroid and 4.12 to 4.91 against spread. These values suggest if the poor intonation samples are excluded the Expanded Sound rating is not greatly influenced by the values of centroid or spread.



(a) Expanded Sound vs Centroid

(b) Expanded Sound vs Spread

Figure 8.8: Mean values of Expanded Sound vs Centroid and Spread

8.3 Discussion

Primary features in musical psychoacoustics are pitch, loudness and timbre (Rasch and Plomp 1999). While the physical correlation between fundamental frequency with pitch and sound intensity with loudness are comparatively well defined, correlations between acoustic metrics and timbral features are less clear. The perception tests undertaken in this research sought to identify if there are absolute acoustic metrics which are physical correlates of the idiomatic descriptions of “Lock”, “Ring” and “Expanded Sound” used to define the desired characteristics of the Barbershop genre.

Each of the proposed physical correlates (intonation accuracy, spectral centroid and spectral spread), for a chord in a steady state, relate to the way the acoustic energy is distributed across the frequency spectrum. The factors which influence these physical measures can be illustrated using one of the perception test samples (PT4) as shown in Figure 8.9. This sample was rated highly by the perception test participants, receiving a mean Lock rating of 6.2, Ring rating of 6.1 and Expanded Sound rating of 6.0, out of a maximum score of 7.0. The tuning of each voice part, measured by the frequency of the harmonic peaks, was very accurate, with a maximum deviation of 12 cents from the ideal frequency in just intonation. The strongest harmonic in the chord, which contained 54% of the total energy, is at 690 Hz (4 x root frequency). It is a combination of the bass' 4th harmonic with the lead's 2nd harmonic and is close to R1 (660 Hz) for an /a/ vowel. The second strongest harmonic, containing 14% of the total energy, is at 1380 Hz (8 x root frequency and an octave above the strongest harmonic). These two harmonics will have a major influence on the chord's values of the centroid and spread. The figure also shows the equal loudness contours in phon. Accurately tuned high harmonic combinations close to vocal tract resonances not only contribute a high percentage of the total acoustic energy, they will also be perceptually louder than lower harmonics with the same dB level.

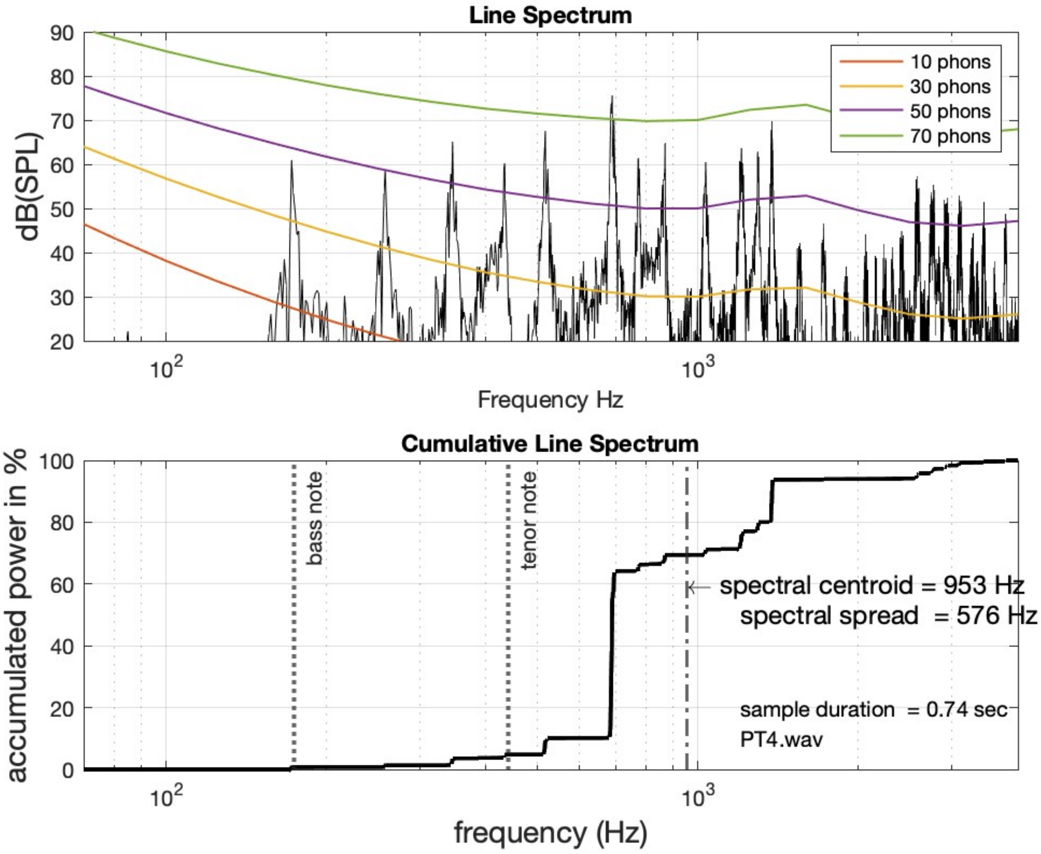


Figure 8.9: Line spectrum and cumulative line spectrum for perception test sample PT4; major triad, 10th voicing, root frequency F3

The spectra shown in Figure 8.9 were obtained using a Fourier analysis which gave a frequency resolution of 1.3 Hz across the whole frequency range. It was selected to enable accurate calculation of the harmonic peaks at low frequencies. In reality the ear doesn't resolve complex sounds so precisely, rather fusing the acoustic energy into wider frequency bands (Howard and Angus 2017). Figure 8.10 shows the results of analysing sample PT4 using a 1/24 octave filter bank which has frequency bands of a quarter tone (± 25 cents) about the centre frequency of a note, which is comparable to the criterion of a maximum interval of 40 cents identified earlier for notes to be considered as in tune. Improving the intonation accuracy between close harmonics in the voice parts will consolidate acoustic energy within the same frequency bands and thereby increase the perception of loudness at that frequency.

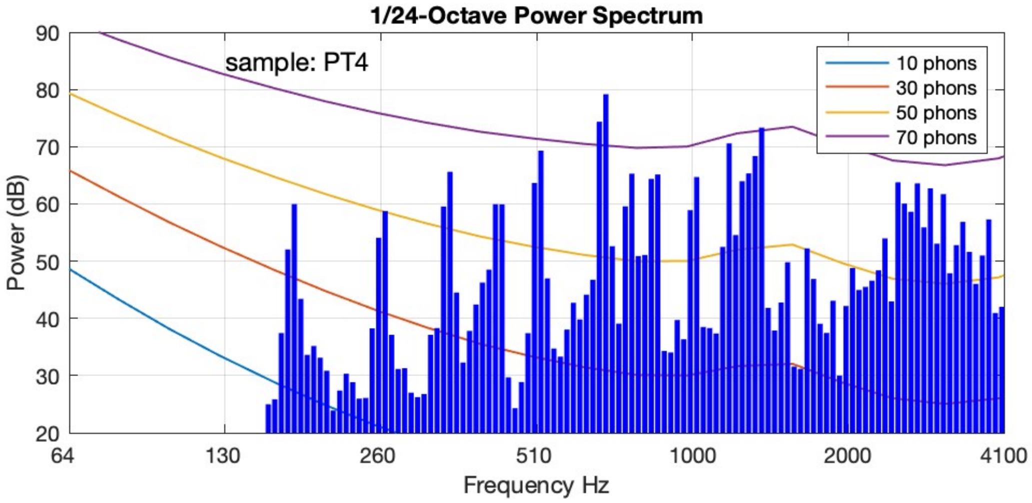


Figure 8.10: Line spectrum using 1/24 octave filter bank for perception test sample PT4; major triad, 10th voicing, root frequency F3

It is important to keep in mind the perception test samples had been normalised. While normalisation will not affect the computation of intonation accuracy, centroid or spread it does prevent a comparison of the strength of harmonics between test samples. Consequently it was not possible to use the perception tests to assess acoustic power output as a physical correlation of Lock, Ring or Expanded Sound.

8.3.1 Lock and Intonation Accuracy

The Barbershop Harmony Society (2023) (BHS) Contest and Judging Handbook defines Lock as "the feeling associated with a justly in tune chord, whose quality is determined by the degree of intonation achieved in and between the individual voice parts". The perception test participants' descriptions of Lock are consistent with this definition where their emphasis was on accuracy of intonation, to the extent that the "chord sounds unified" and "the constituent notes are almost not discernable". This description complements the converse observation by R. Plomp and Levelt (1965) that when there is inaccurate intonation small frequency differences can create beats or roughness in the combined sound. In the participant's responses the Lock rating reduced as the chord diverged from just intonation as shown in Figure 8.4. Samples with a divergence range R of 40 cents or more generally had a Lock rating below 4. While accuracy of intonation has a physical correlation with Lock, Figure 8.5 indicates the Lock rating for accurately tuned chords is independent of the values of centroid and spread.

Although it was not assessed in the perception tests, improved Lock could result in a perception of increased loudness as the voice harmonics combine in the same frequency bands. This effect would be more pronounced for higher harmonics because of the non linearity of hearing.

In summary it is concluded that accurate tuning between the voices in just intonation is a physical correlation of the perceptual description of Lock. Lock is independent of the values of spectral centroid and spectral spread.

8.3.2 Ring and Spectral Centroid

The BHS definition of Ring is “the sound resulting from the production and reinforcement of harmonics in the composite voice parts, derived from the ringing quality contained in the individual voices”. A limitation of this definition is that it doesn’t explain the source of “the ringing quality contained in the individual voices”. The perception test participants’ descriptions are consistent with the first part of the BHS definition and place emphasis on “the reinforcement of overtones from multiple voices singing locked chords”, “low-order harmonics reinforcing each other to generate prominent overtones”, and a “perception of brilliance”. The second part of the BHS definition stating Ring is derived from “the ringing quality of the individual voices” echoes Richards (2001) observation that an ensemble’s Ring “is dependent on the individual voices possessing Ring”. The term Ring could be interpreted as resonance in the voice but the descriptions given place emphasis on evidence of energy in the higher frequencies. In this context Ring would relate to the description of “brightness” used by Schubert and Wolfe (2006) which was shown to have a correlation with spectral centroid. Mathematically the spectral centroid for a quartet is equivalent to a weighted average of the centroids of the individual voices, which would be consistent with the statement that the Ring of the ensemble is a consequence of the ringing in the individual voices. Evidence of high centroids occurring in the quartet sound when high centroids are present in the individual singers voices is supported by the results of the 15-bar exercise in the Sound Hypothesis tests described in Chapter 7.

The perception test samples which had poor intonation were generally given Ring ratings of below 4 (Figure 8.6). If these samples are excluded the least squares fit in Figure 8.6 indicates Ring increases as the centroid increases (and correspondingly so does the spectral spread).

In summary it is concluded that provided the ensemble is accurately tuned the spectral centroid of the sound is a physical correlate of the perceptual description of Ring.

8.3.3 Expanded Sound and Spectral Spread

The BHS definition of Expanded Sound is "the effect resulting from the combined interaction of voices singing with accurate intonation with uniform word sounds in good quality with proper volume relationships that reinforce the more compatible harmonics and combination tones and with precision all producing an effect greater than the sum of the individual voices". The test participants described Expanded Sound as the "perception that the combined sound is greater than the four component voices" with the "creation and reinforcement of strong overtones and undertones", which are compatible with the BHS description. Notably these definitions are for chords in a steady state and make no reference to perceived changes with time which may result from singer adjustments.

Of the three defining attributes of Barbershop singing Expanded Sound is the most difficult to quantify. Lock and Ring can be related to measurable acoustic phenomena. Expanded Sound is a subjective perception which is "greater than the sum of the parts", without having a comparable reference for the sum of the individual voices. Nevertheless the research results suggest some possible origins for the perception of Expanded Sound, which all relate to the position, size and shape of a chord's frequency spectrum:

- the creation of "combination tones", generated by the sum and differences between the harmonics of the contributing voices.
- a broadening of the frequency spectrum of the ensemble's output, relative to the spectra of the individual voices, indicated by a large spectral spread.
- an increase in the amplitude of the spectrum resulting from the matching harmonics in the individual voice parts combining, resulting in a perceived increase in loudness.
- an enhancement in the spectral properties with time as singers make adjustments; improving intonation accuracy, resonance tuning and possibly phase matching

Combination tones are an acoustic effect resulting from the sum and difference between tones of differing frequencies. Combination tones do not appear in the frequency spectrum because they don't contain any rms power. Based on previous research (Rasch and Plomp 1999) described in Chapter 4 it was concluded that combination tones are only clearly discernable for pure tones which are close in frequency, of comparable amplitude and in phase. As the majority of the harmonics in each contributing voice in a quartet will have different amplitudes and phases the conditions for combination tones to be audible will be very limited. Combination tones are also likely to be concealed by masking between the voices (Sundberg and Gauffin 1974). A possible exception to these limitations is the undertone below the notes being sung, created by the strong low harmonics of the voice parts and not masked by the fundamental frequencies. As a result of the special conditions required to hear combination tones they are unlikely to be the principle source of the phenomenon of Expanded Sound.

The spectral spread is a measure of the distribution of the acoustic energy about the spectral centroid and is an indication of the *breadth* of the spectrum. To date spectral spread has not been associated with a specific perceptual effect. The hypothesis formulated at the start of this research was that spectral spread was a physical correlation for the perception of Expanded Sound. The results of the perception test ratings for Expanded Sound are shown in Figure 8.7. When the samples with poor intonation are excluded, the least squares fit indicates the rating of Expanded Sound does not vary significantly with the spectral spread (or centroid) suggesting these spectral measures do not correlate with the perception of Expanded Sound.

The third explanation for the perception of Expanded Sound is that it relates to the *amplitude* of the frequency spectrum. When four voices combine to sing accurately tuned chords the superposition of the matching harmonics will amplify the strength of the harmonics in the individual voices. Using the computer model in Chapter 6 it was demonstrated the effects can be maximised, using the same glottal power input, through the choice of chord structure, root frequency and vowel which match the chords harmonics with the singers vocal tract resonances. The resulting sound may be further enhanced in real time as the singers make vocal adjustments. This possible correlation between spectral amplitude and the perception of Expanded Sound could not be objectively evaluated from the perception tests because the sound samples had been normalised and the relative acoustic power input of the singers was not known.

On balance, but without conclusive evidential support, it is considered spectral amplitude is most likely to be the physical correlate of Expanded Sound as a result of vocal efficiency with accurate intonation of matching harmonics, not simply as a result of greater acoustic power input. The effect will be more apparent at higher frequencies because of the non linearity in hearing and may be enhanced in real time by the singers resonance tuning and improving intonation. With the exception of the "undertone", combination tones are not considered to be a major contributing factor to the perception of Expanded Sound.

8.4 Limitations

The main limitations of the methodology adopted for investigating the correlation between objective acoustic measures and psychoacoustic descriptions are:

- The large variety of sound samples and the limited number of test participants means the results are not statistically robust. The conclusions should therefore be viewed as plausible explanations of the effects
- The chord samples were of limited duration (2 seconds) and with a few exceptions had no lead in sequence making it difficult for participants to assimilate the musical context
- Synthesised chords are not a familiar sound which the participants would be used to judging
- Normalising the sound samples to ensure consistency for the participants prevented an analysis of harmonic amplitude being the physical correlate of Expanded Sound

8.5 Summary

Lock, Ring and Expanded Sound are interrelated but conceptually different psychoacoustic effects with their own determining features. Within the limitations set out in the previous section it is concluded:

- Lock correlates with intonation accuracy. Good Lock is a necessary condition to maximise Ring and Expanded Sound
- Ring correlates with a high spectral centroid which results from acoustic energy being concentrated in the higher harmonics in a chord. Ring in the ensemble sound derives from the Ring in the individual voices
- Expanded Sound does not correlate with spectral spread. Nor is it a result of combination tones. Expanded Sound is more likely to correlate with increasing spectral amplitude as a result of vocal efficiency, not simply as a result of greater acoustic power input. The efficiency derives from accurate intonation between the voices and aligning harmonics with the resonances of the vocal tract. The effects can be enhanced in sustained chords as the singers make real time adjustments to their vocal techniques

Chapter 9

Conclusions

A cappella singing is a very popular musical genre, widely practised by both professional and amateur groups. Singing ensembles can enrich listener experience if they are able to create a diverse palette of sounds with different acoustic characteristics. To realise this goal practitioners need to understand the factors which affect an ensemble's acoustic output. A substantial body of research on the solo voice already exists but for practical reasons comparatively few investigations have been conducted on ensemble singing. By its nature ensemble singing is a complex, multi-variable form of music, which makes it difficult to isolate the contributory factors and objectively quantify their impact on the acoustic output.

In parallel with understanding the practical factors influencing the acoustic output, it is helpful to have a consistent vocabulary for describing how an ensemble's sound is perceived. Having these two ways of classifying the sounds prompts the question "is there a correlation between absolute acoustic measures and perceptual descriptions?" The research described in this thesis is a study of both the objective acoustic metrics and the perceived acoustic features of the musical genre of Barbershop singing and the relationship between the two. While Barbershop is the primary focus of the research, the insights gained are of relevance to a cappella singing in general.

Accurate tuning of note intervals, spectral centroid and spectral spread were postulated as objective acoustic metrics for differentiating between the sounds of Barbershop chords in a steady state. An analysis of 50 chord samples extracted from published recordings of Barbershop quartets demonstrated a wide range of values for these metrics resulted from changing the chord structure, the chord pitch (defined by chord root frequency) and the vowel sung. A computer model of a Barbershop quartet was developed to run parametric studies to understand the causes of the variations in the acoustic features. The computer model produced behaviours which were consistent with the conclusions derived from analysing the archive database. Primary variables affecting the metrics were confirmed to be the chord structure, the root frequency of the chord and the vowel sung. The computer model generally predicted lower values for the metrics than the archive database for the corresponding chords, vowels and chord root frequencies. Closer correlation was achieved with the laboratory test samples and the computer model when more accurate voice characteristics for each voice part had been measured.

As the archive samples had been normalised, a comparison of the absolute acoustic power output of a quartet when singing different chords, root frequencies and vowels could not be made. The influence of these variables was investigated by running the computer model with different chord structures, root frequencies and vowels using while maintaining the same glottal power input for each voice in each sample. The results demonstrated greater vocal efficiency, measured by increased quartet power output relative to power input, was achieved with resonance matching i.e. with chord structures, root frequencies and vowels which produced strong harmonics in the region of vocal tract resonances.

An additional limitation of the archive sample analysis was that the contribution made by the individual singers to the aggregate sound could not be isolated. This limitation was addressed by conducting structured audio laboratory tests with a live quartet. The laboratory tests confirmed the chord structure, chord pitch and vowel sung all had a major influence on the acoustic characteristics of the ensemble. The quartet exercises were used to determine the overall ensemble characteristics as well as the each singer's intonation accuracy and individual glottal opening quotients. The singers also performed exercises solo to determine their vocal characteristics. Accepting the singers may modify their techniques when performing as a quartet, the solo results provided useful insight into their individual features. When singing the same vowel there were differences between the singers' tract resonances, and the resonances also changed with pitch sung. The analysis of a sustained chord recorded during the laboratory tests demonstrated that the singers made vocal adjustments in real time, improving intonation accuracy and resonance tuning, which affected the intensity and spectral properties of the quartet. This behaviour will affect the way a quartet sound is perceived over time.

Perception tests were conducted with qualified Barbershop judges to investigate the correlation between the selected acoustic metrics and the perceptual descriptions of "Lock", "Ring" and "Expanded Sound" which are commonly used terms for describing the characteristics of Barbershop singing. The perception test sample database was compiled from the archive database, synthesised chords using the computer model and samples from the laboratory tests. The samples were predominantly isolated chords of relatively short duration, which excluded the aural context of lead in chord sequences and possible singer adjustments with time. Also, the perception test samples were normalised, which excluded using "loudness" as a variable.

9.1 Restatement of the research hypothesis

The hypothesis posed at the outset of this research was:

The distinctive sound of Barbershop singing can be defined by the absolute acoustic measures of accurate tuning in just intonation and high spectral centroid and broad spectral spread in the frequency domain, which are the physical correlates of the perceived qualities of Lock, Ring and Expanded Sound.

And as a corollary:

For a given chord structure, pitch and vowel determined at the point of composition/musical , a Barbershop ensemble can optimise this unique sound by adjusting, in real time, their individual vocal techniques.

The Barbershop Harmony Society (2023) definition of Lock "is the feeling associated with a justly in tune chord, whose quality is determined by the degree of intonation achieved in and between the individual voice parts". The research supported this definition. The perception tests confirmed the measured accuracy of tuning between the voices in just intonation is a physical correlation of the perceptual description of Lock.

The Barbershop Harmony Society (2023) definition of Ring "is the sound resulting from the production and reinforcement of harmonics in the composite voice parts, derived from the ringing quality contained in the individual voices". Based on the research this definition requires some qualification. It is the production and reinforcement of the *higher* harmonics which creates the perception of Ring. This descriptor for an ensemble could then correspond to the term "brightness" used by Schubert and Wolfe (2006) which they demonstrated scaled with spectral centroid for solo voices. Provided an ensemble achieves good Lock, the spectral centroid is a physical correlate of the perceptual description of Ring. Ring in some, if not all, of the contributing voices is a prerequisite for Ring to be evident in the ensemble sound.

The Barbershop Harmony Society (2023) states Expanded Sound is "the effect resulting from the combined interaction of voices singing with accurate intonation with uniform word sounds in good quality with proper volume relationships that reinforce the more compatible harmonics and combination tones and with precision all producing an effect greater than the sum of the individual voices". Based on the research findings this definition requires some qualification. Combination tones, which can undoubtedly be perceived, don't appear in the frequency spectrum as they contain no time averaged energy, they are a temporal effect. Goldstein J L (1970) found the intensities of combination tones were very dependent on the relative intensities and phase relations of the tones. For the combination of complex sounds in ensemble singing the differences in amplitudes and phases between the harmonics in the contributing voices will be significant. With the possible exception of the undertone below the phonation frequencies of the contributing voices combination tones are unlikely to be discernable and therefore wont contribute to the overall perception of Expanded Sound. The perception tests didn't show any clear relationship between spectral spread and Expanded Sound, thereby excluding spectral *breadth* as a physical correlation of Expanded Sound. From the computer model results it seems more likely that Expanded Sound relates to the *amplitude* of the frequency spectrum, i.e. an increase in the output power of the sound from the production and reinforcement of the harmonics in the individual voices. The opportunity to create Expanded Sound is greatly influenced by the composer/arranger and lyricist choosing chord structures, chord pitch and vowel which create harmonic combinations which lie in the region of the singers' vocal tract resonances and thereby exploit vocal efficiency. Expanded sound is not simply a result of singers generating more glottal input power. The perceived effect will be more pronounced at higher frequencies because of the non-linearity of hearing. The effect can be further enhanced by the singers making real time adjustments to improve intonation accuracy and by resonance tuning.

On the basis of the research undertaken using analyses of archive database, computer modelling of a quartet, laboratory testing with a live quartet and the perception tests with a panel of qualified Barbershop judges it is concluded:

Lock is a perception of tonal consonance associated with a justly in tune chord, whose quality can be measured by the precision of intonation achieved in and between the individual voice parts. Good Lock is a prerequisite for maximising Ring and Expanded Sound.

Ring is a perception of brightness in the sound resulting from the production and reinforcement of the higher harmonics in the composite voice parts. The quality of Ring correlates with a high spectral centroid in the frequency spectrum. Ring in some, if not all, of the contributing voices is a prerequisite for Ring to be evident in the ensemble sound.

Expanded Sound is the effect resulting from the interaction of voices singing with accurate intonation with uniform word sounds that combine the matching harmonics in the contributing voices all producing an effect greater than the sum of the individual voices. Expanded Sound is a result of vocal efficiency in the singing process, it is not simply a consequence of increased vocal effort. The physical correlation of Expanded Sound is sound pressure level and can vary with time in sustained chords.

In summary the research results support the original hypothesis that the sound of Barbershop singing can be represented in the frequency domain by the absolute measures of accurate tuning in just intonation and high spectral centroid which are the physical correlates of the perceived qualities of Lock and Ring. However, the research did not support the hypothesis that spectral spread is the physical correlate of Expanded Sound. It was concluded that Expanded Sound is more likely to be a consequence of vocal efficiency from resonance matching with a physical correlate of increased power.

The compositional choices of chord structure, pitch and vowel are very influential in creating the opportunities for ensembles to produce the distinctive sound of Barbershop. The research supports the hypothesis that an ensemble can optimise this unique sound by adjusting, in real time, their individual vocal techniques.

9.2 Future work

The sound produced by a singing ensemble is the result of a complex process involving multiple variables. For practical reasons it is necessary when conducting research to be selective when choosing the inputs and outputs so the results can contribute to a holistic picture. Starting with a defined musical composition as input each singer's contribution to an ensemble is determined by their individual vocal technique and conditioned by environmental factors such as room acoustics and singer positioning. Singers adjust their vocal production in real time depending on what they hear of the ensemble. The acoustic output for the ensemble can be measured in a number of ways; intensity (loudness), intonation (tuning) and timbre. Methods of quantifying the outputs need to be specified, either as absolute acoustic metrics or qualitative descriptions. The input variables and measured output metrics used for this research have been defined. From the results some areas which merit further investigation can be identified.

The computer programme written to model a Barbershop quartet proved to be a useful tool for running parametric studies. It could be run with alternative inputs; glottal flow models other than the KLGLOTT88 model which have different glottal flow profiles (Doval et al. 2006), different open quotients (Henrich, d'Alessandro, et al. 2005) from breathy to pressed voices; different lip radiation filters (Titze and Palaparthi 2018) which can be more efficient vocally depending on mouth opening; using different vocal tract resonances for the singers and changing with pitch; modelling mixed voice ensembles; having the different voice parts sing different vowels concurrently; changing the phases between singers which could be more significant with small numbers of singers in environments with little reverberation.

The perception test samples primarily used short duration sound samples. Tests using longer duration samples, with lead in chord sequences would provide more musical context when assessing the metrics and also identify how much singers make adjustments while performing. Masking between pure tones has previously been investigated but the masking of the harmonics in individual voices and between voices, which could affect the perception of the complex sounds produced by an ensemble, has received little attention. The samples used in the perception tests were normalised for consistency. Further perception tests could be carried out using the same samples at different loudness levels to see if the rating of Expanded Sound was affected.

Chapter 10

Practical applications of the research findings

Barbershop is a form of a cappella singing with its own distinctive sound and yet, apart from the specific requirements of having only four different voice parts and the melody line being in the second voice, there are no absolute rules which define the genre. Since its origins in the mid-19th century the style has evolved to the extent that nowadays what defines Barbershop can be a source of debate. Nevertheless Barbershop does have preferred perceptual features which, when strongly evident, differentiate this form of a cappella and make it distinctive. These qualitative properties have been described as Lock, Ring and Expanded Sound. The Barbershop Harmony Society (2023) has published descriptions of these characteristics while acknowledging the definitions are evolving. In a document prepared for the Barbershop Harmony Society, Richards (2001) put forward a “lockability” ranking for chords using an empirically derived formula based on the number of combination tones in a chord. Richards’ approach made no allowance for the different strength of the harmonics in a chord, which will be affected by the chord structure, the pitch of the chord and the vowel being sung. This research has demonstrated that in addition to chord structure the pitch of the chord and the vowel do affect the perceptual characteristics of an ensemble’s sound.

The research described in this thesis sought to identify which conditions create the opportunities to realise these perceptual features. From the results it is concluded that Lock, Ring and Expanded Sound can best be achieved by (a) selecting chord structures, pitch and vowel combinations which produce strong harmonic/ harmonic combinations in the region of the singers’ vocal tract resonances (b) singing with precise tuning of the vertical intervals between notes in just intonation which minimises tonal dissonance and concentrates the acoustic energy in the matching voice harmonics into the same frequency bands and (c) applying resonance tuning to optimise these effects.

Resonance tuning is a vocal technique where a singer adjusts the "articulators" (the lips, the jaw opening, the tongue, the velum and the larynx) to alter the resonant frequencies of the vocal tract so they match the harmonics in the voice for the note being sung. Sundberg (1987) sets out the consequences of moving the different articulators. All resonances are lowered by a narrowing of the lips opening and by a lengthening of the vocal tract. R1 is raised by opening the jaw. R2 is sensitive to the shape of the tongue; when the tongue constricts the front part of the vocal tract R2 is raised, if the tongue restricts the vocal tract in the velar region R2 is low. In vocal pedagogy it is suggested similar effects can be achieved by modifying the vowel being sung; for example by moving from /i/, a bright vowel, towards /u/, a dark vowel. However, vowel recognition is about the relative positioning of the formant peaks *not* their absolute values. For example, shifting all the resonances down by lengthening the vocal tract shouldn't impact the vowel perceived.

In this final chapter some strategies based on these conclusions are presented which may help practitioners realise their acoustic objectives. Although the observations are primarily intended for Barbershop singing the principles are of relevance to any form of a cappella singing which is based on just intonation.

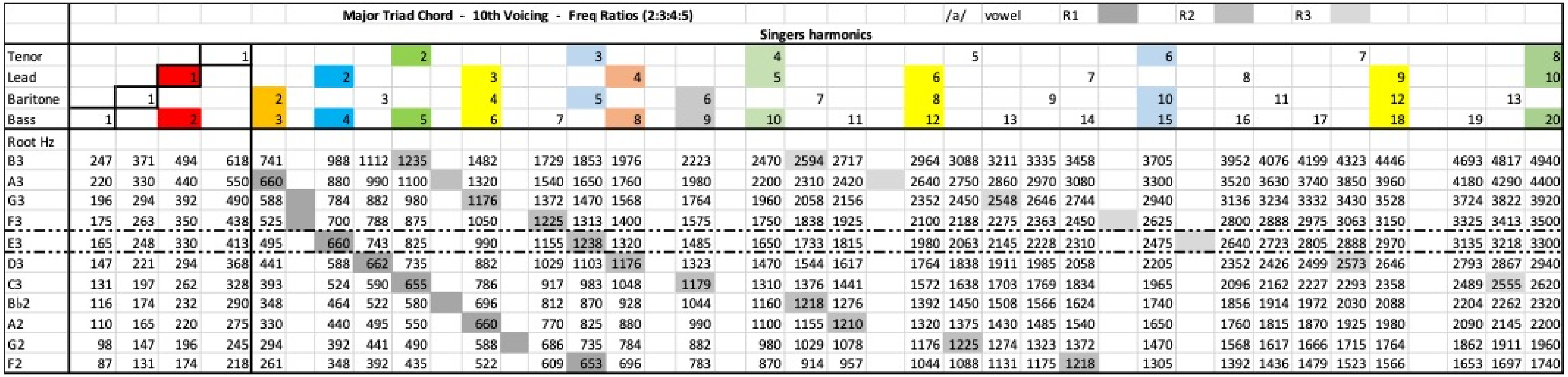
10.1 Practical applications for composers, arrangers and lyricists

The compositional choices of pitch (chord root frequency), chord structure and vowel made by composers, arrangers and lyricists significantly affect the opportunities for performers to produce Lock, Ring and Expanded Sound. Ensemble chords are a fusion of complex tones each of which comprises a set of partials. Neighbouring partials from the different tones can interact to produce interference effects which can be perceived as dissonance (Harrison and Pearce 2020). Chords with small integer ratios between the notes in just intonation will have maximum harmonic reinforcement and consonance; the strongest example being the major chord of root:fifth:octave:third with the frequency ratios of 2:3:4:5. Chords with higher numeric ratios such as a minor 6th (6:7:9:10) will have partials which fall in the same critical band for hearing and will sound dissonant even when accurately tuned and also have less harmonic matching between the voice parts. The arranger's choice of pitch and chord structure defines the notes required of each voice part. Depending on the vowel being sung, the selection will determine the proximity of the harmonics to the resonant frequencies of the singers' vocal tracts and thereby the opportunity to exploit resonance tuning.

As an example, Figure 10.1 shows the frequencies of the harmonics in a major triad chord with 10th voicing for different chord root frequencies. The harmonic combinations close to R1, R2 and R3 for an /a/ vowel are highlighted in grey. If a chord with a root on E3 is selected all the phonation frequencies are below R1. The dominant harmonic combinations will be the (bass' 4th + lead's 2nd) close to R1 (660 Hz), the (baritone's 5th + tenor's 3rd) close to R2 (1200 Hz), the (tenor's 6th + the baritone's 10th + the bass' 15th) close to R3 (2550 Hz). By using such tables, it is possible to identify composition and lyric choices which provide the best opportunities for singers to maximise the ensemble power through resonance tuning to the key harmonic combinations. Comparable charts can be constructed for other chord and vowel combinations.

Conversely, if the voice parts concurrently sing different vowels, the peak intensities of the harmonic combinations will be reduced because the singers' vocal tract resonances will amplify the aligned harmonics differently.

Where the musical context allows, important chords should be sustained for sufficient time to enable singers to accurately tune their notes and enhance the sound through resonance tuning.



													Major Triad Chord - 10th Voicing - Freq Ratios (2:3:4:5)												/a/	vowel		R1	▨		R2	▨		R3	▨					
																				Singers harmonics																			
Tenor				1					2					3					4					5					6					7					8
Lead			1				2				3				4				5				6				7				8				9				10
Baritone		1			2			3			4			5			6			7			8			9			10			11			12			13	
Bass	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19		20
Root Hz																																							
B3	247	371	494	618	741		988	1112	1235		1482		1729	1853	1976		2223		2470	2594	2717		2964	3088	3211	3335	3458		3705		3952	4076	4199	4323	4446		4693	4817	4940
A3	220	330	440	550	660		880	990	1100		1320		1540	1650	1760		1980		2200	2310	2420		2640	2750	2860	2970	3080		3300		3520	3630	3740	3850	3960		4180	4290	4400
G3	196	294	392	490	588		784	882	980		1176		1372	1470	1568		1764		1960	2058	2156		2352	2450	2548	2646	2744		2940		3136	3234	3332	3430	3528		3724	3822	3920
F3	175	263	350	438	525		700	788	875		1050		1225	1313	1400		1575		1750	1838	1925		2100	2188	2275	2363	2450		2625		2800	2888	2975	3063	3150		3325	3413	3500
E3	165	248	330	413	495		660	743	825		990		1155	1238	1320		1485		1650	1733	1815		1980	2063	2145	2228	2310		2475		2640	2723	2805	2888	2970		3135	3218	3300
D3	147	221	294	368	441		588	662	735		882		1029	1103	1176		1323		1470	1544	1617		1764	1838	1911	1985	2058		2205		2352	2426	2499	2573	2646		2793	2867	2940
C3	131	197	262	328	393		524	590	655		786		917	983	1048		1179		1310	1376	1441		1572	1638	1703	1769	1834		1965		2096	2162	2227	2293	2358		2489	2555	2620
B♭2	116	174	232	290	348		464	522	580		696		812	870	928		1044		1160	1218	1276		1392	1450	1508	1566	1624		1740		1856	1914	1972	2030	2088		2204	2262	2320
A2	110	165	220	275	330		440	495	550		660		770	825	880		990		1100	1155	1210		1320	1375	1430	1485	1540		1650		1760	1815	1870	1925	1980		2090	2145	2200
G2	98	147	196	245	294		392	441	490		588		686	735	784		882		980	1029	1078		1176	1225	1274	1323	1372		1470		1568	1617	1666	1715	1764		1862	1911	1960
F2	87	131	174	218	261		348	392	435		522		609	653	696		783		870	914	957		1044	1088	1131	1175	1218		1305		1392	1436	1479	1523	1566		1653	1697	1740

Figure 10.1: Resonance tuning options for a major triad chord 10th voicing on an /a/ vowel. Vocal tract resonances R1, R2 and R3 highlighted in grey

10.2 Practical applications for performers

The practice of boosting vocal output through resonance tuning is well established for soloists in classical singing. Strategies which can be used by sopranos, altos, tenors and baritones have been analysed and documented by Henrich, J. Smith, et al. (2011). Having multiple voices in ensemble singing offers greater opportunities to amplify the sound as the different voice parts can separately tune R1 and/or R2 to different harmonic combinations.

The effect of resonance tuning was illustrated in the time segmented analysis of the final chord in the song "Start of Something Big" performed by Sound Hypothesis described in Chapter 7. A power spectrum for the chord sample as it progresses in 50 msec increments is plotted against time in Figure 10.2.

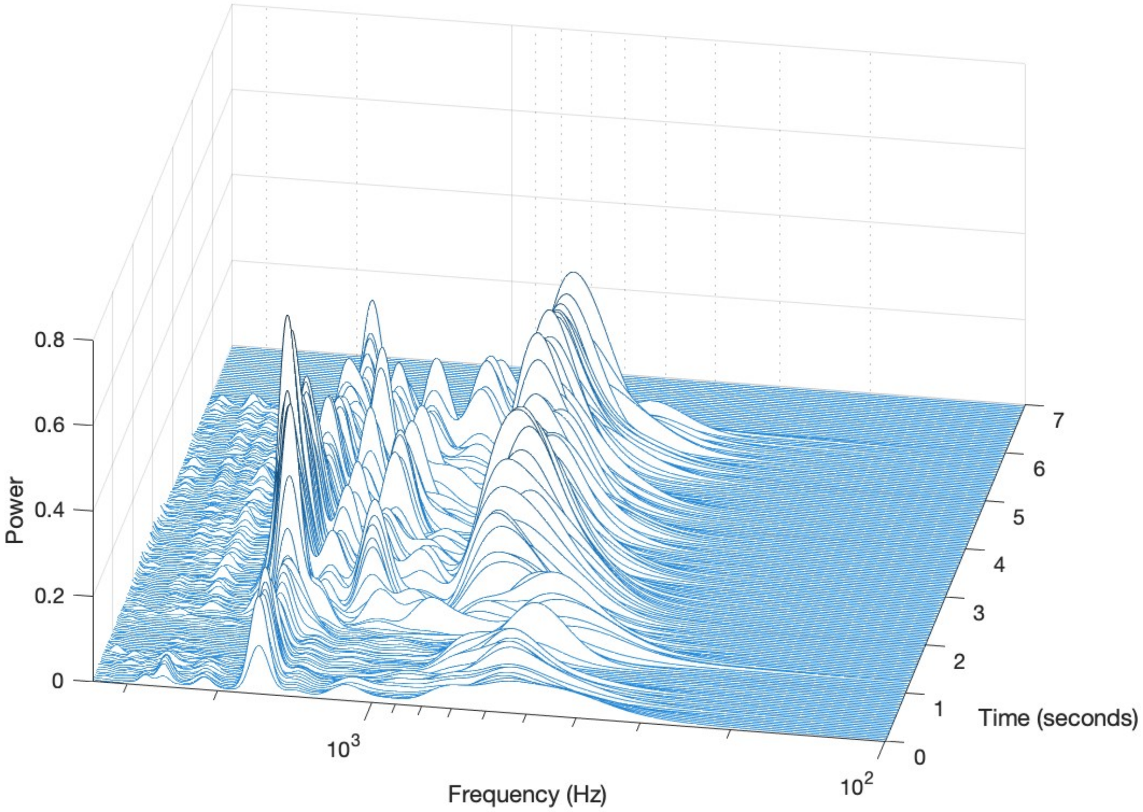


Figure 10.2: Power spectrum vs time for the final chord of "Something Big", Bar 108/109

The central part of the chord is on an /æ/ vowel. The power concentrated around R2 (1660 Hz) is very evident at 2 seconds. From 2 seconds onwards the power around R1 (620 Hz) increases. The lead's fundamental (above the tenor) is at 590 Hz close to R1, the bass' 7th harmonic is at 1660 Hz close to R2 as is the (lead's 3rd harmonic + baritone's 5th harmonic) combination at 1780 Hz. For the relatively short duration of this chord sequence each singer would be able to sustain their total acoustic power production but a redistribution of the total energy from R2 to R1 would cause the peak amplitudes and spectral centroid to shift. This shift, which could result from an adjustment of any of the singer's vocal tract resonances or slight divergence from accurate tuning in combined harmonics, would result in a change in the perception of the timbral qualities and the loudness of the chord with time. Maximising sound expansion can be achieved through accurate intonation and resonance tuning of the aligned harmonics. Ensuring precise synchronisation between the voice parts when changing matching phonemes, vowels, diphthongs and singable consonants will facilitate this aim.

10.3 Implications for male, female and mixed voice ensembles

The research described in this thesis was based on all male ensembles but the conclusions and principles established for achieving Lock, Ring and Expanded Sound apply equally to female ensembles and mixed voice ensembles. However, the resulting timbre for these groups singing the same chords will differ because of differences in the physical morphology of the singers. In general men and women have different vocal tract lengths and as a consequence have different resonant frequencies but the female tract is not a scaled version of the male tract. The differences between men and women's resonances vary considerably between vowels (Fant 1975) with the average difference across all vowels being an increase of 12%, 17% and 18% for R1, R2 and R3 respectively. In mixed voice ensembles where the male and female singers are singing the same note different harmonics will be amplified because of the differences in tract resonances. If the ensemble is all female singing in a higher key the spacing between the harmonics is greater which requires the resonances to be moved more to be effective. As a consequence the perception of the vowel is more likely to change.

10.4 Final Remarks

A cappella singing is a very popular art form widely performed by both amateur and professional groups. All aspire to produce the best results appropriate to the genre being performed. While singing science will never supplant artistic flair in delivering memorable performances, an understanding of the causal effects of compositional choices and vocal techniques can help practitioners realise their acoustic goals. To date the Barbershop singers' objectives of Lock, Ring and Expanded Sound, while understood in general terms, have not been unequivocally defined. It is hoped the research contained in this thesis will contribute to a greater appreciation of these perceptual properties and explains the conditions under which they can be best achieved.

Appendix A

Glossary

Acoustic power The physical measure of the amount of energy produced and radiated into the air per second

Complex tone A sound wave composed of multiple simple tones with different frequencies, amplitudes, and phases.

Corner vowels The vowels at the extremes of the vowel triangle, representing extreme tongue placements and usually including /a/, /i/, and /u/

Critical Band The band of audio frequencies within which a second tone will interfere with the perception of the first tone

Electroglottogram A display of the electrical conductance through the larynx by the glottis

Equal temperament A tuning system in which the octave is divided into 12 semitones of equal size

Formant A peak in the output spectrum envelope radiated from the mouth

Fundamental frequency The lowest frequency in a periodic waveform; also called the first harmonic

Harmonic A component whose frequency is an integer multiple of the fundamental

Intensity A measure of power per unit area

Just intonation A tuning system in which the intervals between notes are based on whole number ratios of their frequencies

Partials The individual sine wave components which combine to create a complex tone

Pitch A perception of highness or lowness of a sound

Pythagorean tuning A tuning system using intervals based only on the perfect fifth (3:2 ratio) and the octave (2:1 ratio)

Resonance tuning Amplifying the sound produced by adjusting the shape of the vocal tract so its resonances align with the fundamental frequency or harmonics of the note being sung

Spectrum A display of the relative magnitudes or phases of the component frequencies of a waveform

Timbre The quality given to a sound by its overtones

Tonal Consonance The perception of musical intervals and chords that sound pleasant and stable to the ear

Undertone A component of a waveform whose frequency is an integer fraction of the fundamental

Vocal tract resonance The poles of the transfer function of the supra glottal vocal tract

Vowel tuning Modifying a vowel as a way of aligning the vocal tract resonant frequencies with the harmonics of the note being sung

Appendix B

Justly tuned note intervals

A diatonic major scale can be constructed in just intonation from three stacked major triads consisting of the 4th, 5th and 6th harmonics of the harmonic series. The three intervals in each triad are the major third, minor third and perfect fifth so the notes in each triad are in the ratio 4:5:6. For example in the key of C the notes are related by

$$\begin{array}{ccccccccccccc}
4 & : & 5 & : & 6 & & & & 4 & : & 5 & : & 6 \\
 & & & & 4 & : & 5 & : & 6 & & & & \\
\mathrm{F} & & \mathrm{A} & & \mathrm{C} & & \mathrm{E} & & \mathrm{G} & & \mathrm{B} & & \mathrm{D}
\end{array}$$

which can be translated to give ratios on a common scale:

$$\begin{array}{ccccccc}
16 & 20 & 24 & 30 & 36 & 45 & 54
\end{array}$$

By doubling or halving some of the ratios a scale within one octave can be expressed as

$$\begin{array}{ccccccccccccccc}
\mathrm{C} & : & \mathrm{D} & : & \mathrm{E} & : & \mathrm{F} & : & \mathrm{G} & : & \mathrm{A} & : & \mathrm{B} & : & \mathrm{C} \\
24 & : & 27 & : & 30 & : & 32 & : & 36 & : & 40 & : & 45 & : & 48
\end{array}$$

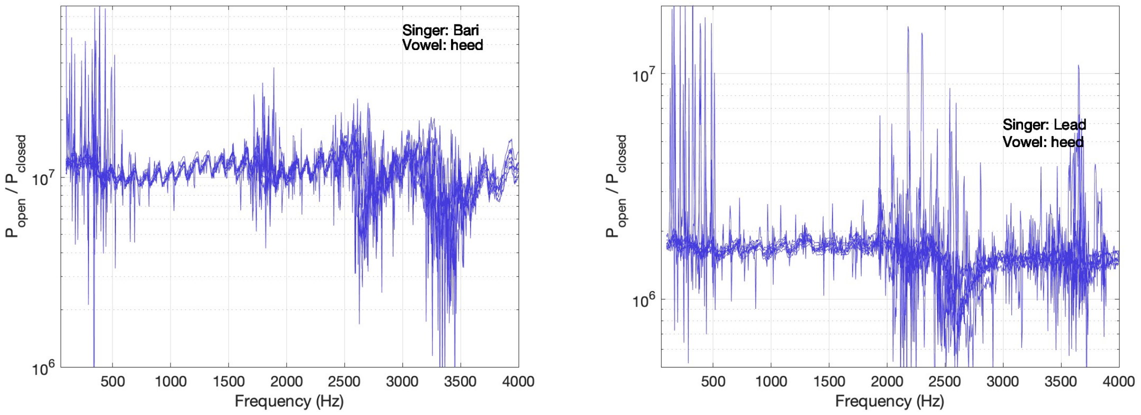
Commonly used consonant intervals which lie within this justly tuned major diatonic scale are shown in Table B.1.

Justly Tuned Interval	Frequency Ratio
Unison	1:1
Semitone	16:15
Minor tone	10:9
Major tone	9:8
Sub-minor third	7:6
Minor third	6:5
Major third	5:4
Perfect fourth	4:3
Augmented fourth	45:32
Diminished fifth	64:45
Perfect fifth	3:2
Minor 6th	8:5
Major 6th	5:3
Harmonic diminished 7th	17:10
Harmonic minor 7th	7:4
Grave minor 7th	16:9
Minor 7th	9:5
Major 7th	15:8
Octave	2:1

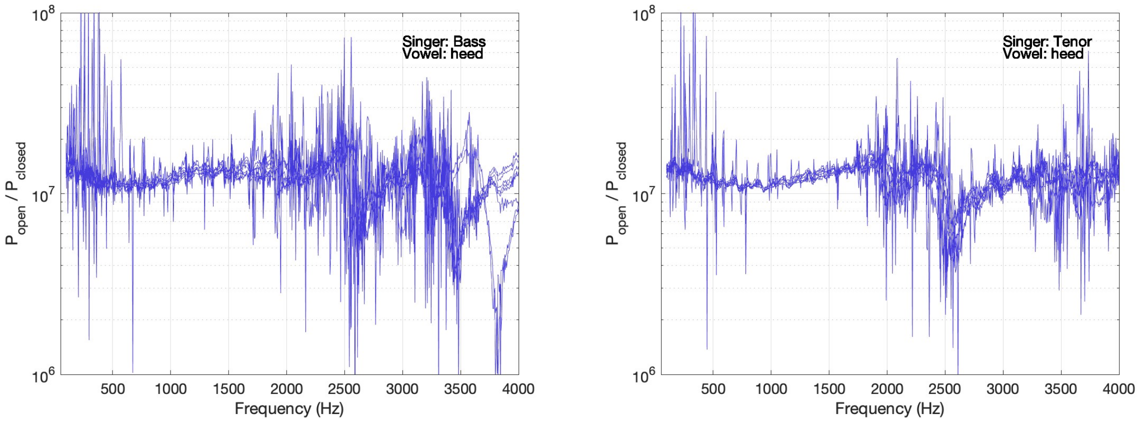
Table B.1: Justly tuned intervals

Appendix C

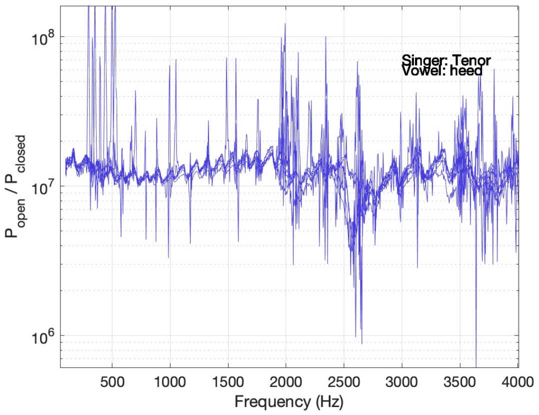
Supplementary figures



(a) baritone (b) lead

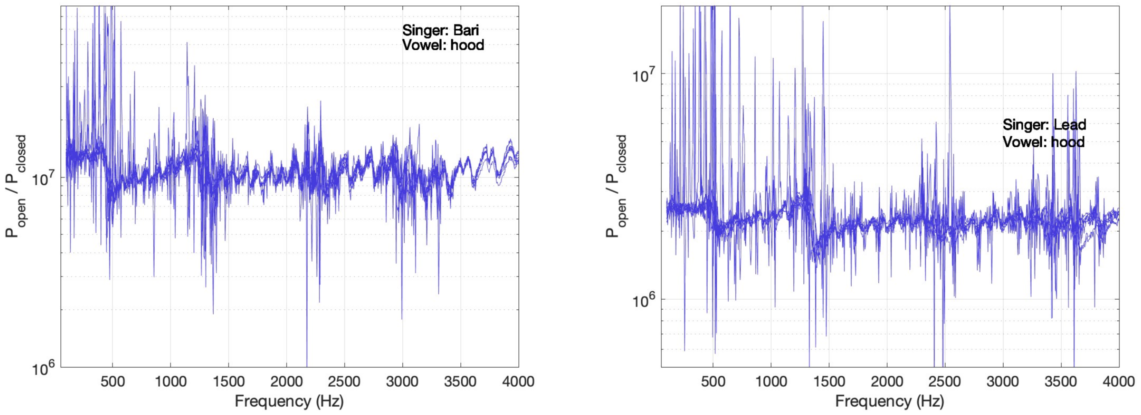


(c) bass (d) tenor

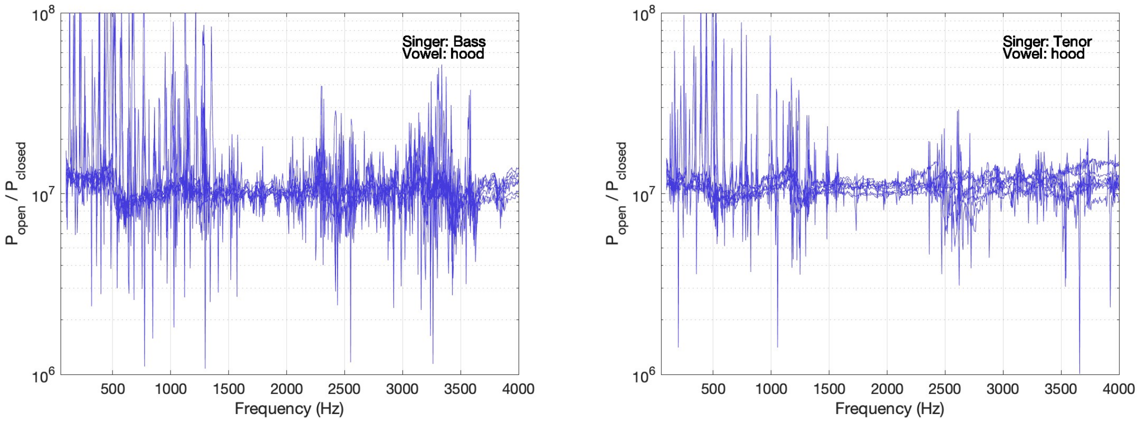


(e) tenor falsetto

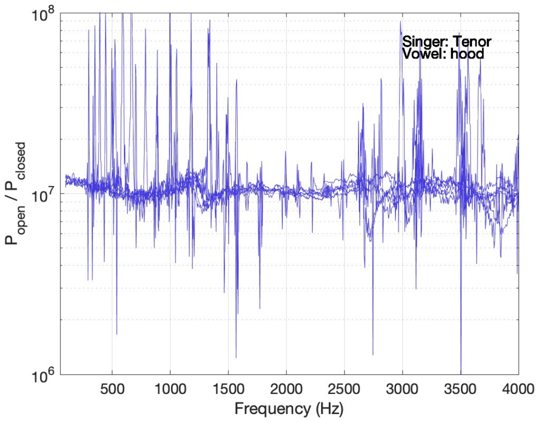
Figure C.1: RMD results on /i/vowel: (a) baritone (b) lead (c) bass (d) tenor (e) tenor falsetto



(a) baritone (b) lead



(c) bass (d) tenor



(e) tenor falsetto

Figure C.2: RMD results on /ʊ/vowel (a) baritone (b) lead (c) bass (d) tenor (e) tenor falsetto

References

Airaksinen, Manu et al. (2014). "Automatic estimation of the lip radiation effect in glottal inverse filtering". In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.*

Andersen, Tue Haste and Kristoffer Jensen (2004). "Importance and representation of phase in the sinusoidal model". In: *AES: Journal of the Audio Engineering Society* 52.11. ISSN: 15494950.

ANSI (1997). *American National Standard Methods for Calculation of the Speech Intelligibility Index.* S3.5. New York: Acoustical Society of America.

Arnela, Marc et al. (2013). "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations". In: *The Journal of the Acoustical Society of America* 134.4. ISSN: 0001-4966. DOI: `10.1121/1.4818756`.

Aspaas, Christopher et al. (2004). "Select acoustic and perceptual measures of choral formation". In: *International Journal of Research in Choral Singing.*

Barbershop Harmony Society (2023). *Contest and Judging Handbook.* Vol. Singing Category. URL: `https://files.barbershop.org/PDFs/Contests-Judging/CJ_Handbook.ver_15.4.pdf`.

Basinger, Lynn (2006). "Acoustical analysis of choral voice matching and placement as it relates to group blend and tone". PhD thesis. Texas Technical University.

Bernini, Antonio and Federico Talamucci (2014). "Consonance of Complex Tones with Harmonics of Different Intensity". In: *Open Journal of Acoustics* 04.02. ISSN: 2162-5786. DOI: `10.4236/oja.2014.42008`.

Bies, David A. et al. (2023). *Engineering Noise Control: Sixth Edition.* DOI: `10.1201/9780367814908`.

Bozeman, Kenneth (2014). *Practical Vocal Acoustics.* Pendragon Press, Hillsdale, NY. ISBN: 978-1-57647-240-8.

D'Amario, S and H Daffern (2017). "Using electrolaryngography and electroglottography to assess the singing voice: A systematic review." In: *Psychomusicology: Music, Mind, and Brain* 27.4. ISSN: 0275-3987. DOI: `10.1037/pmu0000184`.

Daffern, H and S D'Amario (2022). "Understanding expressive ensemble singing through acoustics". In: *Together in Music: Coordination, Expression, Participation*. Ed. by R Timmers et al. Oxford University Press. Chap. 16, pp. 129–138. DOI: `10.1093/oso/9780198860761.001.0001`.

Daugherty, James F. (1999). "Spacing, formation, and choral sound: Preferences and perceptions of auditors and choristers". In: *Journal of Research in Music Education*. ISSN: 00224294. DOI: `10.2307/3345781`.

Dayme, Meribeth Bunch (2009). *Dynamics of the singing voice*. DOI: `10.1007/978-3-211-88729-5`.

Devaney, Johanna et al. (2012). "A study of intonation in three-part singing using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT)". In: *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012*.

Doval, Boris et al. (Nov. 2006). "The spectrum of glottal flow models". In: *Acta Acustica united with Acustica* 92.6, pp. 1026–1046. ISSN: 16101928.

Epps, J. et al. (1997). "A novel instrument to measure acoustic resonances of the vocal tract during phonation". In: *Measurement Science and Technology* 8.10. ISSN: 09570233. DOI: `10.1088/0957-0233/8/10/012`.

Fant, Gunnar (1971). *Acoustic Theory of Speech Production*. DOI: `10.1515/9783110873429`.

– (1975). "Non-uniform vowel normalization". In: *STL-QPSR* 16.2-3.

Giannakopoulos, Theodoros and Aggelos Pikrakis (2014). *Introduction to Audio Analysis: A MATLAB Approach*. ISBN: 9780080993881. DOI: `10.1016/C2012-0-03524-7`.

Gill, Brian P. et al. (2020). "Spectrum Effects of a Velopharyngeal Opening in Singing". In: *Journal of Voice* 34.3. ISSN: 18734588. DOI: `10.1016/j.jvoice.2018.11.014`.

Goldstein J L (1970). "Aural Combination Tones". In: *Frequency Analysis and Periodicity Detection in Hearing*. Ed. by Plomp R and Smoorenburg G F. Leiden, Netherlands: A W Sijthoff, pp. 230–247.

Goodwin, Allen W. (1980). "An acoustical study of individual voices in choral blend". In: *Journal of Research in Music Education*. ISSN: 19450095. DOI: `10.1177/002242948002800205`.

Gregg, Jean Westerman and Ronald C. Scherer (2006). "Vowel Intelligibility in Classical Singing". In: *Journal of Voice* 20.2. ISSN: 08921997. DOI: `10.1016/j.jvoice.2005.01.007`.

Hagerman B and Sundberg J (1980). "Fundamental Frequency Adjustment in Barbershop Singing". In: *Journal Research in Singing* 4, pp. 3–17.

Harrison, Peter M.C. and Marcus T. Pearce (2020). "Simultaneous Consonance in Music Perception and Composition". In: *Psychological Review* 127.2. ISSN: 19391471. DOI: `10.1037/rev0000169`.

Havel, Miriam et al. (2023). "Effects of Nasalization on Vocal Tract Response Curve". In: *Journal of Voice* 37.3. ISSN: 18734588. DOI: `10.1016/j.jvoice.2021.02.013`.

Helmholtz, Hermann von (1877). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Trans. by A J Ellis. 4 edition. Dover Publications, Inc, New York.

Henrich, Nathalie, Christophe d'Alessandro, et al. (2004). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation". In: *The Journal of the Acoustical Society of America* 115.3. ISSN: 0001-4966. DOI: `10.1121/1.1646401`.

– (2005). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency". In: *The Journal of the Acoustical Society of America* 117.3. ISSN: 0001-4966. DOI: `10.1121/1.1850031`.

Henrich, Nathalie, John Smith, et al. (2011). "Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones". In: *The Journal of the Acoustical Society of America* 129.2. ISSN: 0001-4966. DOI: `10.1121/1.3518766`.

Henrich N Gendrot C, Michaud A (2017). *Tools for Electroglottographic Analysis: Software, Documentation and Databases*. URL: `http://voiceresearch.free.fr/egg/`.

Holmes, J. N. (1983). "Formant synthesizers: Cascade or parallel?" In: *Speech Communication*. ISSN: 01676393. DOI: `10.1016/0167-6393(83)90044-4`.

Howard, David M. (2007). "Intonation Drift in A Capella Soprano, Alto, Tenor, Bass Quartet Singing With Key Modulation". In: *Journal of Voice* 21.3. ISSN: 08921997. DOI: `10.1016/j.jvoice.2005.12.005`.

– (2023). "Influence of vowels on note pitching in a cappella vocal quartet performance". In: *Journal of Voice*. ISSN: 18734588. DOI: `10.1016/j.jvoice.2023.11.008`.

Howard, David M. and Jamie A. S. Angus (2017). *Acoustics and psychoacoustics*. Focal press. DOI: `10.4324/9781315716879`.

Jeanneteau M Hanna N, Almeida A (2019). "Open source software for estimating vocal tract resonances using broadband excitation at the lips". In: *Proceedings of the 19th Int Congress of phonetic Sciences*, pp. 2971–2975. URL: `https://www.phys.unsw.edu.au/jw/reprints/ICPhS19.pdf`.

Kalin, Gustaf (2005). "Formant frequency adjustment in barbershop quartet singing". PhD thesis. Stockholm: KTH Royal Institute of Technology.

Klatt, D.H. and L.C. Klatt (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers". In: *The Journal of the Acoustical Society of America* 87.2. ISSN: 0001-4966. DOI: `10.1121/1.398894`.

Klatt, Dennis H. (1980). "Software for a cascade/parallel formant synthesizer". In: *Journal of the Acoustical Society of America* 67.3, pp. 971–995. ISSN: NA. DOI: `10.1121/1.383940`.

Kob, Malte et al. (2011). "Analysing and Understanding the Singing Voice: Recent Progress and Open Questions". In: *Current Bioinformatics* 6.3. ISSN: 15748936. DOI: `10.2174/157489311796904709`.

Lohri, Angela et al. (2011). "Combination tones in violins". In: *Archives of Acoustics* 36.4. ISSN: 01375075. DOI: `10.2478/V10168-011-0049-1`.

Lottermoser, W and F Meyer (1960). "Frequenzmessungen an gesungenen Akkorden". In: *Acta Acustica* 10.3, pp. 181–184.

Marshall, A. H. and J. Meyer (1985). "Directivity and auditory impressions of singers". In: *Acustica* 58.3. ISSN: 00017884.

Mathworks (2021). *Matlab Audio Toolbox Feature Extraction*. URL: `http:www.mathworks.com`.

Morse, P. M. et al. (1970). "Theoretical Acoustics". In: *American Journal of Physics* 38.5. ISSN: 0002-9505. DOI: `10.1119/1.1976432`.

Nyman B, Nykanen A (Mar. 2025). "Loudness matching of complex tones simulating sounds from electric trucks". In: *Acta Acustica* 9. DOI: `10.1051/aacus/2024090`.

Peeters, Geoffroy (2004). *A large set of audio features for sound description*. Tech. rep. arXiv: `0611061v2 [quant-ph]`.

Peeters, Geoffroy et al. (2011). "The timbre toolbox: Extracting audio descriptors from musical signals". In: *The Journal of the Acoustical Society of America* 130.5, pp. 2902–2916.

Peterson, Gordon E. and Harold L. Barney (1952). "Control Methods Used in a Study of the Vowels". In: *Journal of the Acoustical Society of America*. ISSN: NA. DOI: `10.1121/1.1906875`.

Plomp, R. and W. J. M. Levelt (1965). "Tonal Consonance and Critical Bandwidth". In: *The Journal of the Acoustical Society of America* 38.4. ISSN: 0001-4966. DOI: `10.1121/1.1909741`.

Rasch, R and R Plomp (1999). "The Perception of Musical Tones". In: *The Psychology of Music (Second Edition)*. Ed. by Diana Deutsch. Second Edition. Cognition and Perception. San Diego: Academic Press, pp. 89–112. ISBN: 978-0-12-213564-4. DOI: `https://doi.org/10.1016/B978-012213564-4/50005-6`. URL: `https://www.sciencedirect.com/science/article/pii/B9780122135644500056`.

Richards, Jim (May 2001). *The Physics of Barbershop Sound*. Barbershop Harmony Society 110 7th Avenue North Nashville, TN 37203 USA.

Risset, Jean-Claude and David L. Wessel (1999). "Exploration of Timbre by Analysis and Synthesis". In: *The Psychology of Music*. DOI: `10.1016/b978-012213564-4/50006-8`.

Sandell, Gregory J (1995). "Roles for Spectral Centroid and Other Factors in Determining "Blended" Instrument Pairings in Orchestration". In: *Music Perception* 13.2, pp. 209–246.

Scholes, Percy A. (1970). *The Oxford companion to music*. 10th. Oxford University Press.

Schubert, Emery and Joe Wolfe (Sept. 2006). "Does timbral brightness scale with frequency and spectral centroid?" In: *Acta Acustica united with Acustica*. Vol. 92. 5, pp. 820–825.

Siedenburg, Kai, Ichiro Fujinaga, et al. (Jan. 2016). "A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology". In: *Journal of New Music Research* 45.1, pp. 27–41. ISSN: 0929-8215. DOI: `10.1080/09298215.2015.1132737`. URL: `http://www.tandfonline.com/doi/full/10.1080/09298215.2015.1132737`.

Siedenburg, Kai, Charalampos Saitis, et al. (2019). "The Present, Past, and Future of Timbre Research". In: DOI: `10.1007/978-3-030-14832-4{\_}1`.

Smith, Julius O (2007). *Introduction to Digital Filters with Audio Applications*. Center for Computer Research in Music and Acoustics, Stanford California.

Sundberg, J (1974). "Articulatory interpretation of the "singing formant"". In: *The Journal of the Acoustical Society of America* 55.4. ISSN: 0001-4966. DOI: `10.1121/1.1914609`.

– (1987). *The science of the singing voice*.

– (1988). "Vocal Tract Resonance in Singing". In: *The NATS Journal* 44.4, pp. 11–20.

– (Apr. 2019). "Intonation in Singing". In: *The Oxford Handbook of Singing*. Ed. by Graham F. Welch et al. Oxford University Press, pp. 280–296. DOI: `10.1093/oxfordhb/9780199660773.013.36`.

Sundberg, J and Jan Gauffin (1974). "An attempt to predict the masking effect of vowel spectra". In: *Speech Transmission Lab Quarterly Progress Status Report*.

Ternström, S (1991). "Physical and acoustic factors that interact with the singer to produce the choral sound". In: *Journal of Voice* 5.2, pp. 128–143. ISSN: 0892-1997. DOI: `https://doi.org/10.1016/S0892-1997(05)80177-8`. URL: `http://www.sciencedirect.com/science/article/pii/S0892199705801778`.

– (2003). "Choir acoustics : an overview of scientific research published to date". In: *International Journal of Research* 1, pp. 3–12.

Ternström, S et al. (1988). "Synthesizing choir singing". In: *Journal of Voice* 1.4. ISSN: 08921997. DOI: `10.1016/S0892-1997(88)80007-9`.

Titze, Ingo R. (2004). "A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice". In: *Journal of Voice*. ISSN: 08921997. DOI: `10.1016/j.jvoice.2003.12.010`.

– (2018). "Where has all the power gone? Energy production and loss in vocalization". In: *Speech Communication* 101. ISSN: 01676393. DOI: `10.1016/j.specom.2018.05.003`.

Titze, Ingo R., Ronald J. Baken, et al. (2015). "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization". In: *The Journal of the Acoustical Society of America* 137.5. ISSN: 0001-4966. DOI: `10.1121/1.4919349`.

Titze, Ingo R. and Lynn Maxfield (2017). “Acoustic factors affecting the dynamic range of a choir”. In: *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: `10.1121/1.5004569`.

Titze, Ingo R. and Anil Palaparthi (2018). “Radiation efficiency for long-range vocal communication in mammals and birds”. In: *The Journal of the Acoustical Society of America*. ISSN: 0001-4966. DOI: `10.1121/1.5034768`.

Titze, Ingo R. and Brad H. Story (1997). “Acoustic interactions of the voice source with the lower vocal tract”. In: *The Journal of the Acoustical Society of America* 101.4. ISSN: 0001-4966. DOI: `10.1121/1.418246`.

Vos, Rebecca R. et al. (2017). “Resonance Tuning in Three Girl Choristers”. In: *Journal of Voice* 31.1. ISSN: 18734588. DOI: `10.1016/j.jvoice.2016.01.013`.

Vurma, Allan and Jaan Ross (2006). “Production and perception of musical intervals”. In: *Music Perception* 23.4. ISSN: 07307829. DOI: `10.1525/mp.2006.23.4.331`.

White, Glenn and Kate Grieshaber (1980). “On the Existence of Combination Tones as Physical Entities”. In: *Journal of Research in Music Education* 28.2. ISSN: 00224294. DOI: `10.2307/3344821`.

Zwicker, E. et al. (1957). “Critical Band Width in Loudness Summation”. In: *The Journal of the Acoustical Society of America* 29.5. ISSN: 0001-4966. DOI: `10.1121/1.1908963`.