Soprano, Style and Voice Quality: Acoustic and Laryngographic Correlates

by

Laura Anne Bateman
B.Mus, University of Victoria, 1990

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in Interdisciplinary Studies

Co-Supervisors: Dr. W.A. Schloss and Dr. J.H. Esling

ABSTRACT

There are numerous widely varying vocal styles and voice qualities in Western music. Popular music in the 21st century uses a particular voice quality for female voice that is quite different from the trained classical voice quality. Classical voice quality has been the subject of a vast body of research, whereas research that deals with non-classical voice quality and pedagogy is very limited. In order to learn more about these issues, the author chose to do research using a variety of standard voice quality tests to substantiate the existing literature, and perhaps generate new information.

This thesis presents a review of the existing literature on voice quality in various different styles of singing: Classical, Belt, Legit, R&B, Jazz, Country and Pop. In addition, this thesis looks at spectral measurements from a small set of voice samples, elicited from a professional soprano. Laryngographic (LGG) data was generated simultaneously with the audio samples. To limit the data set for the scope of this thesis, singing samples using the vowel [i] are selected. The analysis techniques used in this thesis are Spectrogram, LPC, FFT, and various LGG ratio measurements. The spectral measurements compared include the relative strength of the first two harmonics, the formant locations, relative energy from harmonic strength near the formants, summed energy in two quadrants (0-3000Hz, 3000-5000Hz), and the inharmonic or aperiodic activity seen in each quality. Data from the LGG is used to calculate the contact quotient (the time the vocal folds are in contact divided by the time for one cycle of the vocal fold vibration), speed quotient (the time between maximum contact of the vocal folds and vocal fold separation divided by the time between first contact and maximum contact) and ascending slope (the slope of the contacting phase of the vocal fold wave). The LGG waveform was also visually assessed. The acoustic and LGG data are compared to an auditory analysis by Dr. John Esling (Professor of Linguistics, UVic) and to the subject's descriptions of the physical configurations involved in producing these qualities.

Physiological observations obtained from x-ray fluoroscopy & MRI scans of belt and classical voice qualities are included in Appendix B.

The intention of the thesis is to reveal more about the workings of these voice qualities. The thesis also serves as a prototype for a series of 10 vowels and running samples that were elicited at the same time. Perhaps, even though this thesis presents a limited data set, it may be useful to pedagogues who are struggling to understand the complexities of the non-classical female voice, as well as to computer programmers and engineers who are developing voice enhancement devices and biofeedback tools.

Table of Contents

A copy of the thesis and CD Rom is housed in the University of Victoria, McPherson Library collection in Victoria, BC, Canada. A second copy of the thesis and CD-Rom is housed in the Linguistics Department, University of Victoria, Victoria, BC, Canada.

List of Tables

List of Figures:

Acknowledgments

Most large projects are a synergy of many peoples combined energy. This thesis is no exception. I was fortunate to link up with Lisa Popeil through various National Association of Teachers of Singing conferences. She generously provided the voice samples. These raw samples had to be extracted from the DAT tape and saved as computer files. The computer files were trimmed and, in some cases, levels boosted before they were run through the analyses. Ellen Williams (nee Bjelica) spent many hours helping me prepare the samples for analysis. I also had to acquire a great deal of knowledge outside of my home department. I would like to thank the faculty and staff at the Linguistics and Engineering Departments for helping me, particularly those engineers that helped bridge the gaps in my knowledge between music, acoustics and engineering: Sami Saab, Dr. Dale Shpak, Prasad Sristi & Dr. Steve Mason, as well as the faculty at the Wilbur James Gould Center for Voice and Speech in Denver, Co. including Dr. Ingo Titze and Dr. Brad Story for increasing my knowledge of the physics of voice production and instrumentation for voice analysis. I would like to thank the supervisory team, Dr. John Esling and Dr. Andrew Schloss for their help throughout the project, the technical support staff in the Linguistic Dept., Jocelyn Clayards and Greg Newton. Lastly, I would like to thank my family including Bill and Kay Bateman, Derek and my daughter Emma whose appearance Feb 18, 2001 caused a slight delay in the completion of this project. Without all these people's contributions this thesis would not have been possible.

Chapter 1: Introductory Comments

1.1 Introduction

Anyone who has watched a classical singer struggle with a pop song will conclude that there is something very different about non-classical voice quality. Maultsby (1981; 1990) theorizes about the origin and evolution of musical styles and associated voice qualities in American popular music. She states that the African slaves came with a different musical style and associated voice quality than the European settlers. This seems like a plausible explanation for the evolution of popular music in America. Perhaps the Europeans liked the sound of this new voice quality and tried to imitate it. Maultsby uncovered incidences of recording studios demanding that white musicians copy black musicians in order to develop a new musical style to be marketed to white audiences. This type of theoretical reflection on American musical styles and associated voice qualities may stem from an observation that there has been a shift in listening trends. In general, the North American population of the first half of the twentieth century liked the classical voice quality, they studied the classical voice quality and they listened to the classical voice quality. The majority of show music repertoire for the female voice before 1930 was written for trained sopranos. This could be because a singer could not be heard in a large hall with any other voice quality. It is still the primary voice quality taught in North American singing studios. However, in the latter part of the twentieth century there was a radical shift away from classical voice quality in the listening audience. The development of amplification systems for singers may have been influential in this shift. Thomas Cleveland (personal correspondence) reiterated the results from a survey done in the 1990's revealing that sales in North America of classical vocal music are so low that the category didn't even appear on the ratings chart. Performers and listeners have overwhelmingly chosen the alternative non-classical voice quality and its variations. For example, singers will line up for hours to get a chance to audition for the TV show "Canadian Idol", which features a competition between pop singers. Non-classical voice quality has emerged as the predominant quality to the chagrin of the classical pedagogue. As a result, teachers of singing have been deluged with clients who want to sing popular music and learn the associated voice quality.

However, singing pedagogues have their reasons for teaching classical voice quality. For example, classical voice quality has a large body of literature, scientific research and an over 400-year-old history, whereas popular American non-classical voice quality has a 50-year-old western history, with little research or pedagogical documentation. Also, there is no conclusive evidence that non-classical voice qualities can be done safely. For example, many speech pathologists dealing with the performing arts see a large number of belters with nodules and other vocal abuse injuries. Is this because more singers are belting, or is belting technique more wearing on the voice than classical technique? These questions have not been addressed in scientific studies? Many pedagogues feel uneasy promoting voice qualities that may have associated health risks. Moreover, since the non-classical voice qualities are diverse, many pedagogues are reticent to deal with the diversities. In the last 10 years these trends in voice production have brought about many changes for singing pedagogues, including the upsurgence of clinics, workshops and conferences highlighting non-classical voice technique.

This thesis will deal with voice quality in singing to add clarity to existing non-classical singing pedagogy, compare the workings of the classical voice to the non-classical voice, find better ways to communicate voice quality issues between disciplines, and search for new insights on voice production. The intent is to help young performers learn to sing with non-classical voice quality, and to help singing pedagogues with the linkage between the sound of each voice quality and the physiological approach to producing them. Terminology is considered very important to the discussion of voice quality. In order to describe the non-classical voice quality well, some adjustments may need to be made to existing terminology and classification systems. This thesis does not try to address musical style by discussing articulation and other factors, such as vibrato even though articulation and vibrato affect the voice quality. For example, Country singers may use an American dialect from the southern United States in Country music, or Pop singers may use straight tone in sustained pitches at the start of the note that moves into vibrato at the end of the note. In addition, the issue of long-term vocal health is not discussed. The voice quality research in this thesis does not attempt to establish physiological correlates for the seven voice qualities; however, Appendix B contains x-ray fluoroscopy and MRI photos for Belt and Classical qualities in order to suggest a direction for further research. Instead, this thesis focuses on acoustic and laryngographic observations of seven distinct singing qualities and introduces ideas about taxonomy and voice production that may be useful in developing hypotheses for further research.

1.2 Thesis outline

The thesis presents a discussion of the interdisciplinary nature of voice quality terminology for speech and singing, followed by a discussion of related work on voice quality for classical and non-classical singing and speech. The latter half of the thesis is dedicated to exploratory research. The research focuses on acoustic and laryngographic correlates for seven voice qualities for one subject and one vowel [i]. The voice qualities used in this research were Belt, Classical, Country, Jazz, Legit, Pop, Rhythm and Blues(R&B) and Speech. A brief description of the voice qualities is included in Chapter 3. The thesis also includes four appendices on CD ROM containing voice samples, measurements, tables, and graphs.

1.3 Source Material

For classical voice, a number of texts have been published dealing with both spectral and physiological correlates. The texts used in this thesis were by William Vennard (1967), Richard Miller (1986, 1996), Johan Sundberg (1987) and Ingo Titze (1994).

William Vennard and Richard Miller have a background in singing pedagogy and voice science. Johan Sundberg has a music background with specialization in Acoustics, and Ingo Titze, a background in Engineering; therefore, the focus for these authors is quite different. The two dissertations by Ronald Bevan and Barry Bounous deal with issues related to Belt quality.

The bulk of information about popular singing quality is from conference papers and journal articles. The richest journal sources include Journal of Voice and the Journal of Singing (formerly the NATS Journal). The Journal of Voice is a medical and scientific journal, whereas the Journal of Singing is a National Association of Teachers of Singing (NATS) publication and deals with voice pedagogy. Other articles were obtained in publications such as: New Scientist, Scientific American, Australian Journal of Voice, Journal of Research in Singing and Applied Vocal Pedagogy, Medical Problems of Performing Artists, Journal of the Acoustical Society of America, Voice, and the Annals of Otology, Rhinology and Laryngology

Among the researchers publishing journal articles are Jo Estill, Johan Sundberg, Harm Schutte, Robert Miller, Ingo Titze, Richard Miller, and Thomas Cleveland. Some information is available only to certified instructors of that particular method due to marketing strategies of various individuals. This includes material from Jo Estill, who has marketed her findings under the name of *Voicecraft* and publications are available only to the participants of her workshops; however, a few of her published journal articles are available.

The speech literature, although not directly related to voice quality in singing, often can be extrapolated to provide clues to the acoustics of the singing voice. The texts used in this thesis were by Laver (1980), Ladefoged (1993, 1996), Fry (1979), and Baken/Orlikoff (2000). Literature was also gathered from conference proceedings and journal articles. Journal articles were collected and examined from publications such as: Journal of the International Phonetic Association, Journal of Speech and Hearing Research, Phonetica, Language and Speech, and Journal of Phonetics.

1.4 Delimitations

The research in this thesis is based on only one subject, and on a limited number of voice samples; therefore, it may or may not hold true when taken to a larger sample. However, the subject, Lisa Popeil, is a professional singer and teacher, with many years experience singing different styles for the advertising industry in Los Angeles, California, as well as concertizing with such notable artists as Frank Zappa and Al Yankovic. Therefore, the samples are believed to approximate the norm for the singing voice qualities considered in this thesis. A further consideration is that the literature on singing voice quality is very limited. Only two dissertations by Bounous (1997) and Bevan (1989) were found pertaining to voice quality in non-classical singing, and a small number of journal articles and conference papers. Moreover, scientific and pedagogical texts were used that pertain to classical quality, since it is the only quality that has been well documented in book form.

Another limitation is that stylistic definitions for non-classical voice are still vague in both pedagogical and scientific circles, and there is much crossover between voice quality and musical style. Confusion also arises through the terminology used in many articles, since terminology differs with the researcher's personal background. Terminology is constantly being updated so some terms may even have been adjusted recently. Finally, it is difficult to ascertain which acoustic events are singing quality cues, and which are simply timbral characteristics of Lisa's voice. This will remain a question for future research using a larger sample.

Chapter 2: Terminology

2.1 Introduction

The interdisciplinary study of the human voice involves differing and often inconsistent terminology. Researchers from a variety of backgrounds: medicine, psychology, speech science, articulatory phonetics, theatre, music, and engineering acoustics, and use a range of terms that often describe similar phonetic events, sounds and postures. Voice quality is perceptual. Just as human beings categorize frequencies of light by colour, we also categorize voice qualities. One theory proposes that we categorize in order to try to simplify the filing process in our brain. Titze (1994: 331), a voice scientist at University of Iowa, explains this phenomenon with the term *categorical perception* or " the perception of discrete entities in a physical continuum... This is one way the brain can simplify the number of items it must identify, store and process"

Linguistics categorizes voice quality by using phonetic labels that are based on the concept of *settings.* These settings are either genetically predetermined configurations, or acquired habitual postures that diverge from the *neutral voice*. They are overlaid on, or abstracted from the speech segments, and produce a similarity in the voice quality over the duration of the speech sounds. For example, one person may talk with their lips consistently pursed as they speak, whereas another may have a voice quality reminiscent of a creaky door. The settings are further divided into laryngeal and supralaryngeal settings. Laryngeal settings describe events at the level of the larynx, and supralaryngeal settings describe events above the larynx. The associated voice qualities are given a label. For example, the speaker with creaky door speech would have a voice quality label called *creaky voice* under the category of laryngeal settings; whereas, the speaker with pursed lips would have a voice quality label *labial protrusion* under the category supralaryngeal settings. Linguists may also use scalar degrees between 1 and 6 to describe the quantity of the setting, the scalar degree 1 being very slight deviation from neutral, and the scalar degree 6 being very extreme deviation from neutral. The linguistic phonetic taxonomy is commonly used to describe voice quality and phonatory settings, and is always expanding to improve the accuracy of these descriptions (Esling, 2003).

Singing pedagogues divide voice qualities into groups of voice sounds called "*registers*". Titze (1994) describes registers as "perceptually distinct regions of vocal quality as pitch or loudness is changed" (p. 335). There are four registers in singing: *fry, chest, head* and *falsetto* (male)/*flageolet* (female). The two most used are chest and head registers. Singers often use labels that are linked with body sensations, since they cannot perceive their own sounds with accuracy. For example, it is common to refer to the low register of the voice as chest voice because this type of vocal production causes sympathetic vibrations in the bones of the chest, bronchi and trachea; in contrast, the high register of the voice is referred to as head voice because the resonant sensations are felt in the head, the palate and sinuses in the face. Furthermore, if a voice quality such as chest voice occurs over a pitch range, then it would be called the *chest register*. Terminology surrounding registration is very similar to the linguistic laryngeal terminology because it labels voice qualities that are associated with particular vibratory patterns of the vocal folds. Singers also use various vocal tract settings, or supralaryngeal settings; however, singers don't have a terminology for the supralaryngeal settings. Perhaps they should, since they consistently apply supralaryngeal settings to the laryngeal settings to create the voice qualities we perceive as various singing qualities. Classical and Belt are two common voice qualities in singing that are used in North America. Classical voice quality has a pure flutey sound. Classical voice quality is used in opera, operetta, art song and contemporary art music, and some musical theatre, whereas Belt voice quality has a buzzy, brassy sound. Belt voice quality is used in popular music and musical theatre. For example, Teresa Stratus sings with a classical voice quality, whereas Celine Dion sings with a belt voice quality.

Since there are apparent similarities in the way speech and singing are categorized, the author has chosen to organize the following section of the thesis using phonetic terminology. This section will explain some of the terms common to the North American schools of singing and equate them to the auditory labels commonly used in the British school of phonetics.

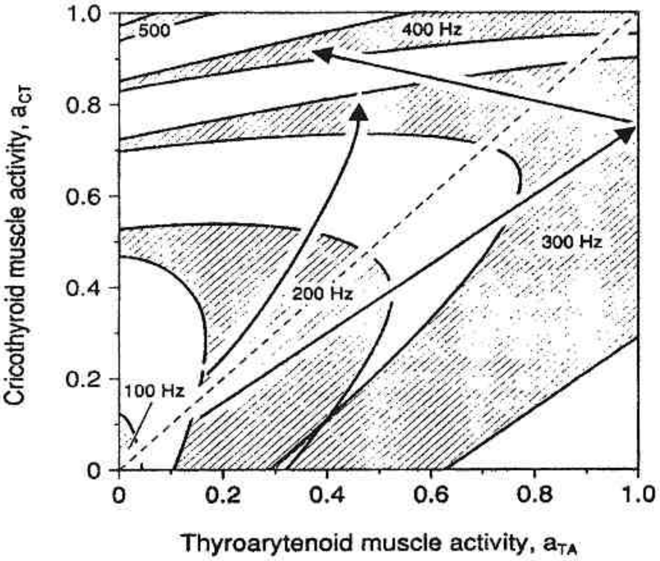
2.2 Laryngeal Terms

One complication in the laryngeal terminology for singing is that categories of voice quality are not universally agreed on. Some pedagogues believe there is only one blended register. Others believe in two or more. Many have elaborate charts with registration events and registers noted for each voice classification or *fach*. However, those pedagogues who teach registers believe that for singers to ascend in pitch they must change register. They describe the process of "gradual register transition" (Miller, 1986) as *vocal registration.*

Voice scientists have provided two theories to explain the concept of registration in singing. The first theory has to do with the coordination between the cricothyroid

(CT), and the thyroarytenoid muscles (TA) (Hirano et al., 1970). For example, singers gradually relax the TA (vocal folds) as the CT (cartilage-tipping muscles) gradually increases activation. The thyroid cartilage tilts forward, gradually stretching the vocal folds, thereby increasing tension, for a smooth ascent of pitch. Figure 1 shows the result of research done with excised dog larynges in a plot of the activation of the CT muscles against the activation of the TA muscles. The shaded sections show a wide variety of combinations that can produce the same pitch. The straight arrows show that when the TA muscles reach a point where they can no longer withstand the tension applied by the CT muscles, the TA muscles release all of a sudden to produce the appropriate pitch. This would be perceived as a registration event or yodel. The yodel is not encouraged in classical singing; however, in folk and popular singing it is often a prized technique.

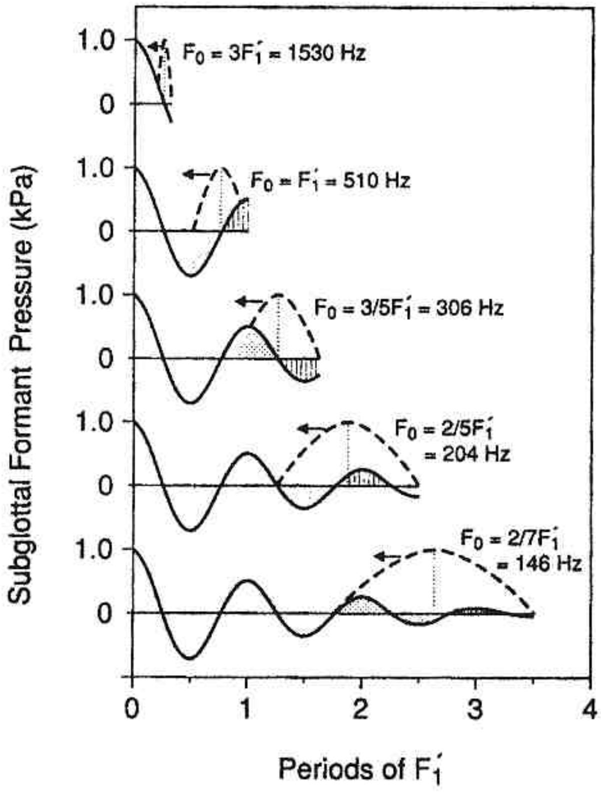
Figure 1: Muscle activation plot (MAP) showing abrupt register transition (two straight arrows) and gradual transition (curved arrow).



(Titze, 1994)

The second theory has to do with subglottal acoustic waves that set up an interference pattern with the glottal wave (Titze, 1983, 1988; Austin, 1992). "Acoustic pressures below the vocal folds can be phased in such a way that they contribute, constructively or destructively to the intraglottal driving pressures of the vocal folds" (Titze, 1994: 263). Figure 2 illustrates the interference between the subglottal formant and the glottal wave.

Figure 2: Phase relationships between the pressure waveform of the first subglottal formant F1 (solid lines) and the glottal area waveform (dashed lines) for systematically increasing fundamental frequency F0 (bottom to top).



(Titze, 1994)

Places of acoustic interference correspond to *passaggi, register breaks, or registration events* in the singing terminology.

Registers are not commonly discussed in acoustic phonetics; however, the voice qualities associated with these registers are used extensively. The labels are different, but the qualities described are similar. In the British school of phonetics Laver (1980) refers to a voice quality and phonation type called *modal voice*. He describes modal voice as the "neutral mode of phonation having moderate adductive tension and moderate medial compression, with moderate longitudinal tension" (p. 111). He states that this phonation type "essentially corresponds to chest voice" (p. 110), but says that it could be differentiated into two sub-types 'chest voice' and 'head voice'. This is a curious statement since many singers perceive head voice as quite a distinct voice quality. Miller (1986) refers to a similar voice quality as *chest voice* or *voce di petto*. Miller states that chest voice is characterized by "heavy action from the TA muscles: wider amplitude of vibration, thicker and shorter folds". He mentions that sympathetic resonant vibrations are felt in the chest and body, especially in the trachea and bronchi and the larger bones of the rib cage. The term chest voice also refers to a phonation type in singing, which is characterized by a thick vocal fold with a vibratory pattern that involves first contact of the lower edges of the fold alternating with contact of the upper edges of the vocal fold. This vibratory pattern produces a complex glottal wave in the mucosal layer of the vocal fold. Titze (1994) describes this phonation type by saying "... it is suspected the bottom of the vocal fold is adducted more in this register than in falsetto and that the thyroarytenoid muscle bulges the vocal fold medially below the level of the vocal processes. This creates a thicker and deeper vibrating structure. The entire cover (including the ligament) is lax and the thyroarytenoid muscle is used to regulate effective tension of the vocal fold." Often this vocal fold adjustment is referred to as modal (Hollien, 1972; Titze, 1994).

In summary, the term chest has been used to describe a register (chest register), a voice quality (chest voice), and a phonation type (chest voice) and a vocal fold adjustment (chest), and this term may relate closely to the term modal such as in Laver's voice quality (model voice) and phonation type (modal) as well as Titze's vocal fold adjustment (modal).

In phonetics, there are two voice qualities that arise from the vocal fold thickness and subsequent mode of vibration: modal voice and *falsetto* (Laver, 1980). In singing, students are encouraged to thin the vocal folds gradually over the pitch range. This may involve an infinite continuum of vocal fold thicknesses controlled by the gradual activation of the CT muscles with a gradual relaxation of the TA muscles. At some given point along this continuum the body of the vocal fold stops vibrating and only the ligament vibrates (Laver, 1980). We perceive this as a quality change because the mode of vibration changes, thereby changing the relative strengths of harmonics (Titze, 1994).

The transition from thick to thin is a difficult thing to study. Usually researchers study muscle activation using an electromyograph (EMG). For this procedure, the researcher inserts hooked electrodes into the various muscles of a singer, and monitors the change in electrical current. This is an invasive technique, and for this reason, much less research has been done. One way researchers have coped with this problem is by using excised dog larynges shown in the research presented in Figure 1. More recently researchers have turned to computer modeling techniques.

The phonetic auditory label for the voice quality produced when only the ligament is vibrating is falsetto. Laver (1980: 118) states that the vertical cross-section of the edges becomes thin and the glottis remains slightly apart … the vocalis muscle is relaxed and only the thin margins of the vocal folds participate in phonotory vibration. This phonetic term does not correspond closely to the term falsetto in singing, where falsetto is reserved for describing the top register of a male voice. Vennard observed two different mechanisms for falsetto (Vennard, 1967). The first mechanism is similar to the phonetic definition. It is thought to be associated with untrained singers, and perhaps may not produce the best quality for singing. The second mechanism uses a partial damping of the vocal folds. There is an increase in the compression of the arytenoid cartilages that causes half of the vocal fold to adduct, and the vibration only occurs in the anterior part of the vocal fold (Vennard, 1967). There are not enough data to predict the numbers of male falsettists who use each technique; however, generally these male falsettists belong to a voice type called *countertenor*. Ladefoged (1993: 141) describes a mechanism in association with creaky voice that he calls *laryngealization*:

> The arytenoid cartilages are tightly together, so that the vocal cords can vibrate only at the other end.

Perhaps damping can occur in all registers. In creak there are other supralaryngeal factors, such as pharyngeal constriction, associated with the phonation type. Perhaps damping can occur apart from the pharyngeal constriction found in creak. The closest female counterpart to falsetto in the singing terminology may be the voice quality produced in the *flageolet* register. Vennard describes the same damping mechanism as for the male singer. If this is true, then before the damped falsetto occurs, perhaps the female singer at some point uses only the ligament for vibration. Could this correspond in females to head voice (voce di testo) above the register change (secondo passaggio)? This explanation would explain the voice quality difference, and also the difference between male and female registration. Would this pure head voice then be falsetto in phonetic terms? Miller (1986: 133) disagrees: " the term falsetto should be reserved to designate the imitation of female vocal quality by the male voice."

Miller (1986) indicates that the range of chest voice (voce di petto) for a soprano is (G3-Eb4) and that the range of head voice (voce di testo) is (G5-C6). However, this leaves over one octave without a designation. Singers negotiate this area using a *mixed registration*. Titze (1994: 275) explains that this occurs in the middle range in female singers, and in the top range in male singers. Keidar et al. (1987) performed some perceptual studies on the nature of vocal register change. A perceptual register change in both men and woman was found around 337 Hz. In females, this would be approximately the beginning of head mix in the middle range. In men, this would be the upper register. However, this was a perceptual test so there were no data linking the perception to the mechanism. Keidar called the perceptual change the transition to falsetto. Vilkman et al. (1995) studied the mechanism with LGG and glottal flow rates, and theorized that when the vocal folds reach a critical mass they can produce a vibratory pattern in the mucosa that is rich in harmonics, and that we perceive as chest voice. If the vocal folds are without this critical mass in the vertical dimension, then the voice quality is perceived as falsetto. These two accounts of falsetto seem to be more related to the idea of the middle or mixed voice (Titze, 1994) or in singing voce mista (Miller, 1986) than to Laver's description of falsetto. Titze (1994) indicates this mixed voice could be called head voice, which could indicate why Laver specifies modal voice as chest or head. He may be including voce mista. Miller, on the other hand, represents the vocal fold variations within this register with the terms chest mix (heavy mechanism) and head mix (light mechanism). These terms indicates the relative thickness of the vocal folds and corresponding sensations in the body. Chest mix would indicate relatively thick vocal folds and sympathetic vibrations felt in the chest. Head mix would indicate relatively thin vocal folds with sympathetic vibrations felt in the head. To complicate matters sometimes singing teachers use qualifiers. For example, a teacher might say, "She sang that note in open chest". This would correspond to Laver's modal voice. The teacher might then say, "Try not to use such a heavy mechanism." This would direct the student to thin the vocal folds. Would this correspond to Laver's sub type modal voice II (head voice)? To directly correspond to Miller's labels, Laver's sub-types would have to have sub-sub types: modal voice I (open chest), modal voice IIa (chest mix), modal voice IIb (head mix), modal voice III (head – male). These auditory qualities are distinguishable and have been used as descriptive terms by voice teachers over the centuries. However, even by expanding Laver's labeling scheme, many pitches can be sung with a variety of mixes. Vennard (1967: 77) also offers these generalizations:

> one as to pitch, one as to intensity, and a third as to quality. First, to develop the widest possible range without a break, the adjustment must be heavy in the lower part of the voice, and the balance should shift smoothly toward the lighter production as the scale is ascended. Second, on any given pitch, the softer it is, the lighter must be the production without breathiness; and the louder, the heavier. Third, to produce "rich" timbre the adjustment should be heavy; to produce "sweet" timbre, it should be light. We have seen that the differences in timbre are differences in degrees of regularity and irregularity in the pattern of each vibration.

Perhaps Vennard uses heavy to describe the chest voice adjustment and light to describe the head voice adjustment. In Classical singing, a lighter mechanism is usually preferred, although mature dramatic singers may use a heavier mix. In popular singing, the Broadway belt quality setting of Ethel Merman or Barbara Streisand may use a heavier mechanism than the corresponding classical quality setting for the same pitch. However, belters still are encouraged to adjust to a thinner vocal fold with ascending pitch. More research is needed in this area to clarify the terminology of register transitions in non-classical voice. Estill (1996: 243) indicates that belt voice uses modal adjustment, but she also indicates that opera uses model adjustment. However, she doesn't describe which pitch range. She also distinguishes the mature dramatic opera quality from the lyric and lieder singer quality, which she calls "sob".

The top register in the female voice is called the *flageolet* or *whistle* register. Flageolet register is used extensively by *soubrette* and *coloratura* voices. The mechanism observed by Vennard is similar to that of the countertenor. Miller (1986) agrees with Vennard. He states that the flageolet voice has a high rate of longitudinal tension in the vocal ligaments, considerable damping of the posterior portion of the vocal folds, limited vibrating mass of the vocal folds, with high subglottal pressure and airflow rate. However, other voice scientists feel that the mechanism for female flageolet register is unclear (Sundberg, Titze, Verdolini, personal correspondence). There are three theories as to the mechanism of this register. The first theory is that the vocal folds may eventually stop vibrating, and the tone is created by a small hole referred to as a *chink* between the arytenoid cartilages. The second theory is that the tone is created by a fine 1 mm slit between the vocal folds with the ligament involved in vibration. The third theory is that females may at some point use damped falsetto. Vennard (1967: 67) questions the glottal chink mechanism:

> in rare cases true whistle does occur through an opening between the arytenoid cartilages (the mutational chink) but this is not very loud, and is not useful for singing. We may ignore it.

If Laver's term falsetto was split into undamped falsetto (untrained male falsetto and perhaps female head voice) and damped falsetto (countertenor and perhaps flageolet) then, the phonetic terminology could be expanded to describe the mechanisms found in singers' top register.

Thus far only two of the four descriptive phonetic auditory labels for voice quality laryngeal settings have been discussed: modal and falsetto. Two other simple phonation types are *creak* and *whisper*. The term creak does not appear in singing terminology. Often this phonation type is referred to as *fry*. Creak or fry vocalizations sound like the labels suggest: a creaky door, hot oil crackling in a pan, a stick being run along a railing. Classical singing never uses this phonation type. It is considered pathological, and much time is devoted to training the voice to coordinate the onset of phonation to avoid fry; however, pop singers use creak extensively as a means of artistic expression, especially as an effect at the beginning of a phrase. Usually in pop songs the onset will start with creak, then move through creaky voice to modal voice or whispery voice. Singers often accompany the creaky voice onset with a scoop, which means the voice starts with a glissando up to the appropriate pitch. Creak is also common in speech, but creak alone is not considered disordered speech (Kent and Ball, 2000: 46). As with modal and falsetto, creak can be used in reference to voice quality or phonation. In voice science, the register for creak is known as *pulse*. Titze (1994) states that it is the lowest register of the voice. He describes the pulse register as the perceptual result of subharmonic or chaotic patterns in the glottal waveform if a subharmonic is below about 70 Hz. He calls it a register with perceived temporal gaps. He means that when the vibration of the vocal folds is less than 70 Hz we can hear the spaces in between the vibrations, and perceive each individual wave of the vocal folds instead of perceiving a constant sound. Other terms for creak are *glottal fry* and *vocal fry* (Hollien, 1972). In phonetics, creak phonation may also occur with modal voice yielding *creaky voice* or, with falsetto yielding *creaky falsetto*. Moore (1971) suggested the following physiological mechanism based on frontal stroboscopic laminagrams:

> vocal fry may……be produced when the mass of the vibrators is increased by the collaboration of the ventricular folds. These structures appear in x-ray photographs to combine the ventricular folds functionally with the vocal folds to form massive bilateral vibrators that move with relatively small amplitude. It is presumed that this mechanism is capable of both impeding the flow of air, even when there is considerable pressure, and of releasing a series of pulses in which the channel is open for relatively short portions of the cycle.

Another interesting finding by Hollien et al. (1969) suggests that in vocal fry, neither length nor thickness of the vocal folds changes as pitch is varied. When creaky voice is analysed with the electroglottogram, the resulting conductance waveform shows a big wave followed by a smaller wave. This could be due to ventricular folds damping the action of the vocal folds; however, more research is needed in this area to clarify these issues.

The fourth simple phonation type in phonetic terminology is *whisper*. Titze (1994) describes whisper as the sound created by turbulent glottal airflow in the absence of vocal fold vibration. Laver (1980: 121) describes the glottal configuration as "a triangular opening of the cartilaginous glottis, comprising about a third of the full length of the glottis". Whisper may be combined with modal voice or falsetto to form whispery voice or whispery falsetto. Whisper is rarely used in Classical singing, however there are occasional occurrences in contemporary twentieth century art music. In popular singing, mixing air into the sound (*air mix*) is common practice and adds another colour to the singer's palette (Popeil, 1996). Whether air mix corresponds to whispery voice or another phonetic auditory label, breathy voice, depends on the supralaryngeal settings that accompany the sound (Esling, 2003). For example, in Country voice quality the high laryngeal position and pharyngeal constriction may produce a sound that corresponds more closely to whispery voice; however, the Rhythm and Blues voice quality may involve a low laryngeal position and an expanded pharyngeal space, which may produce a sound corresponding more closely to the *breathy voice*.

Phonetic laryngeal terminology also includes compound settings. In the preceding discussions some of these terms have already been used. Compound settings are those that combine the simple phonation types (modal, falsetto, creak and whisper) with each other, or with other parameters. For example, whispery voice combines modal voice with whisper. In popular singing, many of the compound laryngeal settings are used. The diversity of sounds made by the singing voice is as varied as with the speaking voice. However, this does not mean that they would be taught in a voice studio. For example, the compound phonation type *harsh voice* is considered disordered in all singing styles. Harsh voice is a phonation type that has aperiodic modes of vibration (Laver, 1980: 127). Acoustic tests show irregularity of the glottal waveform (jitter) and spectral noise. Titze (1994) says that this is a result of constricted glottis and insufficient airflow. The result is, in Titze's words, "a ghastly sound" (Titze, 1995). In some styles, such as heavy metal, an unpleasant, rough, rasping sound may be desirable, but generally the voice teacher's job is to improve the quality of a voice. From a voice teacher's perspective, I would prefer that singing students use signal processing to degenerate the sound quality rather than practicing ways of producing harsh voice. It is useful for a singing teacher however, to identify the auditory quality and be able to prescribe ways to help encourage periodic vibration without any excessive medial compression of the vocal folds. Maybe this is an argument for including harsh voice in the singing terminology. Singing pedagogues use the term *pressing* or *pressed voice* to describe excessive medial compression that results in a harsh timbre.

To summarize, acoustic phonetics has descriptive laryngeal auditory labels for voice quality in speech. They are outlined in Table 1 (Laver, 1980). The simple phonation types and some of the compound phonation types may have equivalent terminology in singing. They are outlined in Table 2.

Table 1: Auditory labels for voice quality in speech

Simple phonation types	Compound phonation types
Modal Voice Falsetto Whisper Creak	Whispery creak Whispery voice Whispery falsetto Creaky voice Creaky falsetto Whispery creaky falsetto Breathy voice Harsh voice Harsh falsetto Harsh creak Harsh whispery voice Harsh whispery falsetto Harsh creaky falsetto Harsh whispery creaky voice Harsh whispery creaky falsetto Ventricular voice

Table 2: Auditory labels for voice quality: a comparison between singing, voice science and phonetic terminology

Singing	**Voice Science**	**Phonetic**
Fry	Pulse	Creak Creaky voice
Chest/voce di petto/Heavy Open Chest	Modal	Modal voice
Mixed Voice/voce mista - Chest mix - Head mix	Mixed voice - Head/Falsetto (female middle range) - Head/Falsetto (male top range)	Modal voice
Head/voce di testo/Light		Modal voice (perhaps Falsetto in females)
Falsetto (male) Flageolet/Whistle (female)	Falsetto/Loft Whistle (female)	Falsetto Falsetto

Key questions remain about mechanisms for various phonation types in singing. First, when does the vocal fold vibration change from the body vibrating to just the ligament vibrating in the female voice? Second, it has been established that the ear perceives a difference from the body of the vocal fold vibrating to the ligament vibrating, but does the ear perceive a difference in quality as the vocal folds move from a near square shaped configuration in cross-section to a more triangular configuration? Third, do male and female singers both use damped falsetto in the extreme top register? If Vennard's (1949) observations hold true in a larger sample, and trained male falsettists and trained female coloraturas use damped falsetto, then there is no equivalent term in Laver's settings. However, Ladefoged's (1993) description of laryngealization may be a way to handle the terminology, but the supralaryngeal constriction and reduced airflow also play a part in the term. Esling and Harris (2003) propose a new system that could explain these relationships. It may not be possible to add to the phonetic terminology to handle the singing voice until there is further research to establish the mechanisms that male and female singers use throughout their full range. In the meantime, I propose a theory that could be tested in future research.

The current research seems to suggest that there could be possible phonetic terms, and associated theories for mechanisms. A new phonetic term, Testo voice, could be added to Laver's labeling scheme as he had proposed with the subtypes of modal voice: chest and head. Since the shorthand for head voice (HV) corresponds with the label for harsh voice (HV), I propose that the word testo be taken from the Italian word for head. I also proposes the terms damped and undamped be used to distinguish the potentially different mechanisms observed for falsetto singing. My classification system for singers would include the following terminology:

> **Creak (C):** voice quality like a rapid series of taps created by a low fundamental frequency and temporal gaps. This voice quality is used extensively in pop singing.
>
> **Modal voice (MV):** a robust voice quality produced when the vocal folds exhibit near square-shaped edges in cross-section. This mechanism works best for low pitches in male and female singing and may be used at higher pitches in non-classical singing such as belting.
>
> **Testo voice (TV):** a pure voice quality produced when the vocal folds exhibit near triangular-shaped edges in cross-section. This mechanism works best for middle register of female classical voice, and the high registers of male singers and female belters.
>
> **Falsetto (F):** an extremely pure voice quality that is produced when the vocal folds exhibit near triangular-shaped edges, and only the ligament vibrates. This mechanism is found in untrained male falsettists and female classical singers singing above the secondo passagio.

Damped testo voice (DTV): a voice quality characterized by a damping of the posterior end of the vocal folds with firm adduction of the arytenoids and anterior vibrations. The vocal folds exhibit near triangular-shaped edges in cross-section. This quality is found in male falsetto singing in the low register.

Damped falsetto voice (DF): a voice quality characterized by a damping of the posterior end of the vocal folds with firm adduction of the arytenoids and anterior vibrations. The vocal folds exhibit a near triangular vocal fold cross-sectional shape where only the ligament vibrates. This quality is found in high male falsetto singing, and in female whistle register singing.

2.3 Supralaryngeal Terms

Supralaryngeal terms describe voice qualities that result from the shape of the vocal tract above the level of the larynx. There are very few of these terms in singing. There may be reasons for this. In traditional classical pedagogy there is a particular shape that will produce the appropriate resonance and voice quality. Most often, singing teachers rely on imagery and analogy to try to describe the shaping of the vocal tract. However, non-classical singing uses a variety of vocal tract shapes that result in a variety of voice qualities. Perhaps a terminology will emerge as the need to communicate different vocal tract shapes and associated voice qualities in singing increases. For example, Popeil (1996) has developed her own terminology including "*ick face*" and "*water in the mouth*" to describe the shapes a singer needs to produce variations in voice quality. "Ick face" is used for classical voice quality and "water in the mouth" is used for Rhythm and Blues (R&B) voice quality. Phoneticists are ahead of the singers in this area. Laver (1980) has proposed a precise way of describing the voice quality associated with vocal tract shape (Table 3).

Table 3: Descriptive supralaryngeal auditory labels for voice quality in speech (Laver, 1980).

Labels for Supralaryngeal Settings	
Settings of the longitudinal axis of the vocal tract	
Labial	labial protrusion labiodentalized
Laryngeal	raised larynx lowered larynx
Settings of the latitudinal axis of the vocal tract	
Labial	Horizontal expansion of the interlabial space Vertical expansion Horizontal constriction Vertical constriction Horizontal expansion and vertical expansion Horizontal constriction and vertical constriction Horizontal expansion and vertical constriction Horizontal constriction and vertical expansion
Lingual	
Tip/blade	Tip articulation Blade articulation Retroflex articulation
Tongue-body	Dentalized Alveolarized Palato-alveolarized Palatalized Velarized Uvularized Pharyngalized Laryngo-pharyngealized
Tongue-root	Advanced tongue root Retracted tongue-root
Faucal	Faucalized
Pharyngeal	Pharyngalized
Mandibular	Close jaw position Open jaw position Protruded jaw position
Settings of velopharyngeal system	
	Nasal Denasal
General labels	
Labial	Open rounding Close rounding Spread lips
Overall	Tense voice Lax voice

Laver (1980: 159) states these terms are "suitable for a precise discussion of the detailed articulatory aspects of the settings". The phonetic terms that may be most useful to the discussion of voice quality in singing are *raised larynx* and *lowered larynx*, and terms describing the shape of the pharyngeal space including *advanced tongue-root, retracted tongue-root*, *pharyngealized*, and *laryngo-pharyngealized.* Other terms that may be useful include *labial*, *mandibular, faucal* and *velopharyngeal* setting labels and overall settings *tense* and *lax*. It is unclear how one would describe ick face or water in the mouth in Laver's (1980) terms; perhaps, as with the laryngeal terms, these settings would have to be expanded to cover the various postures and associated voice qualities common to singing.

Amongst the few supralaryngeal terms in singing are those that deal with the larynx. Singing pedagogues often mention a *stabilized larynx*. This means that the larynx is anchored by the laryngeal musculature so it moves very little with changing pitch. Estill (1988) often uses the term *anchored.* As well as the stabilized larynx, classical singers try to attain a spacious feeling in the throat: *open throat, garden breath, yawny, gola aperta.* The classical quality also encourages low diaphramatic breathing that causes the lungs to exert a downward pull on the trachea. Diaphragmatic breathing encourages the lowered laryngeal setting, thereby increasing space in the pharynx.

The low larynx is usually accompanied by the expansion of the piriform sinuses. The muscles that lower the larynx connect the hyoid bone and thyroid cartilage to the sternum and shoulder blades: sternohyoid, sternothryoid, and omohyoid muscles. Laver (1980: 31) notes "speakers who used lowered larynx often seem[ed] to adopt a posture with their chin 'tucked in', as it were, together with a slight rotation downwards of the head". Classical pedagogues often describe a stretch in the back of the neck that may be a similar phenomenon. Sundberg has observed the lowered larynx in classical voice quality as well as the raised larynx in belting quality. He also noticed that the piriform sinuses were collapsed in many vowels in the belting quality (Sundberg et al., 1993: 309).

There are two mechanisms for raising the larynx. The first mechanism involves stabilization of the hyoid bone and the larynx being pulled up toward it. The second mechanism involves raising both the hyoid bone and the larynx together. Laver's (1980) auditory labels don't differentiate between these mechanisms. One must assume that the two mechanisms produce the same voice quality. The voice quality associated with a higher than neutral larynx is labeled raised larynx. The voice quality associated with a lower than neutral larynx is labeled lowered larynx. However, Laver (1980) does use scale degrees or qualifiers such as slight, moderate and extreme. The voice quality of the lowered larynx sounds darker and richer, whereas the raised larynx quality often sounds strained. Laver (1980: 310) cites that Van Riper and Irwin described the tension inherent in the raised larynx posture:

> When, in phonation, [the larynx] is raised, and the thyroid is tilted in the direction of the position used in swallowing, many abnormal patterns of muscular contraction take place. Certain muscles, which in normal phonation need only make movements of fixation, anchoring with their antagonists any of the laryngeal structures, must now make strong contractions. Other muscles, normally not employed, must be brought into play to operate the displaced larynx. The whole activity is productive of localized tension.

Estill has seen evidence that there is a tilting of the cricoid cartilage in belting which is thought to relieve some of the pressure on the system and reduce tension (Estill, 1996: 243).

Although the shape of the vocal tract in classical singing has been documented in anatomical terms, the singing pedagogue usually doesn't tell the student to expand the piriform sinuses or lower the larynx. Instead, they may use imagery. For example, a singing teacher may say " pretend there is a balloon in your throat". Alternatively, they may have the student perform a natural function like yawning and pay attention to how the vocal tract feels.

The historic Italian School of singing has long advocated *gola aperta* and also a relaxed tongue root. Estill's (1996) research suggested that gola aperta is partially created by a *compressed tongue*, which increases the space in the middle pharynx. She also noticed that Twang and Belt qualities have a high tongue, which is somewhat retracted and Twang qualities have a constricted pharynx (Estill, 1996: 243). Sundberg's research also indicates that belt quality had a constricted pharynx. (Sundberg, 1993: 308) Estill, on the other hand, indicates belt voice has a widened pharynx (Estill, 1996: 243). Perhaps both are possible if the vocal tract is flattened from back to front (retracted tongue) but widened from side to side (piriform sinuses, pharyngeal walls). Laver's (1980) terminology for the tongue body, tongue root and pharyngeal wall movement and the auditory correlates may be useful in the discussion of the pharyngeal parameters of singing qualities. The phonetic term pharyngealized is used in association with tongue body movement. A constriction in the middle pharynx that "is achieved by retraction of the body of the tongue into the pharynx, rather than by the spincteric action of the muscles of the pharynx bringing the back wall of the pharynx forward" (Laver, 1980: 46) is called pharyngealized. Laver also uses the terms advanced tongue root and retracted tongue root to describe tongue root movement. Akan, a language spoken in Ghana, is used as an example. There are two sets of spoken vowels in Akan that differ mainly in the size of the pharynx. The description of these vowels uses the term Advance Tongue Root (ATR). In one set of vowels the tongue is drawn up and forward therefore increasing the size of the pharynx. In addition, the larynx is lowered (+ATR). In the other set, the tongue is bunched up and pushed back, narrowing the pharynx. This may be accompanied by a raised larynx (-ATR) (Ladefoged, 1993: 226). The description of +ATR in speech seems similar to that of compressed tongue in singing except the compressed tongue doesn't appear to move as far forward. The retracted tongue seems to correspond to -ATR. However, in singing, the postures usually also involve the pharyngeal walls. Laver (1980) uses the same term pharyngealized to describe the pharyngeal wall movement caused by the constrictor muscles in the oropharynx. Estill (1988) terms this pharyngeal width, and doesn't necessarily state the contributing physiological factors. Another consideration is that the constriction of the pharynx involves other physiologically and acoustically linked phenomena, such as a "vertical shift of larynx position downwards, giving a lowered larynx component (and a somewhat breathy phonatory setting), a tendency to pull the velum downwards (because the velum and tongue are attached to each other by the palatoglossus muscle), giving some nasalization" (Laver, 1980: 46). An example of such a linkage may be found in non-classical qualities such as R&B and Country as mentioned previously. R&B quality sometimes uses a breathy voice which may be linked to the lowered larynx and expanded pharynx, whereas Country quality often uses a whispery voice which may be linked to the raised larynx and constricted pharynx.

Laver's latitudinal labial settings such as horizontal expansion and vertical constriction may correspond to the subject's "smiley" sensation in Pop and Legit voice qualities. Labial protrusion may correspond to the subject's sensation of "puckered lips" in Country and Jazz voice qualities. Many non-classical singing qualities involve a close jaw in order to keep the vowels as similar to speech as possible, whereas classical and legit singing qualities involve an open jaw at high pitches in order to provide maximum volume through formant tracking. The velopharyngeal term nasal is also useful in the description of singing qualities such as Country, which may be similar to Estill's (1988) Twang voice quality. Nasality is a complex subject because the auditory quality perceived may be caused by two mechanisms. Nasality may result from an open nasal port or from a constriction in the pharynx (McKinney, 1994: 134-135).

Sometimes the terms *tense* and *lax* are used to describe the firmness of the tissue in the vocal tract. Lax tissue damps high frequencies much better than tense tissue and therefore will affect voice quality. The terms tense and lax have sometimes been linked to raised and lowered larynx. For this thesis, tense and lax will be used independently of laryngeal position.

As can be seen from the above discussion, there are many phonetic terms that are useful in describing singing voice quality. Since singing terminology for supralaryngeal settings is so limited, phonetic terminology is used in the description of supralaryngeal settings and associated singing quality. For example, the auditory analysis, which was performed by a linguist, uses terms such as raised larynx. When phonetic terminology (laryngeal or supralaryngeal) was not adequate to represent the event or quality then, other terminology was used.

Chapter 3: Related Work

3.1 Introduction

A review of the existing literature on acoustic and laryngographic correlates for the vowel [i] revealed varying ideas about voice qualities for speech and singing. Some voice qualities have been the subject of extensive scientific studies; others have not. Chapter three outlines related work on acoustic correlates for speech neutral setting, and for the seven voice qualities for singing analysed in the thesis. This chapter also contains some research that links acoustic characteristics with physiological postures, and includes related laryngographic research for singing voice quality. Research results from the existing literature will be compared with the results of the exploratory research in Chapter four.

3.2 Acoustic Characteristics of Speech

The average fundamental frequency for women's speech is 235 Hz (Peterson and Barney, 1952) and the average formant values for the [i] vowel are first formant (F1) 310 Hz, second formant (F2) 2790 Hz, third formant (F3) 3310 Hz. The vowel is characterized by the large spread between the first and second formants (Peterson and Barney, 1952). There is little energy above 5000Hz. The speech source signal harmonic spectrum drops by 12 dB SPL/octave (Sundberg, 1977: 16). Breathy Voice and

Whispery Voice show aperiodic activity in the spectrogram between the partials. Increased subglottal pressure combined with increased transglottal pressure will cause increased amplitude. Speech uses a modal vocal fold adjustment and neutral laryngeal and pharyngeal space in most cases although the vowel [i] has a more +ATR than mid vowels.

3.3 Acoustic Characteristics of Singing

3.3.1 Classical Voice Quality

Classical voice quality is found in classical singing style. It is used in opera, operetta, and art song singing, and is the primary vocal quality taught in North American voice studios, conservatories and universities. For the soprano, it is characterized by a flutey timbre and warm ring. Classical quality in this thesis may correspond closely to Estill's (1988) Sob Quality. Estill (1988) also describes another classical voice quality she calls Opera Quality with "*squillo*". She has found an aryepiglottal constriction in the pharynx with the squillo technique; in contrast, Sob is not a pharyngealized quality.

One important contribution to the voice quality of female classical singers seems to be the ability to track the fundamental frequency (F_o), which is also called the first partial (P1) with the first formant (F1). In female voices, this appears to be the predominant way of increasing amplitude, since in many cases the female classical singer is singing above F1. This is called *formant tracking* or *formant tuning*. A female classical singer does this by opening her mouth and raising the larynx slightly to shorten the vocal tract for higher pitches (Cleveland, 1994; Titze, 1994: 231; Sundberg, 1987:

127). For pitches above D5 (587 Hz), the F1 tracks P1; however, for pitches below D5, F1 may occur between P1 and the second partial (P2) (Schutte, 1993: 147). F1 tracking P1 will increase the amplitude of P1 and may contribute to a purer and a darker timbre when compared to non-classical singing qualities. In male and low female voices, there may be less fundamental formant tracking. In male voices the singer's formant (approx. 3000 Hz) appears more prominent and assumes the role of increasing audibility (Sundberg, 1987: 123; Cleveland, 1994; Titze, 1994: 238-240). Estill's (1989) research indicates an aryepiglottal constriction in classical voice. Based on Sundberg's original research published in 1974, Titze (1994: 231) calculated the dimensions between the larynx and the epiglottis and stated that the singers' formant could be attributed to this secondary tube within the vocal tract

Another contributing factor in classical voice quality is the slightly lower larynx. Generally it is lower than in many other singing qualities. Classical singers lower the larynx in order to lower formant frequencies and aid in the production of the singer's formant at approximately 3000 Hz (Titze, 1994). In male singers, the larynx may also lower with ascending pitch; this is called "*covering*" (Sundberg, 1987: 98). With female singers the larynx may rise to a more neutral position with ascending pitch, for the formant to track the fundamental frequency (Sundberg, 1987: 129).

There also seems to be extra space created throughout the vocal tract in this quality. Sundberg (1993: 308) noted separated sidewalls, wide piriform sinuses, low larynx, visible vocal folds, separated ventricular folds, and a wide laryngeal tube. This spacious vocal tract enhances the lower frequency partials. The expanded and spacious vocal tract could also contribute to the richer, purer quality in classical voice.

Narrow band spectrograms show clear harmonic striations with little energy between each partial (Miller, 1996: 278). The clarity of the harmonic striations indicates the absence of inharmonic activity in the vowels and may also contribute to the purity of tone. In classical voice purity of tone is paramount.

Female classical singers will modify the vowels by lowering the jaw and tongue position toward the neutral schwa vowel as pitch increases. At very high pitches all vowels converge to neutral schwa. The movement of vowels towards schwa is a result of formant tracking. To track the fundamental and increase the amplitude of the fundamental, the jaw must drop to shorten the vocal tract and vowels are lost in the process. This is especially true for close vowels (Sundberg, 1987: 126) and female singers.

Male classical singers also try to find the best tone by balancing the light and dark timbres in their voices (*chiaroscuro*). Researchers have quantified what chiaroscura means in terms of acoustics. A balanced tone occurs when the energy in F1 is matched by the combined energy of the second formant (F2) and the singer's formant (SF). If F1 lacks strength, then the timbre will be too bright, and if the upper formants lack strength, then the sound will be too dark (Miller, 1996: 278). It is still unclear how many sopranos uses primarily fundamental tracking, and how many use a combination of fundamental tracking balanced with increased activity in F2 and singer's formant. The balanced tone has been quantified for baritones. In Bounous's (1997) study of baritone voices, the fundamental needed to be no more than 17.0% stronger than the average of all other partials for good tone. He also found that the singer's formant needed to be at least

85% of the vowel formant. In classical singing quality, partials above the fifth formant F5 show little energy. This would also be consistent with this quality's emphasis on purity of tone.

The vocal tract isn't the only factor in voice quality. The glottal source also plays a part. Female classical singers use a thinner vocal fold adjustment. The new term proposed was "testo voice", or alternatively, in Laver's (1980) terminology, a subtype of modal voice. In contrast, most speakers use a thicker vocal fold or modal adjustment. Sometimes very low female voices, and mature dramatic singers will use more modal adjustment in for their singing quality, but generally there is a predominance of testo adjustment. Thinner vocal folds mean less time spent in the closed phase, and less subglottal pressure needed for vibration, which increases the purity of the timbre by decreasing the amount of high frequency partials. It also allows the female singer to produce higher frequency tones with ease (Titze, 1995: 41). Singers constantly vary the thickness of the vocal folds as the pitch goes up and down. They increase the thickness on lower notes, and decrease the thickness on higher notes. However, the general tendency is towards a thinner vocal fold adjustment for classical singing, especially in lieder and art song.

3.3.2 Belt Voice Quality

Belt voice quality has a full rich, bright and brassy sound. Belt voice quality is found primarily in American music theatre and pop music; however, it is also heard in ethnic folk music around the world. Ethel Merman made Belt voice quality famous.

Formant tracking occurs in belt voice as well as classical, but the tracking may not involve the fundamental frequency. Many researchers have found the amplitude of the second partial higher (Bevan, 1989: 77; Sundberg, 1993: 303; Schutte and Miller, 1993). Schutte and Miller (Schutte et al., 1993: 146) concluded that the higher-lying first formant was tracking the second partial. Bounous (1997) noted that the F1 tracking P2 is usually what happens in baritone singing voices.

There are conflicting reports on the singer's formant for belt quality. Sundberg (1993: 303) found energy in the singer's formant area greater in classical than in belt. Jo Estill found the opposite. Estill's (1984) research suggested belting has the greatest energy in the singer's formant range through the singer's entire range. Bounous found that the singer's formant in belt was 74% of the vowel formants. (Bounous 1997: 63) Estill's research indicated an aryepiglottal constriction, as in classical voice, which might contribute to increased energy in the singer's formant. (Estill, 1989)

The larynx may be higher in belt quality than in classical quality. This would generally increase all the formant frequencies including the singer's formant (Schutte, 1993: 146). Sundberg agrees that formant frequencies in belting are higher than in classical (Sundberg, 1993: 307). Estill (1988) notes that there is more activity in the thyrohyoid and sternothyroid muscles. She suggests that they may have an important role in belt in the stabilization of the larynx.

Belt voice quality seems to be generally less spacious than the classical voice quality. The higher larynx reduces the pharyngeal diameters, and other postures also constrict the vocal tract. Lawrence (1979) describes the vocal tract as having closed ventricular spaces, and constricted pharyngeal diameters including the vallecula on either side, epiglottis often tilted slightly over the larynx (aryepiglottal constriction), and an elevated tongue base. Sundberg (1993: 308) noted constricted sidewalls, narrow piriform sinuses, adducted ventricular folds, constricted laryngeal tube, compressed esophagus entrance, and narrow aditus. This posture could closely relate to the pharyngealized setting in phonetics. However, Estill (1996: 243) has observed an expanded pharynx. EMG studies suggest a lower velum, and relaxed tongue (Estill, 1988). The relaxed tongue, similar to –ATR without the bunching, may contribute to the pharyngeal constriction. Constriction in the vocal tract usually is associated with enhanced higher frequency partials. This adds a brighter and buzzier quality to the tone. If one looks at the acoustics of orchestral instruments, then a flute has a very pure light timbre. It is characterized by a strong second partial and very little upper partial activity. An oboe, on the other hand, has a buzzier timbre and is characterized by a weak fundamental and second partial, and much more strength in the upper partials. The acoustic difference between classical soprano and belter seems similar to the comparison of flute with oboe.

Schutte and Miller (1993: 163) describe less vowel modification than with classical singing to enhance clarity of diction. Sullivan (1985: 75-86) on the other hand, describes in detail vowel modifications need for healthy belt singing. There seems to be some agreement that vowel modification is necessary, but that it's not as severe as the classical quality. Schutte and Miller (1993: 147) noted that D5 was the upper limit for female belters. This coincides with the pitch Estill has said requires increased effort.

Estill (1988: 39) found "the folds remained closed more or less than 70% of every cycle across the two-octave range" examined, which is much higher than in classical singing in both modal (chest) and falsetto (head). This is consistent with both Bevan and

Sundberg. However, other research by Estill (1990) suggests that the vocal folds thin out in belt quality as the singer ascends in pitch. The high closed quotient has the effect of increasing the high frequencies and reducing low frequencies, especially the fundamental (Collyer, 1997). The closed quotient results have been associated with thicker vocal folds and increased subglottal pressure. Computer modeling suggests that acoustical power of vocal fold vibration peaks at a closed quotient of between 40% and 50%. Acoustical power diminishes as the closed quotient increases. This suggested belt singing may be less efficient than classical and therefore may not appear as loud as classical (Titze, 1994: 227). Estill (1990: 171) disagrees and states that belting glottal efficiency rates were higher in belt than in speech or opera. It has been noted pedagogically that belt may not be quite as loud as Opera. Estill however, theorized that belting might be perceived "louder" because there is more energy in the ear's most sensitive frequency range, at around 3000 Hz (Estill, 1988: 39).

Finally, Estill (1988) noted the cricoid cartilage is tilted. Pedagogues call this "laryngeal lean". There are presently no theories on spectral correlates for this posture. Physiologically, Estill's (1988) EMG studies indicate, "the infrahyoids are functioning to stabilize the thyroid cartilage in a vertical position to allow the cricoid cartilage to tilt dorsocranially". She theorizes that this may reduce the vocal fold pressure. Titze (1994) correlates pressed phonation to the introduction of high frequency partials; therefore, this posture may decrease some high frequency partials. However, Bevan (1989) found greater amplitudes of partials between 6-8 kHz when comparing female chest voice to belt. Perhaps high frequency partials are created at the source by something other than an increase in adductory forces.

3.3.3 Pop Voice Quality

Pop voice quality is found in Pop music, American Musical Theatre and in children's musical movies (e.g. Disney). Female pop quality is generally used for the ingénue characters. It is natural and bright in timbre and has a lighter quality than Belt. It can also be found in folk music. It may equate closely to Estill's Speech quality.

Schutte and Miller (1993) were the only team to include Pop as a voice quality in their studies. They found that the main difference between Pop and Belt was what they termed the "falsetto" vocal fold adjustment. This may equate to my testo setting. The thinner vocal fold adjustment in Pop quality may produce a less rich spectrum with less energy in the higher partials. They noted that the larynx was raised in Pop and the result was higher formant frequencies. They also noted that the F1 was not tracking the fundamental. They estimated that F1 was just below the second partial (P2), and F2 was just above P2. Both F1 and F2 supported the second partial. This translated to more energy in the second partial compared to the classical voice production, which showed more energy in the fundamental. They also showed that there was more energy in F3, F4, and F5 than in classical voice quality.

3.3.4 Legit Voice Quality

Legit voice quality is similar to classical voice quality; however, it is more speech-like in the lower register. It is found in musical theatre. For example, many musical movies use this voice quality: Sound of Music, Kiss me Kate, and Oklahoma. Schutte and Miller (1993) also included legit in their study. They found legit to also use a "falsetto" vocal fold adjustment and a high larynx. This laryngeal setting again may equate to my testo voice. This type of vocal fold adjustment should decrease upper partials and the high larynx should raise formants. They found that F1 tracked below the second partial, similar to classical voice and theorized that, in the upper register, F1 would track the fundamental.

3.3.5 Country Voice Quality

Country voice quality is a forward nasal twangy quality common in American country music style. Jo Estill refers to this quality as *Twang*. Estill (Yanagisawa et al., 1991) found the larynx to be very high, moving substantially from a neutral position on low notes to a very high position on high notes. The pharyngeal walls were constricted and the velum was lowered and slightly open, as compared to opera, in which the velum on vowel sounds was closed and arched. The high larynx raises formants and the vocal tract constriction enhances higher partials. The slightly open velum added nasality to the sound. Cleveland (1999) found that in male singers there was no consistent trend toward raised or lowered larynx. Cleveland (1995) also studied the lung capacity used in Country singing and found it to be minimal. Most performers used *costal breathing*. This means that the descending rib cage and the associated intercostal muscles control the airflow somewhat like a bellows. In contrast, classical singers usually use *diaphragmatic breathing*. In this method, the diaphragm and the abdominal muscles work together to control the airflow from below like a piston. Diaphragmatic breathing exerts a downward pull on the trachea and larynx, which is advantageous in qualities that require a lower larynx. Sundberg theorized that costal breathing would provide minimal tracheal pull. Since Country quality demands a neutral or high larynx, low diaphragmatic breathing may not be advantageous. Sundberg suggested that the high larynx and the costal breathing often are found together (Sundberg, 1999, personal correspondence). Cleveland (1999) found that male country singers use similar formant placements for speech and singing. He found that if formant frequencies changed, then they were higher rather than lower than speech.

3.3.6 Rhythm and Blues (R&B) Voice Quality

Rhythm and Blues voice quality is a dark and soulful quality with warmth and character. A literature search did not find any research pertaining to this quality.

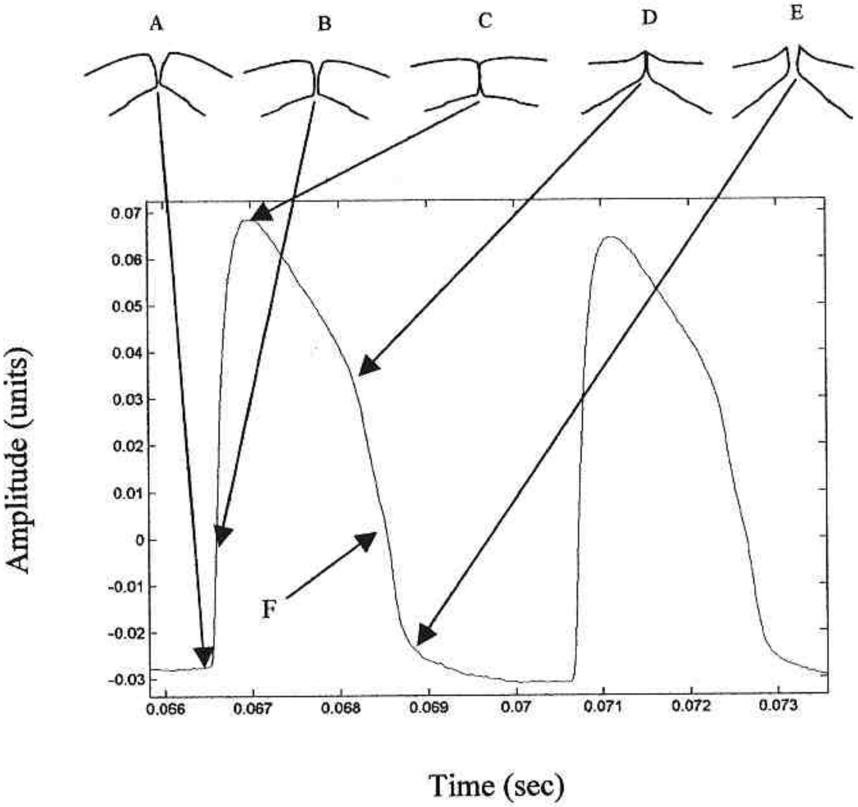
3.3.7 Jazz Voice Quality

The voice qualities used in vocal jazz musical style are often varied and colourful. Two distinctive jazz qualities were analysed in this thesis. The first was a darker quality similar to R&B. The second was a bright quality similar to Belt. The darker quality, which the subject refers to as "Smokey Jazz", was used in the acoustic analyses. Laryngographic samples for both dark and bright qualities were analysed. A literature search did not find any research pertaining to these qualities.

3.4 LGG Characteristics of Speech

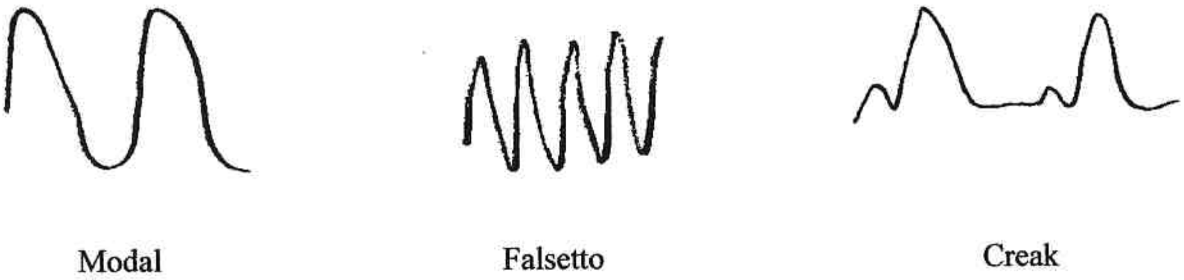
Although the Lx signal only shows the conductance of an electrical current through the vocal folds, researchers have learned that various conductance patterns seem to correspond with various vocal fold activities. Considerable research has been done on linking the important features of the Lx signal to vocal fold function. First, it is generally agreed that the Lx signal is not influenced by the vowel (Fabre, 1958; Valancien & Faulhaber, 1967; Orlikoff, 1995). Second, there is a general consensus among researchers that an Lx signal can be linked to vocal fold contact area. Fabre (1958) stated that the laryngograph primarily shows the progression of the vocal fold contact during each glottal closure. Lecluse et al. (1975) did simultaneous laryngograph and synchronstroboscopy linking the physical vocal fold cycle to the Lx wave. The view obtained by the stroboscope does provide some detail, but doesn't show the complexity of the vocal fold movement. Researchers have now established that the vocal folds roll against each other from the bottom margins to the top margins in the vertical plane, and may close like a zipper in the horizontal plane. The work of Rothenberg (1981) and MacCurtain and Fourcin (1982) helped clarify the link between the Lx signal and the corresponding vocal fold movement. Figure 3 shows the Lx wave and the corresponding coronal view of the vocal folds with the important transition points marked.

Figure 3: Vocal fold movement and corresponding points on the Lx waveform. The Lx waveform is a sample from the subject of this investigation. The diagrams of vocal fold movement were extrapolated from what is currently known based on Rothenberg (1981) and MacCurtain and Fourcin (1982).



Third, the vocal register and phonation type can be recognized from the Lx signal as seen in Figure 4 (Bakan and Orlikoff, 2000).

Figure 4: Three phonation types and the corresponding Lx waveform based on Baken and Orlikoff (2000).



Modal Falsetto Creak

Esling (1984) correlates linguistic phonation types, the associated Lx waveform and RT/FT ratios (where RT = rise time and FT = fall time). In Figure 3 the duration in seconds between points A – C would equal rise time and the duration in seconds between C – E would equal fall time. Esling (1984) measured between 10 % and 90 % of both RT and FT for the RT/FT ratio. The different phonation types all have a distinctive Lx waveform and associated RT/FT ratios. Esling's (1984) research on laryngographic correlates describes seven phonation types: Creaky Voice (CV), Modal Voice (MV), Ventricular Voice (VV), Harsh Voice (HV), Whispery Voice (WV), Breathy Voice (BV), and Falsetto (F).

Titze (1990) describes the basic mechanism for modal phonation and how it relates to the Lx waveform. When the vocal folds are at rest and closed, ready for phonation, then the upper medial edges are assumed to be in contact and the lower edges are convergent. Phonation begins as the air from the lungs is pushed against the vocal folds causing the upper medial edges to separate. As the air moves through the opening, the pressure drops (Bernoulli effect) and the vocal folds can no longer be held open. They begin to move toward the midline at the lower medial edges first. The vocal folds then collide, roll past and deform against each other to the upper medial edges. As the glottis closes, the airflow is cut off, but the wave front above the vocal folds keeps moving, creating a suction their wake that helps close the glottis. The vocal folds return to the convergent shape and the cycle repeats. This movement creates the idealized Lx shape Figure 5. It is not possible to tell from the Lx signal exactly when the vocal folds start to move toward each other, or away from each other.

Figure 5: An idealized Lx waveform and the associated vocal fold behaviour.

Idealized Lx signal

Vocal fold behaviour

The start of the contacting cycle may be the lower margins of the vocal folds making initial contact, or may be a mucus bridge forming between the folds (Fourcin, 1981). In the horizontal plane the Lx signals with a lesser slope at the beginning of the contacting phase are thought to be zippering shut, whereas a consistent slope from zero to maximum is thought to show a more parallel closure. In the vertical plane, the Lx signal with a steep slope suggests parallel closure, whereas a lesser slope suggests a rolling closure or "ramping" as seen in Figure 8. The peaks of the Lx wave signify maximal contact; however, the vocal folds may not actually achieve full closure. The significance of the "knee" found on the descending slope, such as point D on Figure 3, is not definitively known. There is evidence that the knee of the Lx signal on the descending slope may be the mucus link severing. Others argue that it may be due to the "upward excursion of the vocal lip just before opening" (Baken and Orlikoff, 2000: 420). Titze (1990) has also theorized that surface bulging may produce what he terms skirt elevation as seen in Figure 9. The Lx amplitude is not related to vocal intensity. However, there is evidence that the ascending slope may be related to vocal intensity and the adductory presetting of the vocal folds (Titze, 1988; Hacki, 1996). The F_0 of the Lx is an accurate measure of the vocal F_0. In the modal voice the contact quotient (contact duration/one Lx wave period) is between 40% and 60%. Below 40% would indicate hypoadduction and above 60% would indicate hyperadduction.

Titze (1990) indicates that peak widening suggests vocal fold adduction as seen in Figure 6. When the vocal folds are firmly adducted, such as in Harsh Voice, more "opening" must be done. The vocal folds will remain in contact much longer than for normal adjustment. The Lx signal usually shows a greater contacting time. The peaks are visually much wider than for normal adjustment phonation as seen in Figure 6.

Figure 6: Peak widening/Vocal fold adduction

Lx signal

Vocal fold behaviour

The angle of convergence of the vocal folds can be greater in some instances. The vocal folds will contact like the normal adjustment; however, they will take more time to decontact. This skews the waveform as seen in Figure 7.

Figure 7: Peak Skewing/Glottal Convergence

Lx signal

Vocal fold behaviour

The convergence of the vocal folds and the vocal fold thickness play a part in the phasing or time lag between the contacting of the upper medial edges and the lower medial edges. If the angle of convergence is greater, then the time lag will greater between the upper and lower edges as they go through each contact cycle, and the Lx waveform will be more triangular as seen in Figure 8.

Figure 8: Triangularity/Increased Vertical Phasing or Ramping

Lx signal

Vocal fold behaviour

The contact area will increase slowly and decrease slowly. If the angle of convergence is less, then the Lx signal will be more rectangular. The contact area will increase quickly remain contacted longer and decrease quickly. As the vocal folds become thinner, the lag time is decreased and the contact area increases quickly and decreases quickly creating a peaky Lx signal.

A bulging of the surface of the vocal folds causes the contact area to grow quickly in the area of the bulge and then less quickly after the bulge. The reverse is true for the decontacting phase as seen in Figure 9.

Figure 9: Skirt elevation/Surface Bulging

Lx signal

Vocal fold behaviour

The linkages Titze (1990) established between the Lx waveform and the associated vocal fold activity should make it possible for researchers to perform a visual analysis of the Lx waveform and gain useful information about vocal fold function.

Various researchers have quantified these observations with ratios in order to better compare Lx samples. Researchers use points along the Lx signal's contact wave. Ratios used in the past for comparison were as follows:

Table 4: Summary of the various types of laryngographic measurement, associated researchers and relevance to vocal fold function.

Contact-based Measures	**Ratio**	**Discussion**
Relative contact duration	$\%CQ = \frac{CP}{V}$ X 100 Where: CQ = contact quotient CP = contact phase V= vibratory cycle (Rothenberg et al., 1988; Scherer et al., 1988; Scherer, Gould et al. 1988; Orlikoff, 1991; Houben et al., 1992)	Contact Phase = the length of time from zero crossing on ascending slope to zero crossing on descending slope Vibratory Cycle = The length of time for one Lx wavelength. - CQ for speech usually varies from 40% - 60% - Vocal intensity increase is proprotional to CQ increase - Adduction increase is proportional to CQ increase - Pitch increase (women) is proportional to CQ increase
Relative contact rise time	Slope = Rise/Run between 25% and 75% of peak (Orlikoff, 1991)	Rise: amplitude of impedance at .75 of peak – amplitude of impedance at .25 peak Run: time at .75 of peak – time at .25 of peak
	Slope = Rise/Run between 10% and 90% of peak (Fisher, et al., 1992)	Rise: amplitude of impedance at .90 of peak – amplitude of impedance at .10 peak Run: time at .90 of peak – time at .10 of peak
	Increasing contact interval relative to vibratory cycle (Houben et al., 1992; Orlikoff et al., 1997: Weschler, 1977)	- high amplitude is proportional to steep slope - zippering is suggested if the initial slope is greater than the mid slope on the ascending portion of the Lx waveform
Contact symmetry	RT/FT (Esling, 1984)	RT = time at 90% of peak – time at 10% of peak on ascending slope of Lx waveform FT = time at 10% of peak – time at 90% of peak on the descending slope Lx waveform

		RT/FT was used to describe phonation types.
	$SQ = \frac{DT}{CT} \times 100$ Where: SQ = Speed quotient DT = Decontacting time CT = Contacting time (Dickson et al. 1992)	Speed quotient is calculated using zero crossings and peak time. DT is calculated by subtracting the peak time from the final zero crossing time and CT is calculated by subtracting the initial zero crossing time from the peak time. For example, the points used to calculate SQ in Figure 3 would would be as follows: B = initial zero crossing C = peak F = final zero crossing
	Contact Index = contacting – decontacting /full contact phase (Orlikoff, 1991)	0 represents a symmetrical waveform. –1 represents a short contacting phase +1 represents a short decontacting phase
		Finding the most reliable and clinically relevant measure of contact rise will require further investigation. - vocal intensity increase is proportional to slope increase - upper partial increase is proportional to slope increase The degree of contact symmetry is thought to reflect vocal fold tonus and is particularly sensitive to vertical mucosal dynamics.

3.5 LGG Characteristics of Singing

Little research is available on the voice source for various qualities of singing. Sullivan (1989) after observing photography of the vocal folds of belters, states that the vocal folds come together firmly and cleanly. Estill (1990) was the only researcher found who looked at LGG and EMG of the voice source in various singing voice qualities. Estill (1988: 39) stated that in Belt voice quality the LGG analysis showed that "the folds remain closed more or less than 70% of every cycle across the two octave range". Estill (1990) used open quotient ratios instead of contact quotient ratios. The open quotient is closely related to the contact quotient, but it uses the decontacting phase instead of the contacting phase in the ratio. The open quotient describes the time that the vocal folds are opening, or decontacting, in relationship to the vibratory cycle.

Since previous research was limited, I chose to use Estill's samples (1990) and analyse them by hand using the Contact Quotient and Titze's (1990) visual criteria for the Lx waveform. These are two of the LGG analysis methods used in the thesis for the exploratory research. Although the thesis did not examine Falsetto voice quality, both Belt voice quality and Speech voice quality were examined. Tables 5-9 show the results of my analysis of Estill's samples.

Table 5: Contact Quotient for Belt, Falsetto and Speech using Estill's (1990) Lx waveform samples.

Pitch	**Belt**	**Falsetto**	**Speech**
196 Hz	70%	30%	50%
294 Hz	60%	40%	60%
382 Hz	60%	30%	50%
587 Hz	50%	20%	50%

Table 6: LGG visual characteristics for Belt, Falsetto and Speech voice quality for a spoken and sung vowel [i] at a frequency of 196 Hz

Characteristic	**Belt**	**Falsetto**	**Speech**
Peak widening	Wide	Narrow peak/ wide base	Average
Peak skewing	Skewed	Not skewed	Skewed
Triangularity	Rectangular	Triangular	Skewed
Skirt Elevation	Extreme	None	Moderate

Table 7: LGG visual characteristics for Belt, Falsetto and Speech voice quality for a spoken and sung vowel [i] at a frequency of 294 Hz

Characteristic	**Belt**	**Falsetto**	**Speech**
Peak widening	Moderate	Narrow peak/ wide base	Moderate
Peak skewing	Skewed	Not skewed	Skewed
Triangularity	Skewed	Triangular	Skewed
Skirt Elevation	Moderate	None	Moderate

Table 8: LGG visual characteristics for Belt, Falsetto and Speech voice quality for a spoken and sung vowel [i] at a frequency of 392 Hz

Characteristic	**Belt**	**Falsetto**	**Speech**
Peak widening	Narrow	Narrow peak	Average
Peak skewing	Slight Skew	Not skewed	Skewed
Triangularity	Slight Triangularity	Triangular	Skewed
Skirt Elevation	Slight	None	Moderate

Table 9: LGG visual characteristics for Belt, Falsetto and Speech voice quality for a spoken and sung vowel [i] at a frequency of 587 Hz

Characteristic	**Belt**	**Falsetto**	**Speech**
Peak widening	Narrow	Narrow peak/ wide base	Narrow
Peak skewing	Slight Skew	Not skewed	Not Skewed
Triangularity	Slight Triangularity	Triangular	Triangular
Skirt Elevation	Slight	None	None

Belt quality has the largest contact quotients and falsetto the smallest for all pitches (Table 5). Belt, Falsetto and Speech have smaller contact quotients in the upper register (Table 5). A high contact quotient (Table 5) for the lower register and peak widening (Tables 6 and 7), suggest Belt quality uses thicker vocal folds; however, as the pitch ascends there is a decrease in the contact quotient (Table 5) and narrowing of the peak (Tables 8 and 9). A smaller contact quotient suggests thinner vocal folds. In a previous study between Opera, Belt and Speech voice qualities, Estill (1988) had found that Opera showed the smallest contact quotient, and Belt showed the largest contact quotient. Estill (1990) equates glottal efficiency to open quotients. She says, "larger open quotients demonstrates a lower glottal efficiency" (Estill, 1990: 172). Conversely then, does this suggest that large contact quotients have increased glottal efficiency? Estill (1990) states that the Belt voice quality had a high glottal efficiency and Twang voice quality was shown to be the most efficient. Estill also tested airflow rates and found as airflow decreased the sound pressure level (SLP) and efficiency increased.

Belt voice quality also shows a rectangular shaped Lx waveform at low pitch (Table 6), which suggests an increase in vertical phasing and a smaller angle of convergence of the vocal folds. The rectangular Lx shape also suggests square-edged vocal folds in coronal cross-section. As the pitch ascends (Tables 7-9) the shape of the Lx waveform becomes more triangular in all voice qualities, which suggests a larger angle of convergence and a decrease in vertical phasing. A triangular-shaped Lx waveform suggests triangular-shaped vocal folds in coronal cross-section.

The Belt Lx waveform was skewed at low pitches and symmetrical at higher pitches, which suggests a change in angle of convergence. Estill (Estill et al., 1988: 239) refers to vocal fold plane as one of the source voice quality parameters. Belt voice quality also shows more skirt elevation than the corresponding pitch for speech (Tables 6-9). Skirt elevation suggests surface bulging of the vocal folds. Surface bulging can suggest increased TA activity. Estill (1988) also tested Belt and Classical with EMG. These EMG tests of the vocalis (TA) support the LGG results. The vocalis was most active in Belt, moderately active in Speech and least active in Classical. However, Classical voice quality in the high register showed more TA activity than the corresponding pitch in Speech.

Chapter 4: Examination of Seven Singing Voice Qualities and Speech

4.1 Introduction

When this thesis was conceived, there was very little published research on non-classical singing available. Many of the teaching methods had differing views on voice production for non-classical singing voice qualities. Often claims about voice production for non-classical voice qualities were not backed up by scientific studies. The object of this research was to improve my skills as a voice teacher and singer by gaining a better understanding of the acoustics and the associated voice production of common singing voice qualities. However, many more questions than answers emerged from this thesis. Generating questions became an important part of the thesis.

This chapter focuses on the observation and analysis of seven singing voice qualities and speech. The singing voice qualities used in the analyses were Belt, Classical, Country, Jazz, Legit, Pop, R&B and Speech. Each of these seven singing voice qualities was described briefly in Related Works (Chapter three). Voice samples in all seven singing qualities were taken from a professional soprano (Lisa Popeil). The subject was chosen for her ability of producing these qualities consistently and effectively. She has developed a non-classical voice training program, marketed under the name VOICEWORKS, that teaches these qualities through face shapes, and laryngeal positioning. The voice qualities used in this thesis form the basis of Ms. Popeil's teaching. Spectral measurements were taken from the voice samples using standard acoustic analytical techniques such as DFT, LPC and Spectrogram. These analyses are used extensively in speech science. They were chosen because the results could be easily compared with existing research and in addition, could provide clues about associated physiological events and postures. LGG was used to observe the vocal fold contact area (VFCA). VFCA has been linked to vocal fold function. Analytical techniques are discussed in detail in Methods and Materials (Chapter four:4.2).

The exploratory research in chapter four focuses on acoustic and laryngographic observations of seven distinct singing qualities with the hope of clarifying non-classical vocal techniques used by singers. The results of this research are compared to existing research. The comparison uses linguistic terminology to describe the voice qualities and suggest associated voice production. Finally, the observations in chapter four lead to the development of hypotheses that can be tested in further research.

4.2 Methods and Materials

4.2.1 Equipment

Spectrogram analysis: (a) Norgate 486, DOS based computer with Panasonic monitor, (b) Kay Elemetrics Computer Speech Laboratory (CSL) software and hardware Model 4300 (c) TASCAM DAT recorder (d) Hewlett Packard Laserjet Printer.

FFT and LPC analysis: (a) Power Macintosh G3, (b) MATLAB Student version 5.0, (c) Epson stylus color 740 printer, (d) Iomega JAZ drive 1 GB.

LGG CSL analysis: (a) Norgate 486, DOS based computer with Panasonic monitor (b) Kay Elemetrics CSL LGG processor software (c) TASCAM DAT recorder.

LGG MATLAB analysis: (a) Soundforge XP 4.5 (b) Sony VIAO laptop computer (c) MATLAB 6.1 software (d) Epson Stylus Color 777 printer

4.2.2 Sample Collection

The original data were obtained from the subject on July 9, 1998 in the Phonetics Laboratory at the Linguistics Department at the University of Victoria. All samples were recorded on DAT tape in a sound-treated room with the microphone approximately 15 centimeters from the subject's mouth. LGG samples were simultaneously taken by the laryngograph and recorded on a second track of the DAT tape. Target pitch was supplied by a tuning fork. Speech samples of one utterance of ten vowels were taken first (heed, had, hoed, hayed, hod, hawed, head, hid, who'd, hood) followed by sung samples of the ten steady state vowels using three pitches (B3, E4, and B4) in each of the seven qualities (Pop, Belt, Classical, Country, R&B, Jazz and Legit). Vowel samples were also taken within the context of songs and are referred to below as "running samples". A subset of voice samples on the vowel [i] was selected for analysis.

File conversion and transfer were the main challenges in the process of sample collection. Each machine had sampling frequency settings and internal filters. For example, the Tascam DAT recorder sampled at 44,000 samples per second (s/s) for audio to digital conversion and had an internal anti-aliasing filter. The output on the Tascam recorder converted the signal back to analog format; therefore it had to be resampled again into digital format by CSL 4300 hardware and saved as computer files. The sampling rate of the CSL 4300 was 50,000 s/s. The CSL 4300 machine also was equipped with an anti-aliasing filter, but it had a digital output.

Another complexity in the sampling process was that the voice samples were collected in the summer of 1998 and the voice samples were analysed over three summers (1998, 1999, 2002). The significance of the time delay was that the equipment changed midway through the project. The project was started using CSL for the spectrograms. The initial acoustic LPC and DFT analyses were done on a Mac G3 computer using MATLAB student edition. Minitab Student edition was also used for graphing and statistical analyses. One limitation of the MATLAB student edition was that the sample size accepted by the program for the DFT analysis was very small. After the summer of 2000, the remaining analyses were done using a Sony VIAO PCG-F430 computer and MATLAB software version 6.1. Although the newer hardware and software had increased capabilities, there were new challenges introduced for file conversion and file transfer.

Different digital equipment also used different file types. The voice samples transferred as digital output from the CSL 4300 to the CSL software were saved as NSP files. MATLAB, on the other hand, was not compatible with NSP files, so they were converted to WAV files to be analysed in MATLAB software. Soundforge XP 4.5 was used to convert the NSP files to WAV files; however, a 50,000 s/s sampling rate was not an option in Soundforge. Instead, a 48,000 s/s sampling rate was chosen.

Another important aspect of sample collection was the transduction of sound pressure level (SPL) to amplitude readings by the computer. The CSL 4300 hardware assigned values for the amplitude between –32,000 and +32,000. This is called *quantizing*. The spectrogram analyses used the original samples and the amplitude values assigned by the computer quantizing process. However, for DFT and LPC analysis, low amplitude voice samples were boosted to the average quantized amplitude of 32,000 and trimmed to isolate a representative steady state portion of the vowel. Steady state samples have constant, or close to constant, pitch for the duration of the sample. Steady state samples are analysed more easily by algorithms than fluctuating signals. It was hoped that having all samples as steady state and with similar amplitude would make the graphical representations of the harmonic content easier to see and compare. Trimming and boosting levels of samples is a common practice in speech science.

4.2.3 Analytical Techniques

The following sections contain information about each analytical technique, choices that were made during the research, and questions that arose from those choices. These choices have a direct relationship to the accuracy of the results of this research. Analyses fell into four categories: acoustic, auditory, articulatory and laryngograph.

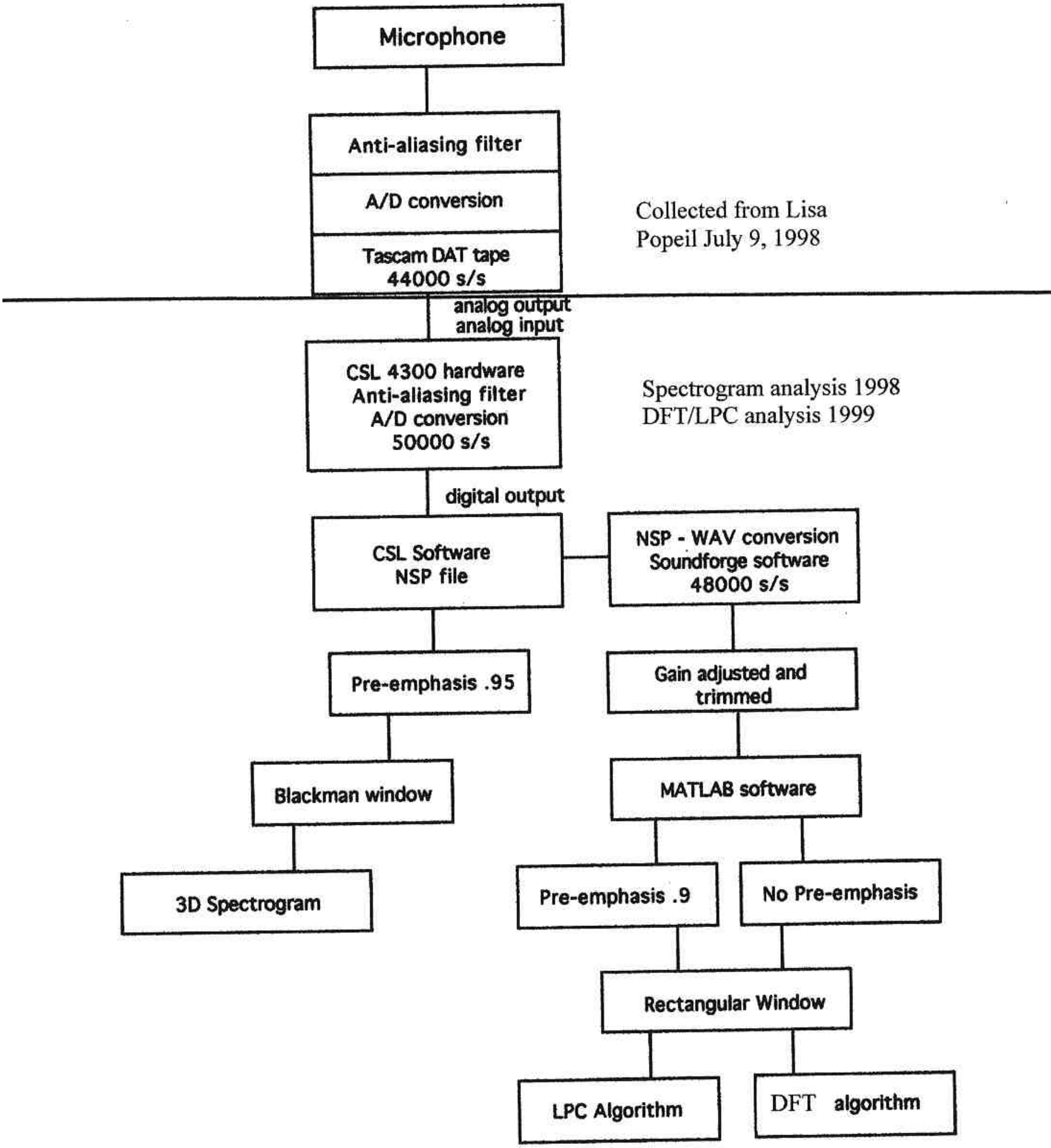
4.2.3.1 Acoustic Analysis

The acoustic analytical techniques used in this investigation were Linear Predictive Coding (LPC), Discrete Fourier Transform (DFT) and Spectrograms. These acoustic techniques are common in speech science; however, the settings for this research were slightly different. I chose to expand the traditional frequency range of the acoustic analyses from 0-5000 Hz, which is common in speech analyses, to 0-24,000 Hz for singing. The initial spectrogram analyses performed on CSL showed energy in some vowel samples up to 20,000 Hz. In order to view this energy in the DFT and LPC analyses, I chose to use MATLAB software. MATLAB provided more flexibility for settings than CSL. The signal path for the acoustic analyses is contained in Figure 11.

4.2.3.1.1 Linear Predictive Coding: algorithm and settings

Formant peak location can be very important to perceived voice quality, and LPC analysis can be a very effective method of extracting information from an

Figure 11: Acoustic Analysis Signal Path



Collected from Lisa
Popeil July 9, 1998

Spectrogram analysis 1998
DFT/LPC analysis 1999

acoustic signal. Linear prediction is an idea that dates back to Gauss in the late eighteenth century, but surprisingly, it was not generally used until 1949. LPC was first used for the analysis of speech in the late 1960's. The method of the algorithm is to predict the next point in a series of numbers with a minimum error. It originally was not designed to track formant location in a voice signal; however, there is a physical connection. The mathematical description of the vocal tract resonance, in matrix form, looks extremely close to the matrix set up in the solution of this linear prediction algorithm; therefore the solution to one mirrors the other. There are three parts to the algorithm: the first part calculates the next point in the series of numbers from the existing data; the result, a summing equation, is transformed via Z transform into an algebraic format which makes it easier to work with, and then the error is minimized through a process called linear least squares. Peaks will appear in the graph when the error is minimized, and these peaks correspond to resonances in the vocal tract (formants).

LPC works best on a finite slice of the signal. A process called *windowing* extracts a time slice of the signal. It is the equivalent to multiplying the signal by, for example, a rectangular pulse, which allows the signal y(n) to be viewed through a window of length "n". Mathematically this process is equivalent to convolving the transform of the signal with the transform of the pulse (Parsons, 1986: 29). This results in the zeroing, or canceling, of the head and tail of the signal and is called rectangular windowing. For example, a signal that contains the series of discrete numbers [5454334876394853942] would become a series of numbers [0000034876394800000] after rectangular windowing. This leaves a finite slice to be analyzed. However, windowing with a rectangular shaped pulse also creates some problems. Anytime a signal drops abruptly to zero the result will contain within it numbers that equate to energy at high frequencies.

The LPC algorithm not only uses a slice of the total signal, but it takes a few points at a time for the calculation. The number of points taken at a time is called the order number. For speech, when sampling at 10,000 samples per second, the suggested order number is 12. One method of calculating the order number is as follows:

- estimate the number of formants expected up to the Nyquist frequency (half the sampling rate)
- double this number, since two roots of the transform polynomial will resolve to form one formant peak
- add two for unresolved roots (DC, and radiation).

Using this method, the order number was calculated at 50 points. By running the LPC algorithm on sample B01 with the order numbers at 5-point intervals between 20-105 the choice was verified. An order number of 50 seemed to provide the best resolution (Appendix A).

In CSL, the window size had a preset limit of 1,024 points and the highest order number available was 20. Conforming to these parameters would have meant downsampling to a top sampling rate of 18,000 samples per second with a Nyquist frequency of 9000 Hz. The spectrograms showed energy in some vowels up to 20,000 Hz. To view this energy, I chose to use MATLAB software instead of CSL. MATLAB software provided complete flexibility of window size and order number.

Pre-emphasis is another common procedure in speech analysis. Since the speech signal drops off by approximately 6 dB per octave, a filter is applied that flattens the low frequencies and boosts the high frequencies. A pre-emphasis filter is applied before the signal is windowed and processed by the LPC algorithm. This type of filter increases the definition of upper formants.

In MATLAB, the pre-emphasis filter used was a "Direct Form II Transposed" implementation of the standard difference equation where b = coefficient (-.9 was used), a = gain (1 was used), x = signal vector, y = filtered signal vector, and n = point numbers:

$$a(1)*y(n) = b(1)*x(n) + b(2)*x(n-1) + \ldots + b(nb+1)*x(n-nb)$$
$$- a(2)*y(n-1) - \ldots - a(na+1)*y(n-na)$$

Pre-emphasis enhances the visibility of the peaks, but it also can change the bandwidth slightly depending on the shape of the window (Parson, 1986). As well, experimental pre-trials suggested that pre-emphasis moved peak locations. The LPC algorithm alone makes the formants appear "peakier" than the actual transfer function of the vocal tract. The CSL manual (Kay Elemetrics Corporation, 1991) states, " if pre-emphasis is not applied, it is likely that an error in the formant frequency calculation will result." This thesis examines two views of the LPC: one with pre-emphasis and one without pre-emphasis. Data was collected from both views.

Pre-emphasis may be more advantageous in speech research than in singing. For example, a common sampling rate for speech has been 10,000 samples per second with a Nyquist cutoff at 5,000 Hz. In these cases the .9 filter may provide a clearer picture of our perception of the sounds. Sound propagation through space may be one factor affecting perception. High frequencies are directional, whereas low frequencies tend to diffract. High frequencies can be heard at a distance much more easily than low frequencies. This is called the radiation effect (Fant, 1960). In fact, our ability to hear frequencies at a distance increases by 6 dB per octave. In addition, our human hearing system becomes gradually more acute as the frequency increases up to about 5,000 Hz. Human hearing intensifies the sounds in the range of 3,000–4,000 Hz and at approximately 9,000-12,000 Hz because these frequencies lie near the human ear canal resonances. Therefore, a smoothly curved filter that flattens the lower harmonics and intensifies the harmonics in the range of 2,000-5,000 Hz may be very close to how we perceive these voice sounds. In this singing analysis however, the Nyquist frequency was increased to 24,000 Hz; therefore a gradually increasing filter may not mirror our perception. Instead, a perceptual filter might have been a better alternative. This avenue, however, has not been explored in this thesis.

LPC analysis in MATLAB used an auto-regressive (AR) model and autocorrelation (solved using the Levinson function). The transfer function was calculated using z-transform digital filter frequency response and then plotted. The settings were as follows:

- samples were trimmed to contain only the steady state portion of the signal
- Gain adjusted in CSL to average quantized value of 32,000
- sampling rate 48,000 samples per second
- order number N=50
- rectangular window
- window size 2,048 points
- view #1 - no pre-emphasis
- view #2 - pre-emphasis (1 - .9)

The data that were collected from LPC analyses are in Appendix A and B.

There were a number of concerns about the accuracy of the LPC processing in this research. The risks of rectangular windowing were mentioned previously. Anytime the signal vector drops from a number to zero the algorithm equates this to the presence of high frequencies. For further research a Blackman window, which filters the signal with a rounded cosine"ish" shaped function could be used. The Blackman window reduces the energy around the edges of the slice, and the phantom high frequencies are reduced. In speech analysis, this process dramatically improves the readability of the graphical output. Windowing with a Blackman or other cosine"ish" window has also been recommended for autocorrelation method of solving the LPC. Autocorrelation minimizes the error over an all time (-infinity to + infinity) and therefore looks outside the window to do calculations. (Kay Elemetrics Corporation, 1991) A better method still would be pitch synchronous analysis, which makes sure that the tail of one piece of the signal is lined up with the head of the next as these pieces of the signal are taken to do the calculations. The effect that no pre-emphasis has on the LPC calculation is unknown. For further research the accuracy of the LPC calculation may be improved by using an

ARMA model, which is slightly more accurate than the AR model. In addition, the mathematical model was based on a vocal tract modal that assumes the vocal tract is a finite series of single tubes. The actual vocal tract is an infinite series of tubes. Recent models of the vocal tract have also included two side branches: the nasal resonance and a pair of piriform sinuses. The resonances of these cavities have been found to act as negative filters to the vocal tract resonance. This means that they will cancel out ranges of frequencies. Whether the newer vocal tract model still matches the old mathematical model is not known. The "goodness" of the old model was questioned from the onset, especially for resolving formants at higher frequencies. Markel et al. (1982: 6) stated that the algorithm was quite accurate from 20Hz to several thousand Hz and pole correction increased the accuracy of the higher poles. It is unclear what effect this might have on the LPC when pushed to 24,000 Hz. Perhaps an experimental test of LPC accuracy should be performed before more research is done.

4.2.3.1.2 Fourier Transform: algorithm and settings

Another important consideration in the analysis of voice quality is the presence or absence of partials and their relative energies. One effective method of measuring harmonics is by using the Fourier Transform algorithm. In the early 1800s Fourier theorized that all periodic complex waves were made up of a collection of sinusoids and developed the transform algorithm. When applying Fourier techniques to signals that have been sampled in time, the Discrete Fourier Transform (DFT) is used:

$$Y(k) = \sum y(n)e^{-i\omega nk} = y(n) = [\cos(\omega nk) + i \sin(\omega nk)]$$

Where $i = \sqrt{-1}$, ω = angular frequency, n = the time-domain sample number, and k= the frequency bin number. Using the DFT, the original function, the voice signal y(n), can be broken down into the odd (antisymmetric) and even (symmetric) parts, and within those parts are found the frequency components, their amplitudes, and their phases. By using Euler's trigonometric identity, we see that the odd part corresponds to the cosine terms and the odd part corresponds to the sine terms. It is a very good thing that we have computers to do these calculations since they involve thousands of multiplications. This could be why discrete Fourier analysis was not popular until the 1960s when it became easy and fast to process the information on computer. The output of the DFT contains real and imaginary parts; therefore, the magnitude must be taken before plotting. The result is that time series data is converted into a frequency array. CSL 3D spectrograms use the Fast Fourier Transform (FFT). The FFT is a computationally efficient implementation of the DFT that can be used when the analysis has a specific number of data points, e.g. 1024.

The DFT analysis used in this thesis was performed on 2666 points of the signal. MATLAB only plots [x,y] vectors of the same length. A frequency vector was generated for the labeling of the x-axis whose length was 1333 points. If a DFT is plotted in the "raw" form it creates a mirror image graph with positive and negative parts. For this reason the half-length was taken for the plot. Certainly any number of points could have been used. A better resolution of the harmonics will result from an increase in the number of points used in the analysis.

MATLAB also has the option of using the FFT algorithm; however, it requires a signal length that is a power of 2 number such as 2048 points. An adjustment in this procedure may be something to consider for the analyses of the other vowels. The first type of Fourier analysis shows three views: the DFT without pre-emphasis, a ZOOM of the signal without pre-emphasis and a DFT of the pre-emphasized signal. The settings were as follows:

- samples were trimmed to contain only the steady state portion of the signal
- Gain adjusted in CSL to an average quantized value of 32,000
- sampling rate 48,000 samples per second
- signal size 2666 points
- view #1 - no pre-emphasis
- view #2 - no pre-emphasis ZOOM
- view #3 - pre-emphasis (.9)

The data that were collected from the DFT analysis can be found in Appendix A and B.

4.2.3.1.3 Spectrogram: algorithm and settings

Spectrograms, in CSL, are generated by a series of FFT calculations presented in a 3D graphical format over a period of time. The time is shown on the x-axis, the frequency on the y-axis and the amplitude on the z-axis (represented in a 2D image with a darkness scale). This spectrographic analysis technique was used to a show a general overview of the acoustic patterns in the samples. The spectrogram, in CSL, does not offer the option to see the actual numeric output from the time series FFTs. Instead, the frequency and time values reported in the spectrogram are computed from the graphic output. Internally generated results are an approximation; therefore, another method of generating formant location data was chosen (LPC). Formant location from the Spectrograms was measured by hand in mm from the x-axis and recorded in Appendix A. For further research, it may be a good idea to add a grid to this graphic output and record the estimated formant values for comparison.

The spectrogram settings were as follows:

- 1024 pt frame (approximating a 71 Hz bandwidth analog spectrogram)
- pre-emphasis at .95
- Blackman window
- Darkness scale: white (below 10.00dB), varying shades of grey (14.88,19.76,24.62,29.53,34.42,39.32,44.21,49.10) and black (above 54.00dB)
- no extra gain
- no normalization
- display range in frequency (0 - 20,000 Hz)
- sampling rate 50,000 samples per second
- linear display

The data that were collected from the spectrogram analysis can be found in Appendix A and B.

4.2.3.2 Auditory and Articulatory Analytical Techniques

Auditory analyses considered phonation, vowel consistency, nasality, pharyngeal constriction, and laryngeal position. The analysis used standard phonetic auditory labels. Information on articulation was based on the subject's descriptions of the singing sensations associated with each quality. Auditory analyses and the singer's comments on sensation are in Appendix A.

4.2.3.3 The Laryngograph: machine and settings

The Laryngograph (LGG) was first developed and used by Fabre in 1957, and now is in widespread use by speech-language pathologists, and by voice researchers. The apparatus contains two electrodes attached to a neck strap. When the strap is fastened around the neck, the electrodes sit on either side of the thyroid cartilage. A small electrical current is sent through neck between electrodes. Depending on the machine, the conductance or the impedance through the tissue is measured. Conductance measures the magnitude of the electrical current through the tissue, whereas the impedance is the magnitude of the tissue's resistance to the passage of electrical current. Human tissue is a moderately good conductor of electricity, so the electrical current can be kept extremely small. This also prevents tissue damage, and avoids extraneous physiological responses. One such undesirable response is the electrical stimulation of nerve fibres. This can cause messages to be sent to the muscles to contract, and can change the density of the tissue. Another consideration is the perceptibility of the current by the subject. High frequency currents tend to be less perceptible to human tissue; therefore, LGG generally use a very small high frequency electrical current. The electrical conductance depends on the chemical composition of the tissue and the shape of the structure. This means the current flow can be affected by changes in the proportion of various tissue types, and their corresponding densities. The larynx is composed of many difference tissue types including muscle and cartilage. When the electrical current flows through the laryngeal region the vocal folds are periodically separated by an air filled space called the glottis. Although human tissue is a good conductor of electricity, air is not. Therefore, electrical impedance through air will be much greater when the glottis is open, than when the vocal folds are in contact. When the electrical current moves through the larynx a rising and falling conductance or impedance pattern can be measured, and a graphical output created. Various measurements can be taken from the data or graphical output that give insight into vocal fold vibration.

A few considerations must be taken into account when dealing with the LGG measurements. First, the larynx is very small, and the electrical current spreads out in all directions. For example, the electricity can flow around the glottis when it is open; therefore, the current flow never really stops. It is only redirected. Titze (1990) took a detailed look at the path of the current when he performed some experimental studies on the shape of the electrical field in the neck during LGG. Second, the current passes through many tissues that have no relevance to vocal fold movement. Any number of changes in these surrounding tissues can influence the LGG measurements. For example, the larynx can move up, down or tilt, thereby altering the cartilages relationship to each other. The muscle tissue may contract, thereby changing the tissue density. The electrodes can also change position slightly. Luckily, LGGs are built to deal with these factors.

The LGG used in this research was a Portable Laryngograph (LGG) (Serial 950904 Reference FLX/MG/95). This machine measures the conductance of the electrical current through the tissue. The LGG also uses an automatic AM detector and filtering to better extract the parts of the conductance signal that are the most useful for voice specialists. The LGG contains a stable high-frequency oscillator, in this case producing a 3 MHz current frequency that travels to two gold-plated, guard-ring electrodes on the neck. The guard ring acts as a ground, shields the electrodes from noise, and restricts the influence of surface current. The neck current is very small, in this case only 3 mA, which is lower than the 10 mA permitted current. The trans-neck voltage varies according to the neck conductance of the user; however, it is likely about 2 Vrms (Fourcin, personal correspondence). If the current path changes from movement of the electrodes on the neck, or movement of the tissues and structures inside the neck, then the conductance changes, and the magnitude of the alternating current changes. This is called amplitude modulation (AM) of a carrier frequency. In order to make these amplitude changes less significant, the LGG has an automatic AM detector that identifies drops in the magnitude of the carrier amplitude and adjusts accordingly. Neck movement may also cause baseline drift. The LGG corrects this with a high-pass filter whose cut-off frequency is approximately 6 Hz. The LGG also has a low-pass filter characteristic that has a gentle roll-off up to about 100 kHz. This filter only reduces high frequency noise in the signal to a limited extent but provides the LGG with exceptionally good transient response. Fourcin (1974) has proposed that the original signal be called "Gx" by voice researchers, whereas the modified signal, after AM detection and filtering, be called "Lx".

The LGG analyses used the analog Lx signal from the laryngograph. The Lx signal was digitally sampled at 48,000 samples per second, saved on DAT tape, then transferred in analog format from the DAT recorder to the CSL 4300 hardware. The CSL 4300 re-captured the signal at 44,100 samples per second. The CSL manual (Kay Elemetrics, 1991) states that when a signal is transferred from a device, such as the DAT player, to the CSL 4300, the signal may show phase distortion if the AC coupling feature on the CSL 4300 is selected. The LGG signal was input to the CSL 4300 with the default setting of AC coupling OFF. The digital signals were trimmed and saved in computer files in NSP format; however, there was baseline drift observed in the Lx signals even after the LGG filtering processes.

CSL EGG software filters the Lx signal again with high and low pass filters before processing:

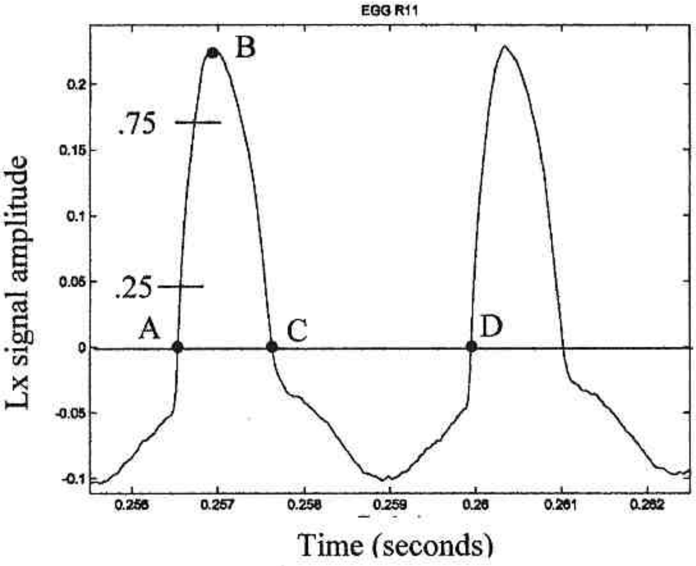
- High Pass Filter: In order to remove the baseline drift caused by movement of the larynx during data capture the software uses a two-pass moving average filter of 10 msecs duration to estimate the low pass component, which is then removed from the original signal by subtraction.
- Low Pass Filter: the low-pass filter that is used to reduce broadband noise is a moving average filter of duration 0.3 msecs. This filter is applied to the entire captured waveform. The cutoff is approximately 1500 Hz at all the supported sampling rates.

A visual inspection of the Lx waveforms found that the high pass filter had removed most of the baseline drift.

A few wavelengths of each Lx signal were saved in graphical format in CSL as well as the graphical output in TIFF format of speed quotient and contact quotient analyses over time (Appendix B). The NSP Lx files were converted to WAV Lx files and MATLAB was used to generate a graphical output to visually compare the waveforms and to measure the slope of a representative waveform for each sample (Appendix B). The numerical results from CSL and MATLAB were transferred to spreadsheet format (Appendix B).

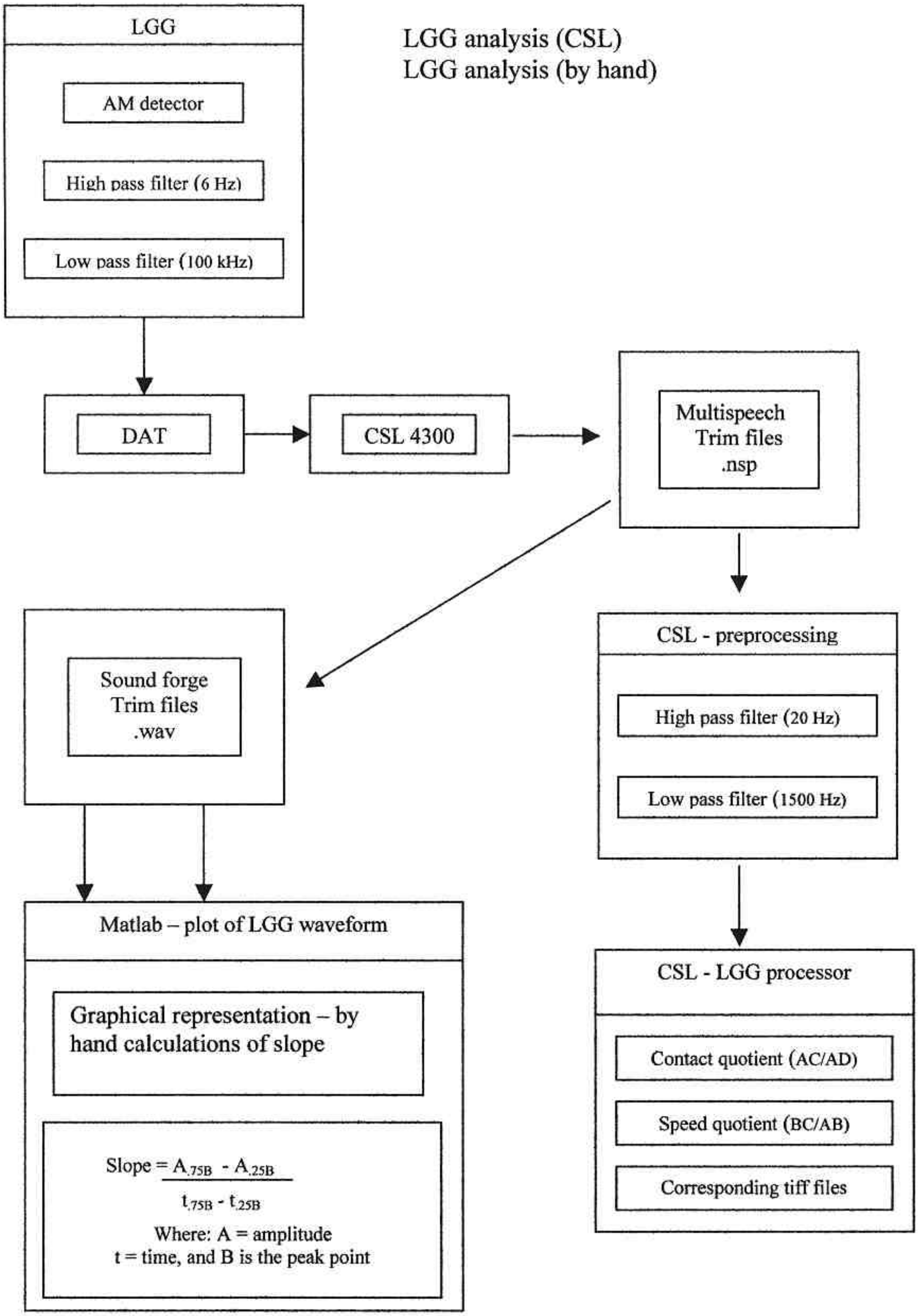
Figure 10 is an example of an LGG signal. CSL finds various points in the signal and calculates ratios between points. The ratios considered to be most important were the Contact Quotient $\frac{t_C - t_A}{t_D - t_A}$ and the Speed Quotient $\frac{t_C - t_B}{t_B - t_A}$ The slope between A – B from .25-.75 of the peak and from 0-.25 of the peak was estimated for a representative wave from each sample.

Figure 10: A sample of the Lx waveform for R11 with points A, B, C and D assigned and the ascending slope portion .25-.75 marked.



There were some concerns about filtering the signal with a filter a 20 Hz cutoff. According to Baken and Orlikoff (2000) a waveform will be distorted if a filter with a 20 Hz cutoff is applied to the signal. They cite early research by Michel et al. (1970) that illustrates how the shape of the Lx waveform obtained from a single male subject varied with high pass filtering. The signal in Michel et al.'s research was filtered with an analog filter; the analog filtering process drastically altered the slope at a cutoff of 20 Hz. CSL uses a digital filtering process. The affect of CSL filters on ascending slope in the Lx waveform is unknown; however, a comparative study should be done in the future to look

Figure 12: Laryngograph signal path



at unfiltered vs. filtered samples. The signal path for the laryngograph analysis is shown in Figure 12.

4.2.4 Measurements

Spectral measurements were taken from voice samples of all seven singing qualities performed by the professional soprano. The spectral measurements compared include the relative strength of the first two partials, the frequency of formant peaks, relative energy of partial strength near the formants, summed energy from 0-3000 Hz and 3001-5000 Hz, and the inharmonic or aperiodic activity seen in each quality. These particular measurements were chosen because previous studies by Estill (1988), Bevan (1989) and Schutte (1993) indicated that key areas, such as the relative strength of particular partials and the location of the formants, could be distinguishing features of singing voice qualities. Data from the laryngograph were used to calculate the contact quotient, speed quotient and ascending slope. The contact quotient and the speed quotient were chosen because these were ratios available from CSL. Baken and Orlikoff (2000) also recommended the ascending slope as an indicator of vocal fold zippering. The slope measurements were taken and calculated by hand, and the Lx waveform was analysed visually using Titze's (1990) criteria. Finally, the acoustic and LGG data were compared to an auditory analysis by Dr. John Esling (Professor of Linguistics, UVic) and to the subject's descriptions of physical configurations involved in producing these qualities.

The measurements for the relative strength of the first two partials were taken to establish the likelihood of formant tracking. Each singing sample is a complex sound that can be broken down by various algorithms to show its harmonic content in terms of *partials*. The first partial (P1) can also be referred to as the fundamental (F_o). If a partial appears very much stronger than adjacent partials, then it was assumed that it must be in the vicinity of a formant. Formants are resonances of the vocal tract. These resonances appear on the graphical representation of the signal as portions of the harmonic spectrum with increased amplitude. Formants are caused by pressure wave interference patterns within the vocal tract, and created when glottal pressure waves reflect and interact constructively, somewhat like enhanced echoes. The lining up of a formant with a partial is called *formant tracking.*

. P1 and P2 strengths were compared in both vowel samples and running samples for each quality using both spectrogram and DFT data. Formant tracking was analysed by looking at the relative strengths of the first and second partials (P1 and P2), and by examining the LPC/DFT overlays. LPC is considered to be very good at predicting formant location; however, it is less reliable at predicting bandwidth. For this reason, bandwidths were assessed by observing the energy in the partials adjacent to the formants. Energy levels for the partials were found in the DFT numerical results (Appendix B) for each sample. An example of the analyses for spectrogram, DFT and LPC/DFT overlay is included for speech (Figures 13-15). Data for P1:P2 formant tracking analyses were compiled in Table 10. Graphs and tables for the other seven singing voice qualities can be found on the CD ROM in the Appendices A and B.

Measurements were also taken to establish the relative strength of F2 and the singer's formant. F2 is referred to as the *vowel format* because it is the acoustic cue for the vowel quality. Researchers have attributed the vowel formant placement to the tongue position. The [i] vowel has a close front tongue position, which expands the dimensions of the pharynx while forming a constriction in the oral cavity. The widened pharynx lowers F1 and raises F2 (Titze, 1997). The *singer's formant* (SF) is the clustering of F3, F4 and F5 between 3000 Hz and 4000 Hz (Sundberg, 1987; Miller, 1996). The singer's formant is the ring in the voice. This energy distinguishes the singing voice from other instruments and makes it more audible (Sundberg, 1987). According to the model for singer's ring, F3 usually moves up and F4 and F5 move down (Titze, 1994). In the [i] vowel, F1 and F2 are spread, which now brings F2 in the vicinity of the cluster. This may add extra energy to the singer's formant. It has been suggested by Titze (1997) that it may also pull F2 higher toward the cluster. Sundberg (1987: 118-119) attributes the singer's formant to a portion of the vocal tract between the glottis and an aryepiglottal constriction specific to some qualities of singing. He suggests that the optimal conditions for resonance occur if $A1/A2 <1/6$ where A1 is the exit area of the epilarynx and A2 is the expanded area into the lower pharynx including the piriform sinuses. Typically, the sub-tube in the lower pharynx responsible for singer's ring is about 1/6 the length of the vocal tract. In an average female, this portion should ring at 3,500Hz (F1) and 10,500Hz (F2).

Formants for the singing qualities were compared with the subject's speech and to statistics for the [i] vowel; Barney and Peterson recorded resonances for the [i] vowel in an average woman at 310 Hz (F1), 2790 Hz (F2) and 3310 Hz (F3), as cited by (Ball, 1993). For comparison, I also made predictions of formant locations for the subject based on the vocal tract model. The formant predictions were calculated by assuming that F5 was in the same location in the [i] vowel as in a vocal tract model of the same length. The vocal tract model is based on a uniform diameter closed one-end tube. The closest vocal tract shape to this model would be found in the schwa. Titze (1994:146) outlines an equation $Fn = (2n-1)(F1)$ Hz that can be used to estimate the formants. If $n = 5$ and $F5 = 5339$ Hz, then by dividing 5339 by 9 should give an approximation of F1. If F1 is returned to the equation, then the estimated vocal tract resonances for the schwa can be calculated. These resonances were predicted at 593 Hz (F1), 1779 Hz (F2), 2865 Hz (F3), 4151 Hz (F4), and 5337 (F5) Hz. The cluster of formants called the singer's formant should be between F3 and F4, which is approximately 3500 Hz. To check the predicted location of the singer's formant another calculation was made. The length of the subject's vocal tract was not measured in this thesis; however it was estimated by the using the formant locations and Titze's (1994: 143) equation $Fn = (2n-1)(c/4L)$ where n = formant number, c is the speed of sound in moist warm air (350m/s) and L = the length of the vocal tract. By using F5 = 5339 Hz, the estimated length of the subject's vocal tract was calculated as 14.8 cm for speech. The approximate frequency for the singer's ring can be calculated by using 1/6 of the vocal tract length as the length of the resonating "sub-tube". The calculations for singer's ring would be as follows:

Calculating length of singer's ring sub- tube: 1/6 x 14.7 cm = 2.5 cm or .025 m

Calculating the first formant of the singer's ring sub-tube (Titze, 1994:143):

$$\begin{aligned} Fn &= (2n-1)\ c/4L \\ SF &= (1)\ 350\ \text{m/s}\ /4(.025\text{m}) \\ &= 3500\ \text{Hz} \end{aligned}$$

The vowel formant and singer's formant placement were analysed by using DFT and LPC. Numerical results from the DFT analyses were used to compare energy in the partials adjacent to the formants. Mean formant frequencies from the LPC analyses were used to compare formant locations. The summed energy was used to compare the relative strengths of the vowel formant and the singer's formant. A graph of singer's formant energy (Q2) vs. vowel formant energy (Q1) is found in Figure 16. Mean formant locations for F2–F6 were computed by averaging the results from LPC and mean energy was computed by averaging the results from the DFT numerical results for the samples for each quality (Appendix B). These results are given in Table 11. The detailed data analysis is in Appendix A.

LPC numerical results of formant frequency and quantized amplitude were extracted with the ZOOM function on full-length samples and 2500-point samples. Scatterplots (F2 vs. F1, F3 vs. F1, F4 vs. F1, F5 vs. F1, F6 vs. F1, F7 vs. F1 and F8 vs. F1) were made comparing formant frequencies from LPC analysis for various singing voice qualities (Appendix A). DFT numerical results included formant frequency, quantized amplitude and relative energy in the adjacent partials as a percentage of P1. For summed energy 0-3000 Hz (Q1) vs. summed energy from 3001 Hz-5000 Hz (Q2) graph, the energy in the fundamental was subtracted from Q1 before plotting. The total summed energy of each sample was set at one. This procedure allowed the fraction of energy for each sample to be represented by the axis coordinates. In further research it would be recommended to subtract the energy in both the first and second partials from the Q1 value; in voice qualities with high energy in P2, the summed energy of Q1 cannot be attributed only to the vowel formant.

DFT analyses were graphed using the fundamental frequency and index as a guide in a stem plot. The computer searched within an area (.4 x index) around multiples of the fundamental index. For example, if the maximum was found at point 10, then the computer looked for the highest value between point 6 and point 14, and again between 16 and 24, 26-34, 36-44 etc., until it generated a vector containing these points and their amplitudes. The frequency was calculated from the index. A list was made of frequency, amplitude and percentage amplitude. Dividing the amplitudes of the partials by the fundamental amplitude and then multiplying by 100 calculated the percentage amplitude. The relative amplitudes were compared among the samples. Energy between the partials was zeroed. Finally, DFT stem plots were overlaid with LPC. The stem plot was used to observe formant tracking.

Laryngograph measurements were organized into spreadsheet format in Appendix B. Graphical representations of frequency vs. closed quotient and frequency vs. speed quotient were produced in MATLAB, and representative sections of the Lx waveform were printed out to calculate the slope and perform the visual analysis for each sample (Appendix B). Lx waveform samples were inspected for evidence of peak widening, peak skewing, triangularity, and skirt elevation, and the results linked to other observations of acoustic, auditory, articulatory data and vocal fold function.

The speed quotient (SQ) calculations showed the skew of the glottal wave pulse numerically. SQ is the decontacting time divided by the contacting time of the vocal folds (Figure 10). Higher SQ values indicate a greater skew and suggest a greater angle of convergence of the vocal folds (Figure 7). The contact quotient measurement is the contact time divided by the period. (Figure 12) This calculation shows the width of the glottal pulse numerically. CQ is thought to relate to the closed phase of the vocal folds; however, it is more directly linked to the contact area. Higher CQ values would indicate greater adduction of the vocal folds (Figure 6).

Activity between the partials on the spectrogram and activity without the presence of visible partials was considered inharmonic or aperiodic activity. It was judged on a scale of 0-5. Zero corresponded to no inharmonic activity and 5 corresponded to extreme inharmonic activity on the spectrograms. The mean of each sample set was taken.

The auditory analysis measured parameters with various scales. Pharyngeal constriction was measured by assigning the following labels: no, slight slight constriction, slight constriction, slight pharyngeal, pharyngeal, 2 pharyngeal. For simplification each setting was given a number between zero and five, zero being no constriction and five being 2 pharyngeal. The auditory analysis for phonation used the following labels: Creaky Voice (CV), Harsh Voice (HV), Modal Voice (MV), Whispery Voice (WV) and Breathy Voice (BV). The qualifiers used in the phonation analysis were as follows: slight slight (1), slight (2), one (3) and two (4). For simplification, numerical qualifiers from one through four were used. Laryngeal height was measured by assigning the following labels: 2 LL, LL, sl. LL, neutral, sl. RL, RL, 2RL. The auditory analysis didn't take into account the vocal fold adjustment of falsetto vs. modal. Any balanced phonation exhibiting neither hyper nor hypofunction was considered modal.

4.3 Results

4.3.1 Acoustic Results

4.3.5.1 Formant Tracking: fundamental (P1) with the first formant (F1)

Spectrogram analysis for speech showed strong energy in both P1 and P2 for the [i] vowel (Figure 13). The DFT analysis numerical results of the spoken vowel [i] measured the P2 to be 76.3 % of P1 (Figure 14). The LPC/DFT overlay indicates that the F1 fell between P1 and P2; therefore F1 was not tracking P1 or P2 (Figure 15).

Figure 13: The spectrogram of spoken vowel [i] shows strong energy in both the first and second partials.

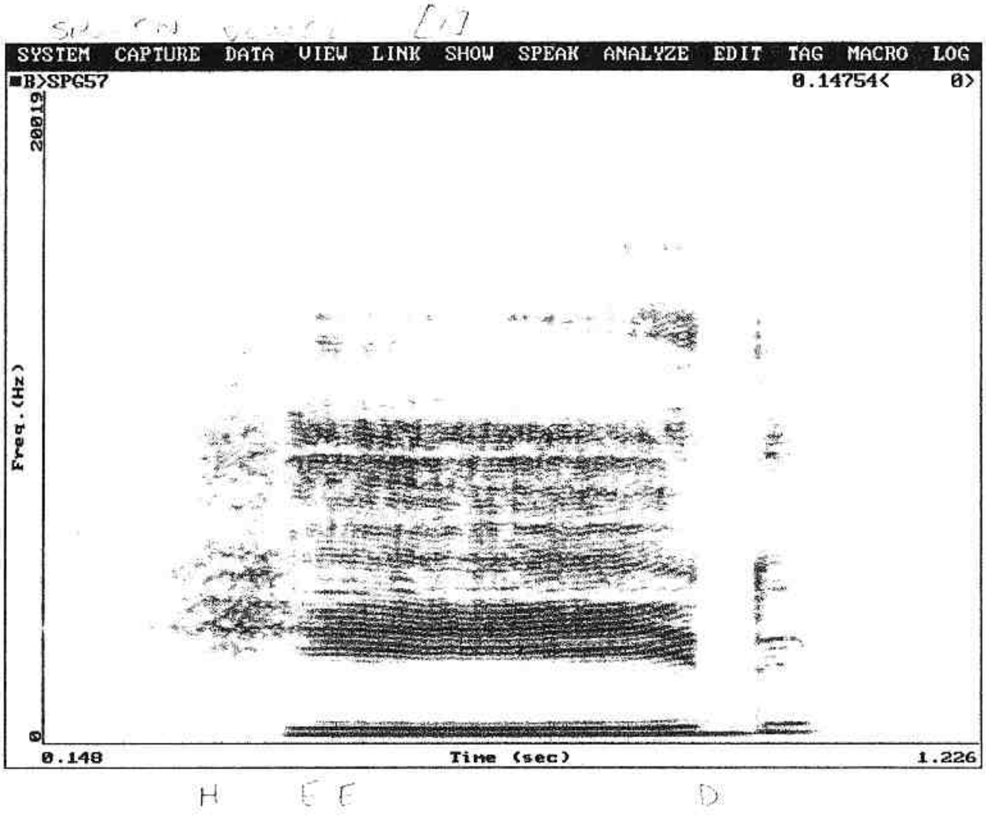


Figure 14: A DFT analysis of spoken vowel [i] with no pre-emphasis shows the relative energy of P1 and P2.

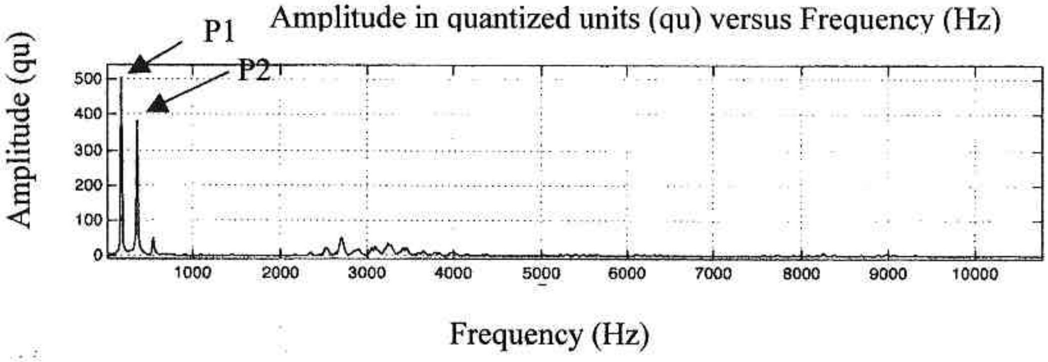
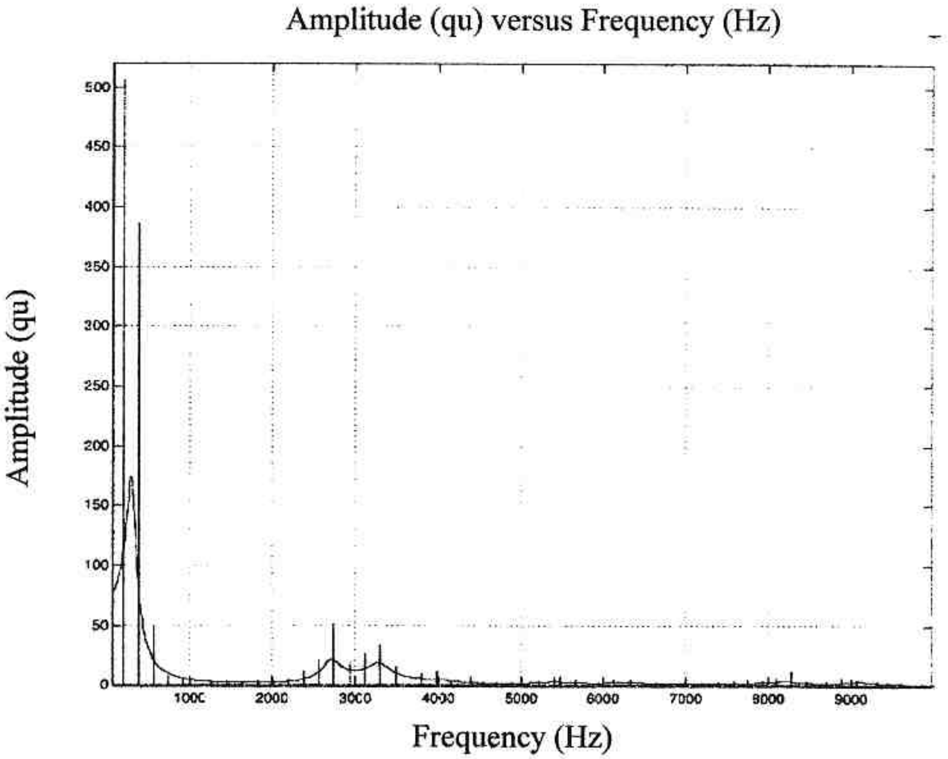


Figure 15: A DFT analysis without pre-emphasis for the spoken [i] vowel is shown in a stem plot graph. The stem plot has been overlaid with a LPC analysis without pre-emphasis (red). The DFT/LPC overlay shows poor tracking of P1 by F1.



Spectrogram analysis of sustained vowel samples for Pop showed strength in P1 and P2 for the lowest pitch (252 Hz), but less strength in P2 as the pitch increased. In the running samples, P1 was stronger than P2. The DFT results found that P2 was strong (41.5 %) for the lowest pitch, but for the higher pitches P2 was weak. The mean P2 energy was 5.0 %. The LPC/DFT overlay showed that most samples had exact tracking with two showing extremely slight negative misalignment. The F1 bandwidth was narrow; however, it showed greater width for the lowest pitch Pop sample.

Country samples showed substantial energy in both P1 and P2 in the spectrogram. As with Pop, the energy in P2 decreased for higher pitches. P1 had more energy than P2 in the running sample. DFT results also showed that P1 had more energy than P2, and that P2 had less energy at higher pitches; however, the steady state middle pitch (E4) showed the lowest P2 energy. Country and Pop had same pattern middle pitch showed lowest P2 energy. In addition, three samples of similar pitch (samples C1, C31 and C37) showed varying degrees of energy in P2 (31.0 %, 34.2 % and 15.1 %). In one instance a higher pitch showed greater energy than a lower pitch. Sample C21, whose frequency was 486 Hz, showed energy at 20.8 %; whereas sample C37, whose pitch was 252 Hz, showed energy at 15.1 %. The LPC/DFT overlay showed very good tracking with two samples showing exact tracking, two samples with slight negative misalignment and one sample with slight positive misalignment. Generally, there was more energy in P2 for this quality than for Pop. The mean energy in P2 was 21.0 % of P1.

Spectrograms for Jazz quality showed more energy in P1 for both steady state vowel samples and running samples. DFT results showed that P1 had more energy than

P2. P2 was showed strong energy of 30.5 % for the lowest pitch, but weak energy for the higher pitches. The mean P2 energy was 6.0 % of P1. The LPC/DFT overlay showed that one sample had exact tracking with two samples showing extremely slight negative misalignment. The F1 bandwidth was narrow for most Jazz samples; however, the bandwidth was wider for the low pitch sample.

Spectrograms for R&B quality showed more energy in P1 for both steady state vowel and running samples. Sample R21, however, showed more P2 energy than the other R&B samples. DFT numerical results showed P2 energy for sample R21 was 22.3 %. LPC/DFT tracking showed extremely slight (negative) misalignment for four out of five samples. One sample showed slight (negative) misalignment. The mean P2 energy was 9.0 %; however, the range was from 2.6 % – 22.3 %. There was not the same consistency as with some of the other qualities. For example, the three steady state vowels showed varying energy for P1 (pitch B3 = 9.0 %, pitch E4 = 2.6 % and pitch B4 = 22.3 %). Running samples showed more consistency than the steady state vowels (7.1 % and 5.2 %). Excluding R21, which was a steady state vowel sample with pitch B4, the mean dropped to 6.0 %, which is similar to the mean energy of Jazz.

Spectrogram and DFT for Classical quality showed more energy in P1 than P2 for both vowel and running samples. The level of energy in P2 was generally weak. The lowest pitch showed more energy in P2 than the higher pitches; however, this was still quite low when compared with the other qualities. For example, the Classical steady state vowel on pitch B3 showed P2 energy to be 19.8 %; whereas, the range of the other samples was between 3.7 % and 6.4 %. The tracking was excellent for this quality any misalignment was extremely slight and on the negative side. The mean was 9.0 % for all samples and 5.0 % for just the higher pitches. This is very similar to the R&B quality.

Spectrograms for Legit quality showed more energy in P1 for both vowel and running samples. The DFT results showed that P2 was of medium strength for the lowest pitches (29.2 % and 20.3 %), but for the higher pitches P2 was weak. The mean energy for P2 was 6.0 %. LPC/DFT tracking showed extremely slight negative misalignment for five samples and excellent formant tracking for two samples.

Spectrograms of Belt steady state vowel samples showed strength in both P1 and P2 for the lowest pitch sample and the running samples. The higher pitch steady state vowels showed less energy in P2. DFT results showed similar findings. Generally, there was more P2 energy in Belt quality than in any other qualities studied in this thesis. The mean P2 energy for these samples was 27.0 % of P1, and without the highest pitch samples (468 Hz and 486 Hz) the mean rose to 32.0 %.

4.3.5.2 Vowel Formant (F2) and Singer's Formant (F3-F5 cluster)

The LPC analysis showed the subject's [i] vowel for speech had formants at 316 Hz, 2683 Hz, 3246 Hz, no F4, and 5339 Hz. The DFT revealed there might be F4 energy around 4390Hz. When the formants for the subject's spoken [i] are compared with the predicted formants for a single diameter tube as calculated in section 5.1.4 Measurements, then F1 moved down, F2, F3, and F4 moved up and F5 remained the same. LPC showed the mean F2 for speech at 2683Hz. The DFT estimated F2 at 2736 Hz. The DFT analysis found that the mean energy at the partial closest to F2 was 10.0 %

Table 10: A summary of F1: P1 formant tracking for various voice qualities. Voice samples were analysed with SPG, DFT and DFT/LPC.

SPG = Spectrogram
DFT = Discrete Fourier Transform
LPC = Linear Predictive Coding
DFT/LPC = DFT stem plot overlaid by LPC

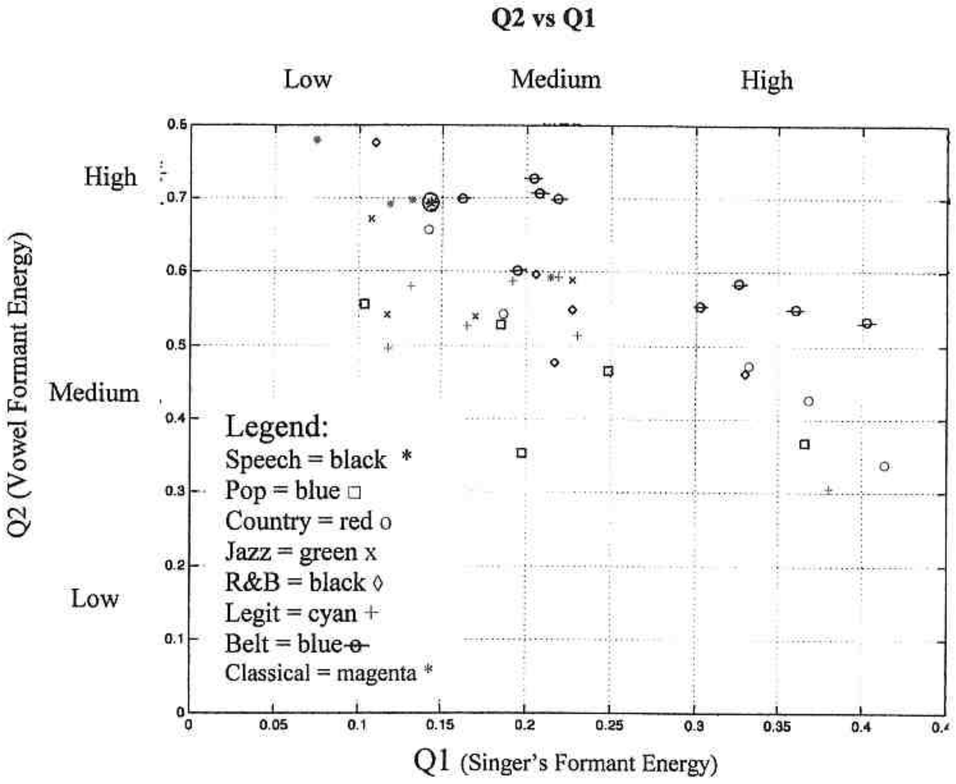
F1 = first formant
P1 = fundamental frequency
P2 = second partial

Voice Quality	**Sample** (pitch)	*** SPG Analysis:**	**DFT Analysis:**	**DFT/LPC Overlay:**
Speech	Spk7	$P_1=P_2$	$P_1>P_2$ Energy: $0.76P_1 = P_2$ Range: non	Poor
Pop	P1(B3) P12(E4) P22(B4) P33 P39	$P_1=P_2$ on low pitches $P_1>P_2$ as pitch increased	$P_1>P_2$ Energy: $0.12P_1 = P_2$ Range: 0.03-0.42	Excellent Some + misalignment
Country	C1(B3) C31(B3) C11(E4) C21(B4) C37	$P_1=P_2$ on low pitches $P_1>P_2$ as pitch increased	$P_1>P_2$ Energy: $0.21P_1 = P_2$ Range: 0.04-0.34	Very good
Jazz	J1 (B3) J21 (B4) J33 J38b&c	$P_1>P_2$	$P_1>P_2$ Energy: $0.13P_1 = P_2$ Range: 0.06-0.31	Good
R&B	R1(B3) R11(E4) R21(B3) R31 R37	$P_1>P_2$	$P_1>P_2$ Energy: $0.09P_1 = P_2$ Range: 0.02-0.22	Good
Classical	O11(B3) O21((E4) O31(B4) O44	$P_1>P_2$	$P_1>P_2$ Energy: $0.09P_1 = P_2$ Range: 0.03-0.20	Excellent
Legit	L1(B3) L2(B3) L12(E4) L22(B4) L34 L35	$P_1>P_2$	$P_1>P_2$ Energy: $.12P_1 = P_2$ Range: .05-.29	Excellent
Belt	B1(B3) B11(E4) B21(B4) B31(B4) B40b, c&d B41 B45	$P_1=P_2$ on low pitches $P_1>P_2$ as pitch increased	$P_1>P_2$ Energy: $0.27P_1 = P_2$ Range: 0.06-0.51	Good More + misalignment than other qualities (formant between P_1 & P_2)

* Note SPG had default pre-emphasis .95

of the fundamental. When comparing all qualities on the summed energy plot (Figure 16), it was found that 10 % constituted a relatively high energy in the vowel formant. The LPC also found formant energy at 3246 Hz and 5339 Hz. The DFT found energy in partials at 3978 Hz and 4390 Hz. This was labeled as F3 (3246 Hz), F4 (3978 Hz) F5 (4390 Hz) and F6 (5339 Hz). The mean F3-F6 energies were F3 (7.0 %), F4 (2.0 %), F5 (2.0 %) and F6 (2.0 %). The summed energy plot found the energy in the range of the singer's formant to be relatively low (Figure 16).

Figure 16: Q2 vs. Q1: A summary of the energy in the vowel formant and the singer's formant was estimated by calculating the summed energy in the range of 0 Hz to 3000 Hz-P1 (Q1), and the summed energy in the range 3001 Hz to 5000Hz (Q2).



LPC showed the mean F2 for Pop at 2750 Hz. This is a little higher frequency than Speech. The higher pitched steady state samples showed energy ranging from 3000 Hz – 3100 Hz. A scatterplot also showed F2 for Pop equal or higher than Speech. The DFT analysis found that the mean energy at the partial closest to the F2 formant was 4.2 %. When comparing all qualities on the summed energy plot, it was found that the energy in Q1 (0-3000Hz minus P1) showed medium energy in the vowel formant. The LPC also found mean formants at 3147 Hz and 4464 Hz. The DFT found partials located near targeted formants at mean values of F3 (3215 Hz), F4 (3616 Hz), F5 (4488 Hz), and F6 (5400 Hz). F3 seems to be approximately at the same location as for Speech. The singer's formant was present at the expected location of approximately 3500 Hz. F4 and F5 seem to be in similar locations to Speech. The mean F3 – F6 energies were F3 (6.4 %), F4 (2.2 %), F5 (0.6 %) and F6 (0.6 %). The summed energy plot found that the energy between 3001 Hz-5000 Hz was variable; therefore, the mean energy values for summed energy in the vicinity of the singer's ring may not be as useful.

LPC showed the mean F2 frequency peak for Country at 2631 Hz. This value is a little lower frequency than the F2 for speech. However, the LPC showed a range of over 400 Hz for F2 locations in the Country samples. The lowest steady state pitch showed F2 lower than Speech (2484 Hz) while the two higher steady state pitches showed F2 higher than Speech (2766 Hz, 2906 Hz). The scatterplot also showed Country with variable F2. The DFT analysis found that the mean energy at the partial closest to the F2 formant was 8.0 % of the fundamental. The summed energy plot showed medium energy in the vowel formant (Figure 16). LPC found mean formant locations at F3 (3473 Hz) and F4 (4378 Hz). The DFT found partials located near the formants at mean locations of F3 (3575

Hz) and F4 (4331 Hz). There were also partials with some energy near 3172 Hz and F5 (5486 Hz). F3 was a little lower than Speech in Country samples. F4 was a little lower than for Speech and F5 a little higher than for Speech. The mean F3 – F6 energies were F3 (8.8 %), F4 (7.8 %), F5 (5 %) and F6 (0.8 %). Energy in the area of the singer's formant was present at the expected location. The majority of Country samples (3/5) showed high summed energy between 3001 Hz –5000 Hz. One sample showed low energy and one sample showed medium energy. Generally, Country quality showed some of the highest energy in the singer's formant area (Q2).

LPC showed the mean F2 frequency peak for Jazz at 2525 Hz. This value is a little lower frequency than F2 for speech. The scatterplot also showed F2 as a little lower than speech. The DFT analysis found that the mean energy at the partial closest to the F2 formant was 5.3 % of the fundamental. The summed energy plot showed the energy for the vowel formant was medium (3/4 samples). One sample showed high energy in vowel formant. The LPC found mean formant locations at F2 (2883 Hz), F4 (3803 Hz), and F6 (5396 Hz). However, in the 2500 point LPCs, F3 and F5 were missing and SF and F4 only appear in the half the cases. The DFT found partials located near the formants at mean locations of F3 (3290 Hz), F4 (3684 Hz), F5 (4488 Hz) and F6 (5400 Hz). The mean F3 – F6 energies were F3 (2.3 %), F4 (3.0 %), F5 (1.3 %) and F6 (0.3 %), which is relatively low energy. The summed plot confirmed these findings showing low-medium summed energy between 3001 Hz-5000 Hz (Figure 16).

LPC showed the mean F2 frequency peak for R&B at 2152 Hz. This value for F2 is much lower than Speech. LPC showed that F2 rose with ascending pitch from B3-E4, but then dropped for the highest pitch B4. The scatterplot also showed F2 to be much lower than Speech. The DFT analysis found that the mean energy at the partial closest to the F2 formant was 6.8 % of the fundamental. The majority of samples showed low energy; however, the steady state sample on B4 (R21) showed very high energy (23 %) so the mean results (6.8 %) were a bit higher than for the median (4.0 %). The summed energy for the vowel formant (Q1) was medium (4/5 samples). One sample showed high energy in vowel formant (R21). The LPC found mean formant locations at 3006 Hz and 3961 Hz. The 2500-point LPC showed the mean F3 at 2976 Hz. The DFT found partials located near the formants at mean locations of F3 (2999 Hz), F4 (3543 Hz), F5 (4323 Hz) and F6 (5283 Hz). The mean F3 – F6 energies were F3 (7.6 %), F4 (3.2 %), F5 (1.4 %) and F6 (0.4 %). The high energy in R21 tended to make the mean a little higher than the median for F3 and F4. The median was 4.0 % for F3 and 1.0 % for F4. Generally, R&B quality showed low energy above the fundamental. The summed energy plot showed medium energy between 3001 Hz-5000 Hz in 3/5 samples (Figure 16).

LPC showed the mean F2 frequency peak for Classical at 2187 Hz. This is very low compared to Speech. Numerical results for frequency placement in LPC analyses ranged from very low (1863 Hz) to slightly low (2454 Hz) as compared to Speech. LPC (2500 point) showed a mean of 2023 Hz and DFT showed a mean of 2075 Hz. Classical and R&B had the lowest F2 values. The scatterplot also showed F2 as lower than Speech. The DFT analysis found that the mean energy at the partial closest to the F2 formant was 4.5 % of the fundamental. As with R&B, the highest steady state pitch (B4) showed considerably more energy than the other steady state samples. The median energy for F2 was 2.0 %. Classical quality the summed energy in Q1 (0 – 3000 Hz minus P1) showed there was high energy in the vowel formant (3/4 samples). One sample showed medium energy. The LPC also found formant energy at 3172 Hz and 3875 Hz. LPC (2500 point) found mean formant placements at 2900 Hz, 3785 Hz, 4031 Hz and 5880 Hz. DFT partial placements near these formants were recorded at 2931 Hz, 3795 Hz, 4113Hz and 5634Hz. F3 and F4 seem a little lower than speech. The singer's formant and F5 are a little higher than expected. The mean F3 – F6 energies were F3 (3.5 %), F4 (2.3 %), F5 (0.8 %) and F6 (0.3 %) The summed energy plot found that the energy between 3001 Hz-5000 Hz was generally low (3/4 samples) (Figure 16). This quality showed the lowest energy above the vowel formant.

LPC showed the mean F2 frequency peak for Legit at 2571 Hz. This F2 value is similar to speech. Numerical results for frequency placement in LPC analyses ranged from slightly low 2484 Hz to slightly high 2719 Hz compared to speech. LPC (2500 point) showed a mean of 2565 Hz and DFT showed a mean of 2523 Hz. The scatterplot also showed F2 as similar to speech. The DFT analysis found that the mean energy at the partial closest to the F2 formant was 8.7 % of the P1, which was the second highest vowel formant energy of the qualities examined. The median energy for F2 was 7.0 %. When comparing all qualities on the summed energy plot, it was found that the energy in Q1 (0 – 3000 Hz minus P1) showed medium energy in the vowel formant (7/7 samples). The LPC also found formant energy at 3232 Hz and 4037 Hz. LPC (2500 point) found mean formant placements at 3220 Hz, 3773 Hz, and 4121 Hz. DFT partial placements near these formants were recorded at 3132 Hz, 3687 Hz, 4368 Hz and 5742 Hz. F3 location was similar to speech. F4 was a little high. F5 was a little lower than speech and F6 was higher. The mean F3 – F6 energies were F3 (6.0 %), F4 (2.6 %), F5 (1.7 %) and F6 (0.9 %). The summed energy plot found that the energy between 3001 Hz-5000

Hz was generally low-medium (6/7samples). One sample showed high energy in Q2 (Figure 16).

LPC showed the mean F2 frequency peak for Belt at 2044 Hz. This is very low compared to Speech. Numerical results for frequency placement in LPC analyses ranged from very low (1840 Hz) to slightly low (2297 Hz) as compared to Speech. LPC (2500 point) showed a mean of 2325 Hz and DFT showed a mean of 2353 Hz. The scatterplot also showed F2 as lower than Speech. The DFT analysis found that the mean energy at the partial closest to the F2 formant was 25.6 % of the fundamental. Belt, therefore, has the strongest vowel formant of the qualities examined in this thesis. When comparing all qualities, it was found that the summed energy plot showed high energy in the vowel formant in 5/9 samples and medium high energy in 4/9 samples. The LPC also found formant energy at 2478 Hz, 3075 Hz and 3697Hz. LPC (2500 point) found mean formant placements at 2964 Hz, 3583 Hz, and 4248Hz. DFT partial placements near these formants were recorded at 2898Hz, 3626Hz, 4246Hz and 5295Hz. F3 is a little lower than speech.

The singer's formant in Belt quality is in approximately the predicted spot (3500 Hz). F4 was a little low and F5 was missing in all samples of LPC (2500 point). The mean F3 – F6 energies from the DFT numerical results were F3 (16.2 %), F4 (15.8 %), F5 (7.1 %) and F6 (0.8 %). The summed energy between 3001 Hz-5000 Hz was medium in 5/9 samples and high in 4/9 samples. Generally, energy in this frequency range was greater for this quality than any of the others examined in this thesis (Figure 16).

Table 11 shows the activity in the area of the singer's formant from the spectrogram measurements, the mean formant locations for F2-F6 for each quality using data from both LPC calculations and DFT calculations from the nearest partial, and the mean energy in the vicinity of the singer's formant from the DFT analysis in a percentage of the fundamental. Table 11 is an attempt to reduce all the data for F2 and SF to one number for each formant and each voice quality; therefore, the data only shows generalizations (see Appendix A for more detail).

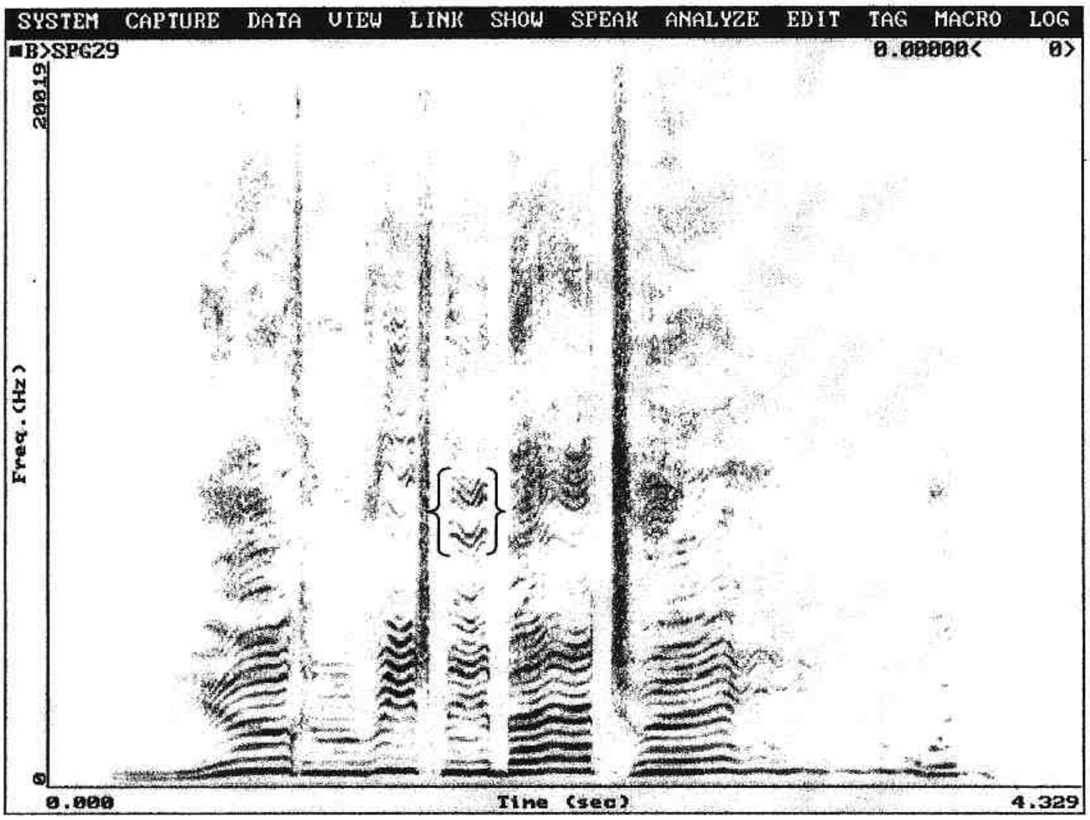
Table 11: A summary of mean formant locations and associated energy for F2 – F6 in seven singing voice qualities and speech.

		Vowel Formant (VF)	Singer's Formant (SF)				
Quality	**SPG** # of samples with SF energy	**F2**	**F3**	**F4**	**F5**	**F6**	**Singer's Formant** Clustering of F3-F5 energy (3500Hz)
Speech	0	2710 Hz Energy: 10%	3270 Hz Energy: 7%	3978 Hz Energy: 2%	4392 Hz Energy: 2%	5415 Hz Energy: 2%	
Pop	1/6	2676 Hz Energy: 4%	3215 Hz Energy: 6%	3616 Hz Energy: 2%	4482 Hz Energy: 1%	5400 Hz Energy: 1%	F3 closer to F4 than spoken
Country	3/5	2666 Hz Energy: 8%	3239 Hz Energy: 7%	3565 Hz Energy: 8%	4333 Hz Energy: 5%	5514 Hz Energy: 1%	SF area boosted F3 closer to F4 than spoken; F4 and F5 moved down closer to F3
Jazz	6/6	2578 Hz Energy: 5%	3290 Hz Energy: 2%	3717 Hz Energy: 3%	4561 Hz Energy: 1%	5400 Hz Energy: .3%	F3 closer to F4: F3 moved up and F4 moved down
R&B	2/5	2207 Hz Energy: 7%	2987 Hz Energy: 8%	3627 Hz Energy: 3%	4327 Hz Energy: 1%	5421 Hz Energy: .4%	F3 closer to F4
Classical	3/4	2049 Hz Energy: 5%	2916 Hz Energy: 4%	3789 Hz Energy: 2%	4086 Hz Energy: 1%	5757 Hz Energy: .3%	
Legit	5/7	2544 Hz Energy: 9%	3176 Hz Energy: 6%	3625 Hz Energy: 3%	4286 Hz Energy: 2%	5742 Hz Energy: 1%	F3 closer to F4;
Belt	8/9	2318 Hz Energy: 26%	2927 Hz Energy: 16%	3635 Hz Energy: 16%	4246 Hz Energy: 7%	53295 Hz Energy: 1%	SF area energy boosted

Upper Partial Activity

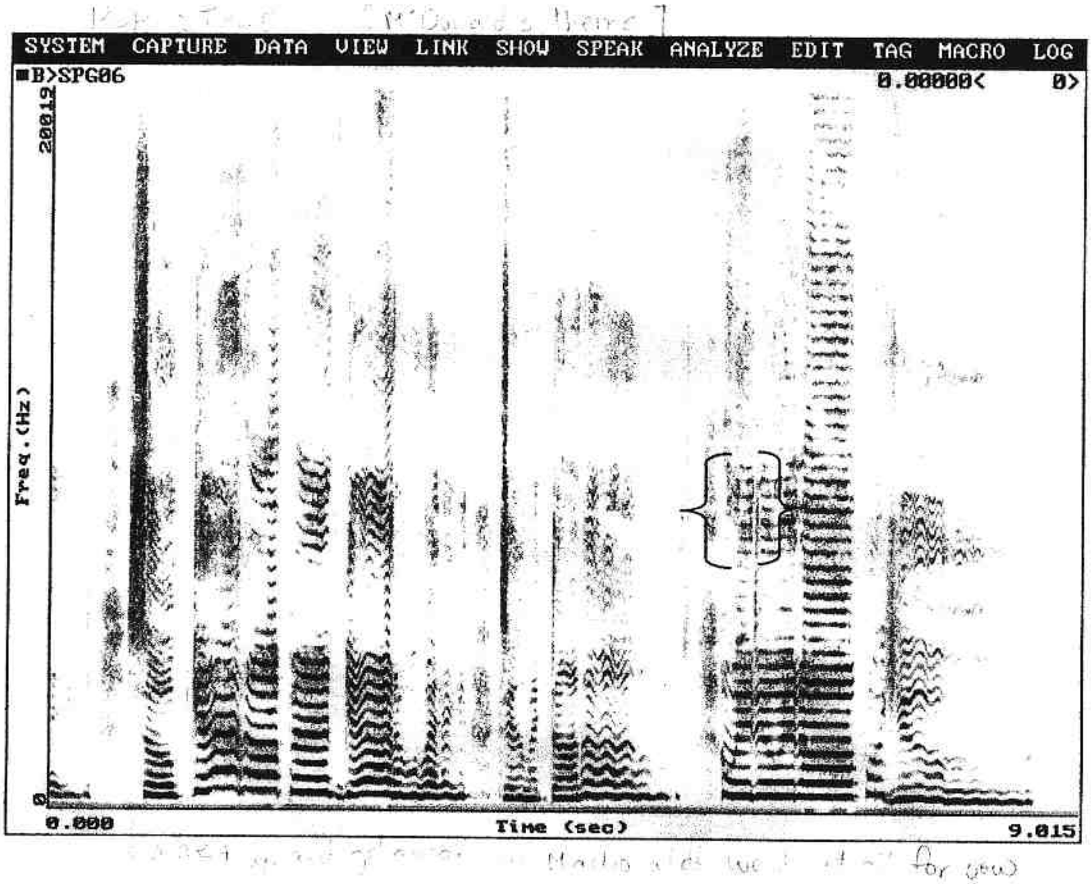
Speech showed upper partial energy at 6318 Hz (1.1 %), 8280 Hz (2.4 %) and 8892 Hz (1.1 %). Classical quality had no partials with energy over 1.0 % in the 6000 Hz – 20000 Hz range. R&B and Jazz were also fairly low in upper partial energy. Only one sample of R&B showed energy between 6480 Hz and 7866 Hz at approximately 1 % (Figure 17).

Figure 17: A spectrogram of sample R37 in R&B quality showing low energy in upper partials on the vowel [i].



Jazz also had only one sample that showed energy of 1 – 3 % in partials between 6066 Hz – 8982 Hz. On the other hand Pop, Legit, Country and Belt showed energy in the upper partials. Pop showed a band of energy on the spectrogram between 6246 Hz and 9360 Hz. Figure 18 shows the upper partial activity in P33, where the strongest upper partial was located at 8082 Hz and the energy was 1.9 % of P1. The spectrograms for Pop quality indicated both harmonic and inharmonic activity (see 5.2.9 Inharmonic Activity). Legit had slightly less inharmonic energy in the spectrograms than Pop, Country or Belt

Figure 18: A spectrogram of sample P33 in Pop quality showing a band of energy in the upper partials between 6246 Hz and 9360 Hz.



qualities, although the auditory analysis showed Whispery Voice and Breathy Voice in Legit quality. The energy in the Legit was concentrated in a slightly narrower band from 6084 Hz – 8820 Hz, and there was more evidence of formant activity. Mean formant placements for Legit were 6337 Hz, 7441 Hz, and 8434 Hz. Spectrograms for Country showed clean partial definition in the spectrograms. The auditory analysis showed very little Whispery Voice or Breathy Voice in this quality. There were two formants approximately 6500 Hz and 7500 Hz and a cluster at approximately 9000 Hz. The mean for the samples showed formants at 6257 Hz, 7410 Hz, 8527 Hz and 9146 Hz; however, the energy in the 8500 Hz range was lower than in the 9000 Hz range. The DFT showed mean energy in the 6500 Hz range at 3.4 % and in the 7500 Hz range at 2.7 %. The mean energy for Country quality around 9000 Hz was 2.3 %. Belt has a similar formant configuration to Country. Belt quality showed formants at approximately 6500 Hz and 7500 Hz and a clustering of energy approximately 8500 Hz – 8900 Hz, which was a little lower than Country, where energy clustering was approximately 9000 Hz – 9300 Hz.

4.3.5.4 Laryngeal Position

In the auditory analysis, the subject's natural speaking voice showed a slightly low laryngeal position. Pop samples showed mostly a neutral larynx, although two of the sustained vowel sounds at E4 and B4 indicated a slightly raised larynx. The subject indicated that she was attempting to sing with the larynx in a neutral position. Country showed a neutral to raised laryngeal position. The subject commented that she didn't feel these samples had a high enough laryngeal position. In the running sample, the quality was identifiable as Country, but the laryngeal position sounded as if it was neutral to slightly low. Analysis of Jazz quality suggested a lower laryngeal position as did R&B and Classical. R&B showed a generally low larynx, but in the sustained vowel samples the larynx raised with pitch. The subject noted R&B had a low larynx. Classical samples showed a generally low larynx in all samples except the first attempt at B3. This sample was repeated with the laryngeal position lower. The vowel samples in Legit varied from neutral to raised larynx in the auditory analysis. The subject indicated that she felt the larynx was in a neutral position for legit. In the running samples however, the auditory analysis indicated the larynx was lowered. Belt showed a generally slightly raised laryngeal position. The exceptions were the pitch B3, where the larynx was neutral, and the running samples of "There are guys", where the laryngeal position was variable. The subject noted the feeling of width in the vocal tract for the Belt timbre.

4.3.5.5 Pharyngeal Constriction

Speech and the Jazz qualities showed zero constriction in the auditory analysis. Legit showed generally little constriction except on the two higher pitches (E4 and B4) of the sustained vowels. The constriction on higher pitches for Legit was assessed at 3. Country, Classical and Belt showed some constriction (0-2). A trill, which is sometimes called growl, was noted in one of the Belt sustained vowels at the pitch B4. Pop showed variable constriction (0-5) depending on the sample. The running samples showed little constriction (0-2), whereas the vowels showed a greater constriction (3-5). R&B samples showed no constriction (0) on the running samples, but some constriction (2-3) on the vowels.

4.3.5.6 Vowel Consistency and Nasality

The auditory analysis showed Speech was a consistent [i] with no nasality. Pop showed some indication of the tongue being back; however, the subject indicated her tongue was forward. Both running samples had varying degrees of nasality. Country samples showed vowel modification of [e] on higher pitches and nasality in the running samples; however, nasality was not present in all of the vowel samples. Jazz showed a consistent vowel throughout, with possibly a little modification on the top note. The running samples showed nasality. The subject noted a low tongue, although this didn't seem to affect the vowel. R&B showed vowel modification on the top note, B4 of the sustained vowel samples, to [e], and no nasality in either the sustained or the running samples. The Classical sustained vowels showed nasality, but it was not present in the running sample. There was no vowel modification noted on the auditory analysis for Classical; however, in the spectrogram F1 and F2 moved closer together as the pitch increased. F1 in closer proximity to F2 is usually an acoustic cue of vowel modification. The Legit samples were quite inconsistent. Vowel quality changed throughout the sustained vowels. The running samples showed no nasality, which is consistent with the literature and the pedagogy. The auditory analysis found Belt voice had –ATR, with the tongue pulled back and slightly lowered. Nasality was noted in the running samples. It is not clear from this auditory analysis whether the perception of nasality was a result of an open velum or from a constriction somewhere in the pharynx. Two of the samples with audible constriction also were nasal sounding. However, there was one sample that was nasal without an audible constriction. Generally, there was little vowel modification with ascending pitch in any quality.

4.3.5.7 Phonation

Creaky voice (CV) was used only twice by the subject. It occurred at the start or end of a sample, which is characteristic of many musical styles: Country, Pop, Jazz, Belt and R&B. It was only seen in Pop quality and Legit quality in this series of samples. Generally, it is not characteristic of Legit, Classical or Speech in English.

Slight Harsh Voice (HV) (1-2) was heard in samples of Belt (B21), Country (C11, C21), Jazz (J21), Legit (L22), Pop (P22), and R&B (R1). Examples like B21 and C21 were redone on request of the subject. In B21 and C21, she was bothered by the HV quality. In C21, she also remarked that the larynx should have been higher. Other instances of HV occurred on the higher B4 pitch. It was often accompanied by constriction or raised larynx.

Breathy Voice (BV) or Whispery Voice (WV), on the other hand, was used more often in many of the qualities. The subject also used WV (2) in her speech. It was most evident in Jazz and R&B, and is characteristic of these qualities. Most often, an airy quality accompanied by a constriction or raised larynx results in WV. WV was also heard in conjunction with HV. Whispery Harsh Voice (WHV) was heard on higher pitches. Spectrograms showed WV as inharmonic activity in the upper partials; however, spectrograms showed BV as inharmonic activity between partials in the midrange. The steady state samples were all produced with an audible constriction and all have varying amounts of WV heard. The running samples in R&B showed BV and no audible constriction was heard. Pop showed the presence of WV and BV, as did Classical and

Legit. HV was more evident than WV or BV in Country, however WV occurred a few times in the Belt samples.

Modal Voice (MV) was the predominant phonation type in Classical, Belt and Country. Pop used MV as well, but also a variety of other phonation types: CV, WV, BV, HCV, and HWV.

4.3.5.8 Inharmonic Activity and Articulation

Table 15: The table shows the mean inharmonic activity levels for seven singing qualities and speech.

Inharmonic Activity	**Mean Activity Level (0-5)**		
High	Speech (4)	R&B (4)	Pop (3.2)
Medium	Legit (2.7)	Jazz (2.5)	Country (2.2)
Low	Belt (2.1)	Classical (1.6)	

Table 15 indicates that Speech and R&B had the most inharmonic activity of the qualities, whereas Classical, Belt and Country showed the least.

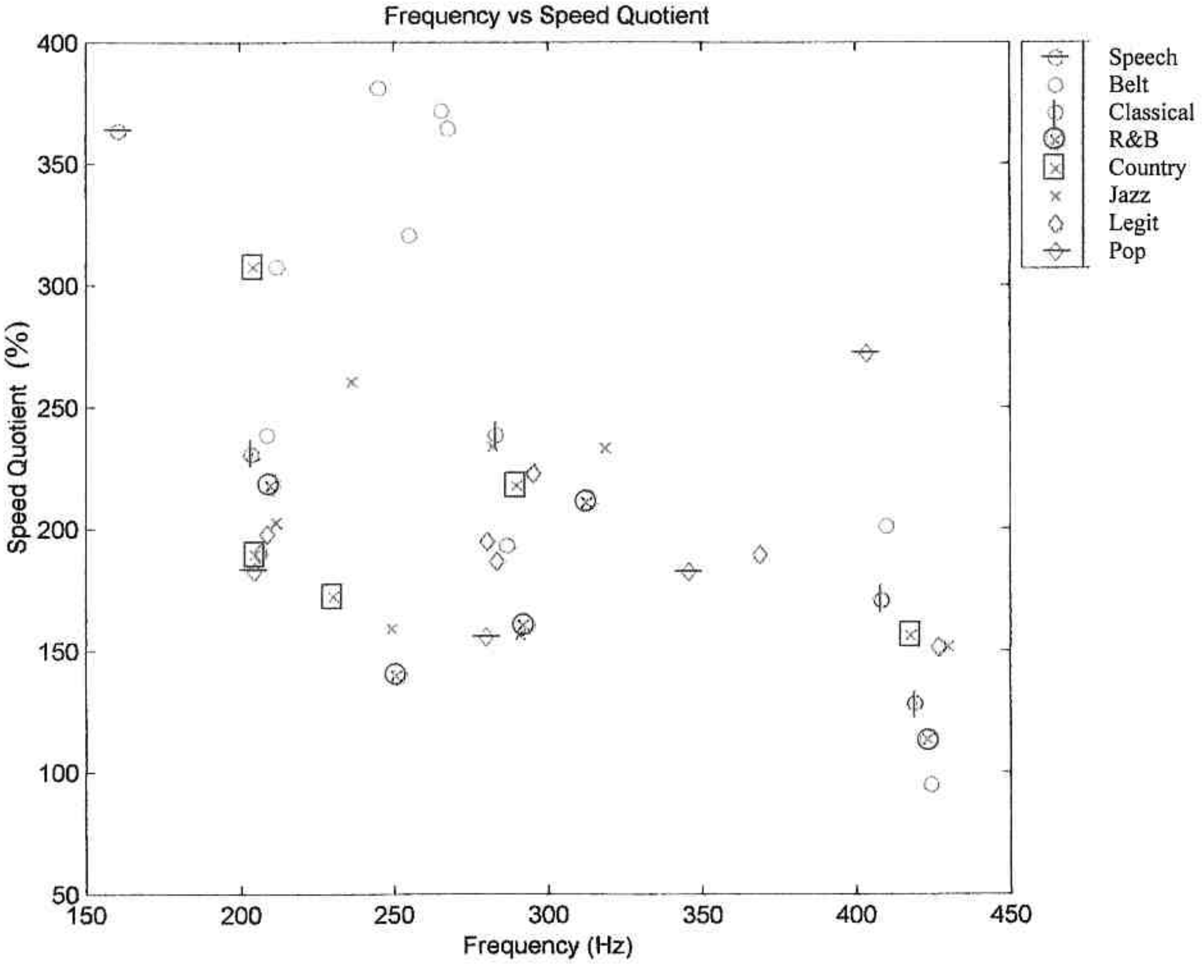
The subject's impressions were recorded in Appendix A. She advocated that Pop and Legit qualities be produced with horizontal expansion and vertical constriction (smile) Country and Jazz be produced with "puckered" lips, R&B and Belt be produced using extra width in the oral cavity: "water in the mouth" and "molar mouth". She described the "water in the mouth" posture was created by increasing space around the lower mandible as if you were holding water in your mouth. The "molar mouth" posture was created by increasing the space between the top and bottom molars. She also indicated firm lips in the Belt quality and a wide flat tongue in the R&B quality. She suggested using a longer space in the vocal tract for Classical quality with cheeks raised.

4.3.2 Laryngographic Results

4.3.2.1 Speed quotient analysis

Figure 19 shows low frequency samples clustered around 225 Hz. Belt, which is a robust voice quality, showed the highest SQ values. These SQ values were very similar in magnitude to speech; however, the speaking pitch was lower (160 Hz). For the lower frequency Belt samples, SQ of Belt quality was similar to speech in 5/6 cases; however, as the pitch increased, the SQ for Belt quality dropped substantially. On the other hand, one sample of Belt quality that was over 400 Hz showed the lowest SQ of all the qualities studied, which was consistent with my analysis of Estill's LGG samples. Those samples showed a more triangular shaped wave for Belt in the upper frequencies. Most other points for low pitch were clustered between 175 Hz and 250 Hz. In this cluster, from lowest SQ to highest SQ, were Country (2/3), Pop (1/1), Legit (2/2), and Classical (1/1). Jazz, R&B and Country were more variable with Country (2/3 low), R&B (1 low and 1 mid), and Jazz (1 low, 1 mid and 1 high). The speed quotient analysis indicates that Classical, Belt and Speech have more skew and a higher slope than the other qualities; however, Belt and Speech showed a much greater skew than Classical.

Figure 19: Frequency versus Speed Quotient graph showing seven qualities of singing and speech. Speed Quotient is related to the skewing of the Lx waveform and indicates the angle of convergence of the vocal folds.



The middle pitches were clustered around 300 Hz. Classical quality showed the highest SQ; however, this SQ was a very similar value to the low pitch for the same quality. Belt voice in the low and higher pitches dropped into the cluster. In this cluster, from lowest SQ to highest SQ, were Pop (1/1), Legit (2/3), Belt (1/1), Country (1/1), Jazz (2/3), and Classical (1/1).

The higher pitches at around 425 Hz showed a curious movement of the Pop quality upward above the remaining qualities. On the other hand, the SQ for Belt quality moved down to a magnitude of middle to low (1 mid and 1 extremely low). The samples in the higher frequency range were less clustered, and showed the following SQ from low to high. Belt (1 mid and 1 extremely low), R&B (1/1), Classical (2/2), Legit (1/1) and Jazz (1/1), Country (1/1), Pop (1/1).

Generally, the samples fell into four categories: those SQ that moved down with an increase in pitch (Belt, Classical), those SQ that moved upward with an increase in pitch (Pop), and those SQ that stayed fairly constant within a midrange with an increase in pitch ((Legit (7/7), Country (4/5; 1 outlier), Jazz (7/7), R&B (5/5)). If the Pop sample with the high SQ on the higher frequency were considered an outlier, then Pop would have been with the group of qualities that showed fairly constant SQ. A visual analysis of the SQ graph showed that the Belt voice had the highest overall SQ similar to speech, next was Classical; however, there was a substantial drop in SQ from Belt and Speech to Classical. On the low end were R&B and Pop. The rest of the qualities were in the middle.

4.3.2.2 Contact quotient analysis

The Contact Quotient graph (Figure 20) shows that the benchmark speech sound fell within the norms for speech at 42%. Most of the Belt quality samples fell above this level, however they were still within the norms under 60%. Another curiosity is the consistency through the pitch range. Most of the Belt samples hovered around between 40% and 55% (6/9 with 2 others just outside those boundaries and one outlier). Pop was another quality with consistently high contact quotient with all the samples found between 45% and 55%. Jazz was very consistently near speech with all the samples between 35% and 45%. Classical was below speech with all closed quotients between 35% and 41%. Legit was very low with closed quotients between 20% and 35% (6/7 with one outlier). R&B was also low between 30% and 40% (4/5 with one outlier). Country was the most inconsistent with a range between 25% and 50%.

4.3.2.3 Slope Analysis

A manual calculation of the ascending slope between 0-.25 of the peak, and the ascending slope .25-.75 of the peak, was used to assess the closing behaviour of the vocal folds. The calculations were plotted in MATLAB in two graphs: Frequency vs. Slope (Figure 21) and Slope vs. Initial Slope (Figure 22).

In Figure 21 all of the slope values were relatively high although the range was from 250 to 2500. Belt quality was on the lower half of the graph for 9/10 samples whereas, Classical was on the higher side for 4/ 4 samples. This indicates Classical quality had a steeper slope and more rapid closure of the vocal folds than Belt. Most of the other qualities were found on the bottom half of the graph. Country (4/5), Jazz (6/8 with 1 point midrange and 1 outlier), Legit (6/7 with 1 outlier), Pop (4/5 with one point midrange), R&B (4/5 with one outlier) all had slopes that were less steep than Classical. Speech was the other quality represented on the top half of the graph. Slopes did not appear to be linked to frequency. There also were no clear divisions between the other qualities represented on the bottom half of the graph.

Figure 20: Graph of Frequency versus Contact Quotient in percent for seven singing qualities and speech. Contact quotient is an indicator of vocal fold adduction.

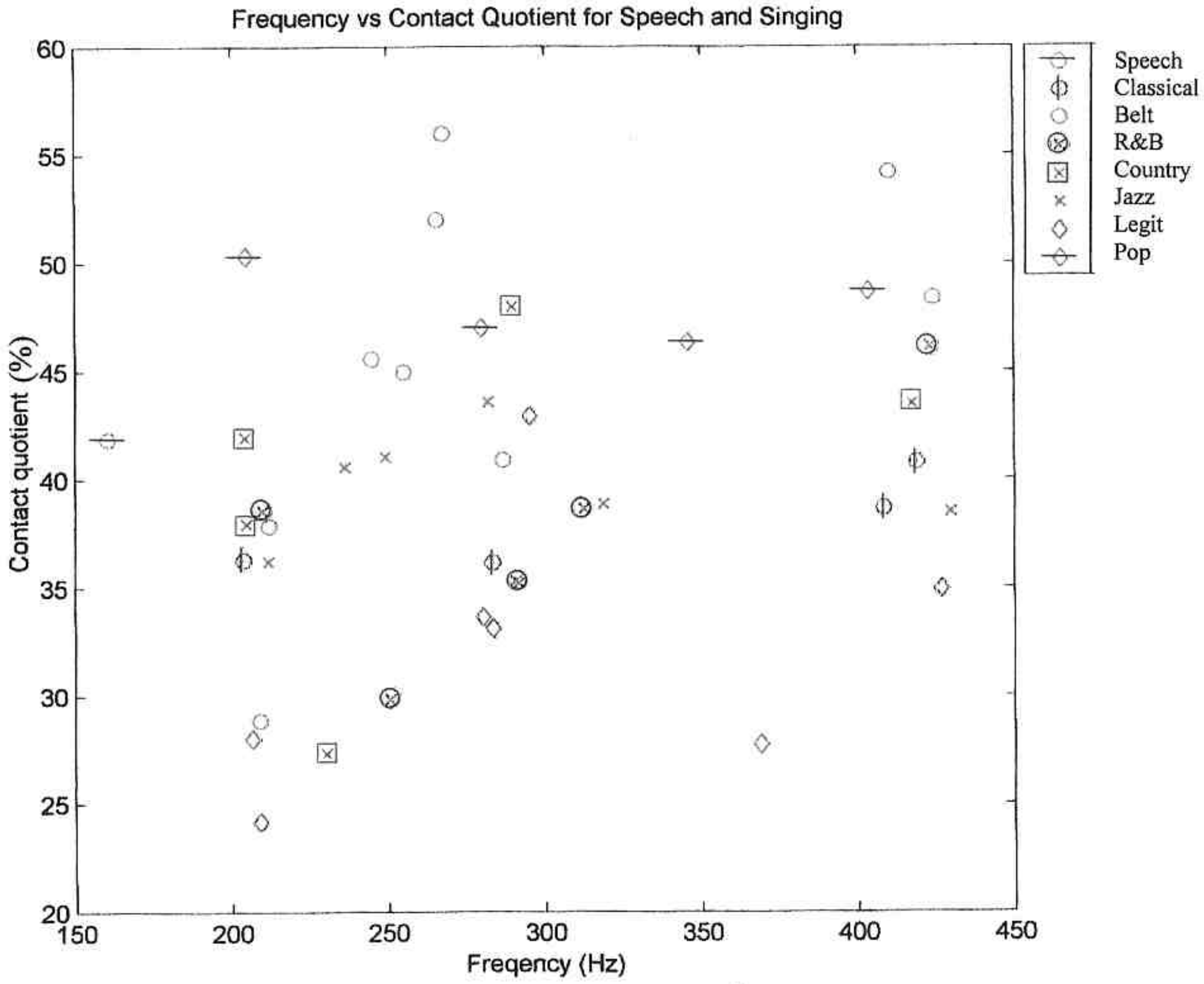
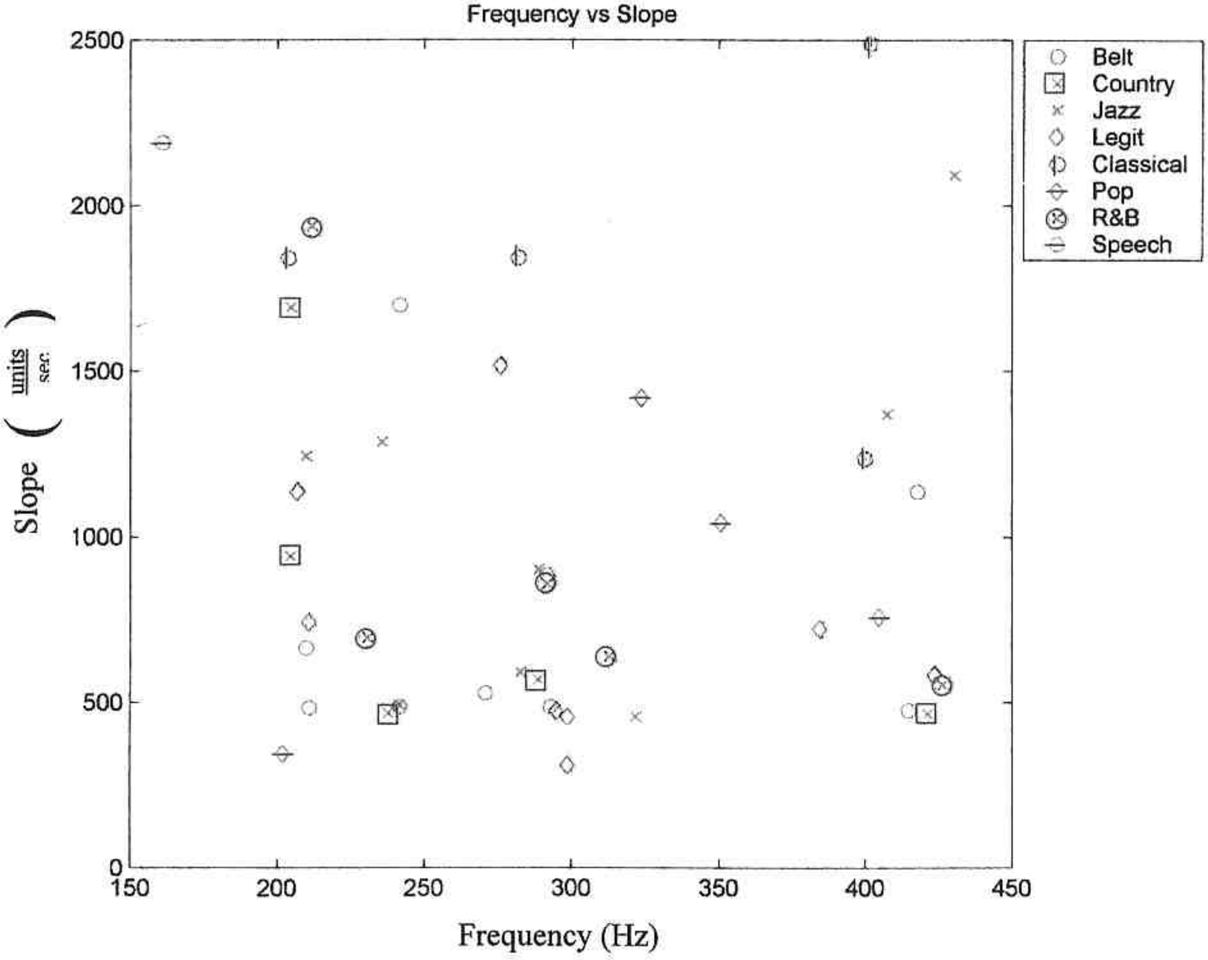


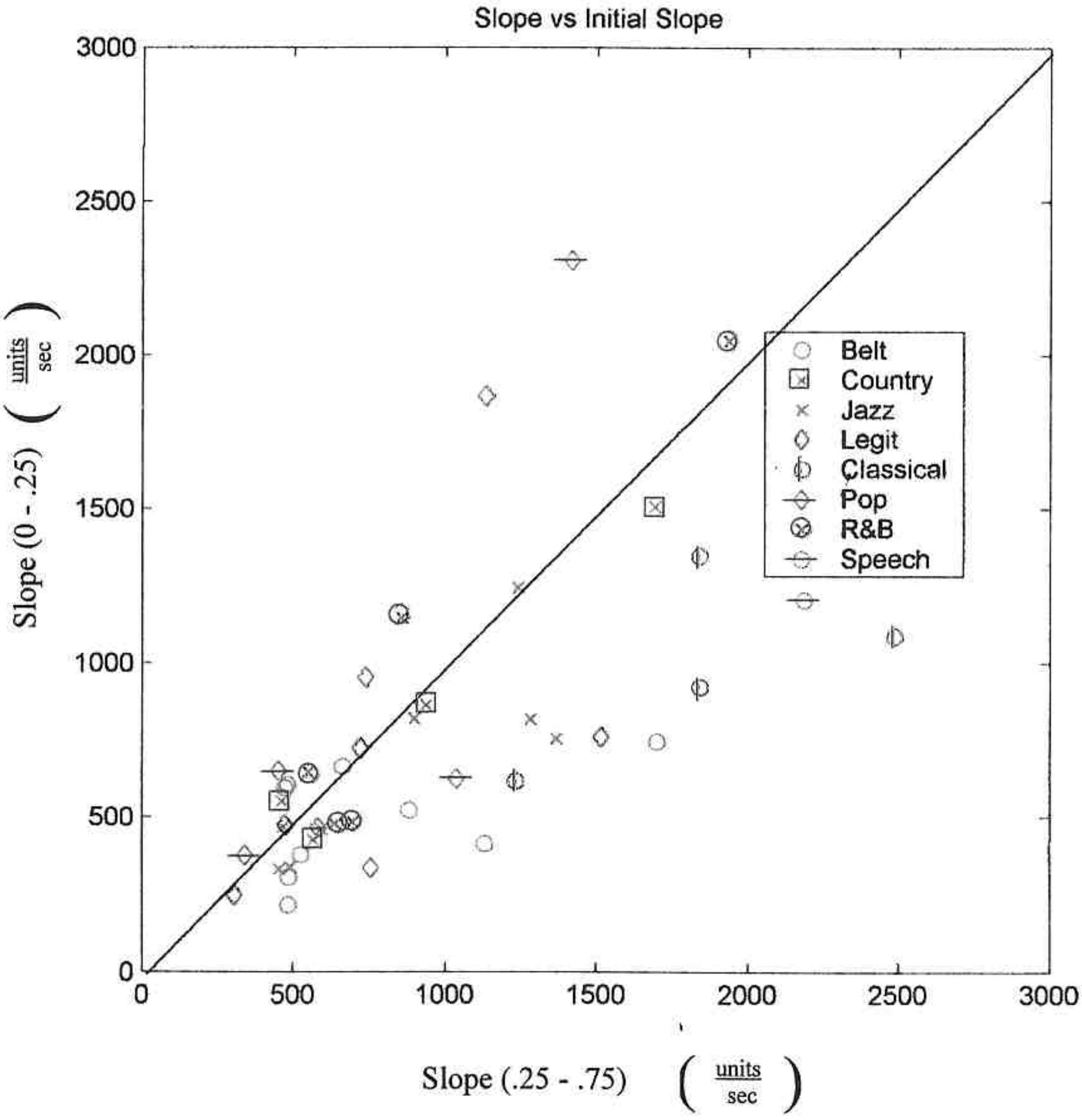
Figure 22 also showed a separation of Classical and Speech qualities from the other qualities. The qualities that showed the most difference between the slope and initial slope were Classical (4/4) and Speech (1/1). This characteristic in the Lx waveform was linked to zippering (Baken and Orlikoff, 2000). There were two outliers,

Figure 21: Graph of Frequency versus Slope for seven singing qualities and speech. Slope is an indicator of amplitude.



one in Pop and one in Legit that showed a negative difference between the slope and the initial slope. In these cases the initial slope was very steep but then it tapered off a little and the mid slope was substantially less. The slope analysis seems to indicate more vocal fold slap in Classical and Speech but also more zippering. On the other hand, the other qualities seem to show less zippering. This may mean that it is not the presence of the slap that causes a vibratory pattern with increased partials, but the absence of zippering. Further research is needed to confirm this finding. It may be possible use inverse filtering to establish the harmonic components of the various source waveforms.

Figure 22: Graph of Slope versus Initial Slope for seven singing qualities and speech. A greater slope than initial slope indicates zippering of the vocal folds.



4.3.2.4 Visual LGG Analysis

The visual analysis was done using Titze's (1990) criteria for LGG analysis cited in Baken and Orlikoff (2000). The samples were divided by quality and by frequency. The following tables 12 – 14 were compiled from the data found in Appendix A. The numbers represent the mean for the sample points in each category.

Table 12: Mean LGG Analysis Calculations for Low Frequencies around 200 Hz

Lx Characteristic	**Physiological Characteristic**	**Belt**	**Classical**	**Country**	**Jazz**	**Legit**	**Pop**	**R&B**	**Speech**
Peak Widening 1 = wide 5= narrow	Adduction	3.3	3	2.3	2.3	5	3	2	3
Peak Skewing 1=none 5=extreme	Convergence	3.5	3	2	1.7	2	3	1.5	3
Peak Shape 1=rectangular 5=triangular	Phase lag	3.5	5	4	4.7	5	4	5	4
Skirt Elevation 1=none 5=extreme	Surface bulging	2.5	1	1.7	1	1	1	1	1

Table 13: Mean LGG Analysis Calculations for Medium Frequencies around 300Hz

Lx Characteristic	**Physiological Characteristic**	**Belt**	**Classical**	**Country**	**Jazz**	**Legit**	**Pop**	**R&B**	**Speech**
Peak Widening 1 = wide 5= narrow	Adduction	2.3	3	4	3	4.3	2.7	2.5	N/A
Peak Skewing 1=none 5=extreme	Convergence	4	2	2	2	1.7	2.7	1.5	N/A
Peak Shape 1=rectangular 5=triangular	Phase lag	2.7	5	4	4.7	5	4.7	5	N/A
Skirt Elevation 1=none 5=extreme	Surface bulging	3.3	2	1	1	1.3	1.3	1	N/A

Table 14: Mean LGG Analysis Calculations for Medium High Frequencies around 400 Hz

Lx Characteristic	Physiological Characteristic	Belt	Classical	Country	Jazz	Legit	Pop	R&B	Speech
Peak Widening 1 = wide 5= narrow	Adduction	3	2.5	5	2	4.5	3	2	N/A
Peak Skewing 1=none 5=extreme	Convergence	2.5	2	3	1.5	2	3	1	N/A
Peak Shape 1=rectangular 5=triangular	Phase lag	4.5	4.5	4	5	4	3	5	N/A
Skirt Elevation 1=none 5=extreme	Surface bulging	2	1.5	5	1.5	1.5	4	1	N/A

The results were arranged in sequences of voice qualities to correlate Lx signal to vocal fold production:

Low Frequency

Peak widening from wide to narrow: R&B, Country/Jazz, Speech/Pop/Classical, Belt, Legit.

Peak skewing from none to extreme: R&B, Jazz, Legit/Country, Speech/Pop/Classical, Belt.

Shape from rectangular to triangular: Belt, Speech/Pop/Country, Jazz, Classical/R&B/Legit.

Skirt elevation from none to extreme: Speech/Pop/Jazz/Classical/R&B, Country, Belt

Medium Frequency

Peak widening from wide to narrow: Belt, R&B, Pop, Jazz/Classical, Country, Legit.

Peak skewing from none to extreme: R&B, Legit, Classical/Country/Jazz, Pop, Belt.

Peak shape from rectangular to triangular: Belt, Country, Jazz/Pop, R&B, Legit and Classical.

Skirt elevation from none to extreme: Country/R&B/Jazz, Legit/Pop, Belt/Classical.

Medium High Frequency

Peak widening from wide to narrow: Jazz/R&B, Classical, Pop, Belt, Legit, Country.

Peak skewing from none to extreme: R&B, Jazz, Legit/Classical, Belt, Country/Pop.

Peak shape from rectangular to triangular: Pop, Legit/Country, Belt/Classical, Jazz/R&B.

Skirt elevation from none to extreme: R&B, Jazz/Legit/Classical, Belt, Pop, Country.

4.4 Discussion

There was evidence to suggest that F1 was tracking P1 in all sung samples (Table 10). Speech was the only quality to show poor tracking. The results of this research do not support F1 tracking P2 for Belt quality as seen by Bevan, 1989; Sundberg, 1993; Schutte and Miller, 1993. However, the results are consistent with Sundberg's (1987) observation that female voice uses fundamental tracking more than the male counterpart to increase audibility. Qualities with the smallest P2 values were Classical (mean = 9.0%) and R&B (mean = 9.0%). Qualities with medium energy in P2 were Legit (mean = 12.0%), Pop (mean = 12.0%) and Jazz (mean = 13.0%). Qualities with strong energy in P2 were Country (mean = 21.0%) and Belt (mean = 27.0%). The pitch of P2 is an octave above the fundamental frequency; therefore, the proportion of P2 in the spectrum affects the perceived "brightness" of the quality. The "darker" qualities, Classical and R&B, both had very little energy in P2, whereas the brighter qualities, Country and Belt, had high energy in P2. There was no evidence in this study to conclude that F1 tracks P2 in any of the voice qualities produced by this subject and analysed in this thesis.

However, the variable in P2 energy suggests that the bandwidth of the formants may be different in some singing qualities than others. Bandwidth may be a consideration in the high P2 energy in some qualities. Bandwidth may be responsible for P2 decreasing in most qualities as pitch increased. The muscular activity in the pharynx is known to affect the bandwidth of the formants. Laver (1980: 62) states tensed walls of the pharynx damp the sound wave less than relaxed walls, and the resultant decrease in damping is reflected in narrower formant bandwidths. Tense pharyngeal wall are associated with some vocal techniques. For example, Classical vocal technique requires singers to make a longitudinal stretch and slight "yawny" space in the pharynx. Classical singers must begin the stretch before the onset of singing, and maintain it for the duration of the sung line. Perhaps this stretch firms the pharyngeal walls, and helps narrow the bandwidth of the formant for better resonance tuning. Higher P2 energy levels in Legit suggest a wider bandwidth, which is consistent with Legit voice production. Legit quality requires less longitudinal stretch, and less advance tongue root than classical singing and may result in reduced tension in the walls of pharynx. More research is needed to clarify the linkage between wall tension and bandwidth in singing voice quality. Another consideration is that formants in close proximity can interact. A wide bandwidth can affect the degree of interaction between formants. The influence of the bandwidth on the location of partials and formants is a question for further study. Wider bandwidths are usually more attenuated than peaky bandwidths.

Vowel formant location in speech was a little lower in the subject than the Peterson and Barney mean value of F2 (2790Hz) for the average woman (Ball, 1993:146). The auditory analysis noted that the subject's larynx was slightly low. Low laryngeal position suggests lower formants. Lower than average formant locations are also found in subjects with longer vocal tracts. The vowel formant for Pop was slightly higher than speech. The auditory analysis suggested that the subject's larynx was raised. A raised laryngeal position will shorten the vocal tract and raise formants. The subject also noted a smile mouth position for this quality. Both RL and smile can shorten the vocal tract and raise formant values.

Energy was found at approximately 9000 Hz in Belt and Country. This energy may be related to second formant of the singers ring cavity; however, in both cases the frequency of the energy is lower than expected. For example, the resonating cavity that produces the singer's ring had a predicted F1 calculated at 3500 Hz. The predicted F2 of the singer's ring tube should be about 10.5 kHz (4.2.4 Measurements). A speech formant was seen at 8900 Hz; therefore, it is possible that energy from the vocal tract formant (F8), and second singer's ring formant (F2), are interacting. This interaction may be enhanced because of the wide bandwidths seen in these qualities. The DFT showed some energy in the 10.5kHz region but it was generally lower than 1.0 %.

Both Pop and Country showed some energy in the upper partials. Country showed moderate energy. Pop, on the other hand, showed weak upper partial energy. The auditory analyses observed that Country samples used MV or HV, whereas 4/5 Pop samples used either WV or BV. These results suggest that BV and WV may reduce the upper partials. Laver (1980) suggested that overall energy in the spectrum is reduced in Breathy Voice. Lx waveform shows that Country and Pop have a more gradual slope than other qualities in 3/4 samples; however, Country showed an initial slope that was very close to the slope, which suggested that although the speed of the vocal fold closure was not as rapid as some of the other qualities, there was no zippering indicated. On the other hand, initial slope values for Pop samples differed from slope values in 4/5 cases. These results suggest more zippering in Pop quality than Country. The presence of zippering may contribute to the attenuation of upper partials. Zippering was also suggested by slope analysis for Classical quality where initial slope was consistently less than the midslope. Classical quality showed some of the lowest energy in the upper partials. In addition, Pop, Country and Belt all showed relatively high CQ and showed some energy in upper partials, whereas Classical, R&B, and Legit showed relatively low CQ and little energy in upper partials. These results suggest that qualities with weak upper partials may have less adductive forces.

Titze (1994) suggested that larger pharyngeal spaces enhance lower partials and attenuate higher partials, whereas constricted spaces enhance higher partials. Opening side branches of the vocal tracts, such as the piriform sinuses, can remove upper partials from the spectrum (Titze, 1997). Generally, the qualities with less constriction also are perceived darker and purer than those with constriction. R&B, Classical and Jazz sound darker than the other qualities studied. The acoustic analyses found lower partials in R&B, Classical and Jazz strong, and upper partials weak. These results suggest R&B, Classical and Jazz qualities are produced with a spacious pharynx. X-ray fluoroscopy and MRI comparisons of Belt and Classical showed Classical quality as more spacious than Belt, with –ATR and low laryngeal position. These results are consistent with previous studies by Sundberg (1993) and Estill (1988).

A lower laryngeal position is one posture that can increase the pharyngeal space. The auditory analysis detected a low larynx in Classical, Jazz and R&B. A lower larynx also suggests lower formants. Classical and R&B also had the lowest F2. Alternatively, laryngeal elevation can cause formants to rise. The subject described a high laryngeal position in Country. She also felt the larynx rising with pitch; however, the auditory analysis found that laryngeal rise was not linked with pitch in Country. Laryngeal position in Country was variable from sl. LL to sl. RL. The auditory analysis found that the lowest pitch sample had a sl. RL, but the acoustic analyses showed a low F2 frequency; on the other hand, the auditory analysis of the high pitch sample showed neutral larynx, but the acoustic analysis showed a high F2 value. The results of this research confer with Stone et al.'s (1999) findings, which indicated that the laryngeal position was variable for Country and not necessarily linked to pitch. The curious movement of F1 and F2 in Country could be a result of pharyngeal constriction. Esling (1994) showed that pharyngeal constriction should bring F1 and F2 closer together. The auditory analyses also found F2 low with the larynx raised for Belt quality in the majority of cases (5/7). Belt showed F2 below 2000 Hz. Again, pharyngeal constriction may be affecting the formants. There was very little evidence of pharyngeal constriction in the auditory analysis for Belt; however, the saggital view of the larynx with x-ray fluoroscopy showed RL, an aryepiglottal constriction and –ATR. (Appendix B) The coronal view showed very little pharyngeal constriction, which corresponds to the subject's sensation of "widening".

A spacious pharynx has also been linked to BV. Esling (2003) has shown that BV requires an open epilaryngeal space, whereas the WV requires a sphinctered epilaryngeal space. BV was found in Jazz, Classical, Legit and R&B. All these qualities also showed a low laryngeal position in the auditory analyses. In addition, Classical, R&B and Legit pedagogy advocates a more open-throated voice production (gola aperta) and a low larynx. Jazz pedagogy also advocates a low larynx (Popeil, 1996). The auditory analyses also found either WV or BV in all qualities. Speech showed WV, no constriction and a LL. The speech results do not support the theory that pharyngeal constriction is linked to WV. The auditory analyses results were consistent with the inharmonic activity found in the spectrograms. Spectrograms showed inharmonic activity for WV between high frequency partials. On the other hand, spectrograms showed inharmonic activity for BV between the midrange partials. Since enhancement of lower frequencies is a characteristic of a spacious pharynx, inharmonic energy in partials in the midrange suggests less pharyngeal constriction than inharmonic activity in the upper frequency range. Esling (1994) also found breathy voice to have a smaller contact quotient than modal voice. This may be why R&B and Legit had lower CQ values however, Pop and Jazz also judged as breathy or whispering in the auditory analysis and the aperiodic analysis showed a relatively high closed quotient in this analysis.

Country and Belt voice showed the least BV or WV. Belt also had the highest CQ ratio. Country had a variable CQ. The CQ has been linked to vocal fold adduction. Low CQ results indicate BV or WV, whereas high CQ ratios suggest HV. HV was found in the auditory analyses of Belt and Country. HV is more tolerated in the esthetics of the belt and country musical styles than in the classical musical style. HWV was heard on higher pitches in Jazz and Pop. These results suggest a pharyngeal constriction. The Pop sample showing HWV corresponded with a sl. RL and sl. pharyngealization. Esling et al. (1994) also suggested the raised larynx and pharyngealization represent the same articulatory setting. Although all Lx samples analysed with CQ were still in the normal range between 40% and 60%, the samples that had HV and HWV showed CQ values above speech except for one sample of Jazz quality. Belt and Pop quality showed the highest CQ in the LGG analyses. Belt and Pop also showed energy in the upper partials. The presence of W suggests a constriction in the pharynx and constrictions enhance higher frequencies. The presence of H suggests firm adduction. HV was only found in a few samples; however, it may be more of a hazard in Belt and Pop than other more spacious qualities such as Classical, R& B and Jazz.

Qualities known for their volume are Classical, Belt and Country. The chapter on related work explained that modal voice was more efficient than the other phonation types mentioned. The auditory analysis found MV more prevalent in Classical, Belt and Country than the other qualities. In addition, the speed quotient analysis indicated that Classical, Belt and Speech have more skew and a higher slope than the other qualities. However, Belt and Speech showed a much greater skew than Classical. Titze related the slope in an Lx pulse to the energy in the acoustic signal (Titze, 1989, Titze 1990). He showed that a steeper slope is usually associated with louder sounds. Louder vocal sounds are generally produced with an increase in subglottal pressure. An increase in subglottal pressure is usually associated with higher adduction rates (Titze, 1994; Estill, 1990).

The skew in the Lx waveform has also been related to glottal convergence of the vocal folds. Estill's (1988: 239) vocal fold plane and Titze's (1990) angle of convergence may be linked. Convergence is the angle of tapering of the vocal folds toward the medial edges whereas; the vocal fold plane is the angle of the vocal folds away from the horizontal. However, establishing this link is a question for further research. In speech, this convergent shape produces a slap and roll effect. The vocal folds slap shut and roll open. The slap phase of the vibration introduces high partials into the voice quality. The samples of Speech and Belt showed a high speed quotient, and as a result may have a greater slap, more adduction, and a greater angle of convergence. This would be consistent with the acoustic analysis, which showed more high frequencies in Belt than the other qualities. Different phonation types have a characteristic skew. The ranking of phonation types according to decreasing skew is creak, modal, falsetto, breathy voice (Childers et al., 1991: 2397). In addition, although peak skewing of low frequency Lx samples did not correspond closely to voice quality, peak skewing of medium frequency samples did correspond. In medium and high frequency Lx samples, the brassier qualities all show a greater skew than the purer qualities. The Lx waveform also showed all samples had a certain degree of triangularity at higher frequencies. This result suggests vocal folds thickness is related to register transition. The shape of the Lx waveform suggests the vocal folds in Classical and Legit qualities stay in a wedge shape throughout the range, whereas the vocal folds in Belt quality change shape as the frequency ascends the range from rectangular to triangular. Belt seems to stay in the rectangular configuration through the low and medium pitches and only move to a triangular shape in the higher pitches. There was a decrease in SQ from low pitch to high pitch, which may also signal a change in shape of the vocal folds. Baken and Orlikoff (2000) also stated that the CQ for speech in women tends to increase with pitch. Legit, Classical and R&B qualities conferred with this speech trend. All other qualities did not. The auditory analysis also detected vowel modification to [e] on the highest pitch (R21). Vowel modification from [i] to [e] will cause F2 to drop. Classical also showed vowel modification from [i] to [e] as pitch rose, which is consistent with Classical vocal technique. Generally, there was little vowel modification with ascending pitch in all qualities except Classical, which may be due to the small pitch range chosen for analysis. The pitches were all within one octave. Further research should look at pitches higher than D5. It is interesting that Belt, Legit and Country qualities used the least vowel modification. Music theatre and Country singing place more importance on the intelligibility of the words than Classical singing. Vowel modification affects the intelligibility of the vowels. One reason could be that Classical singing was developed for volume. Physical postures needed to create volume in Classical singing, such as open jaw and +ATR, may reduce intelligibility of the vowels.

Titze (1990) indicated the shape from rectangular to triangular of the Lx waveform suggests associated phase lag and the shape of the medial edges of the vocal folds. A visual analysis of the Lx waveform using Titze's critieria suggests that the vocal fold shape of Belt, Speech, Pop, and Country is more rectangular than the vocal fold shape of Classical, R&B and Legit. The brassier qualities all have an Lx waveform that suggests a rectangular medial edge of the vocal fold. Alternatively, the purer qualities all have an Lx waveform that suggests a triangular medial edge of the vocal fold. These results indicate that the shape of the medial edges of the vocal folds plays a very important part in the perception of voice quality, however further research is needed.

Baken and Orlikoff (2000) suggest that contact quotient of the Lx waveform is related to contact area and adduction of the vocal folds. (Scherer et al., 1988) suggests that the contact area can indicate vocal fold thickness or shape of the medial edges of the vocal folds. The CQ analysis suggests the following level of adduction from most to least:

Belt, Pop, Speech/Jazz, Classical/R&B, and Legit. Country was variable. It has been found that different phonation types have characteristic LGG pulse width. Childers et al. (1991) measured open quotient pulse width and found the ranking of phonation type according to decreasing contact width was creak, modal, falsetto, breathy voice. (Childers et al., 1991: 2397) Falsetto and breathy voice appeared to have no closed phase.

Peak widening of the Lx waveform also suggests adduction of the vocal folds. In low pitch Lx samples suggests that the forces of adduction in the vocal folds from firm to weak may also be as follows: R&B, Country/Jazz, Speech /Pop/ Classical, Belt, Legit. Adductory forces may not be a good indicator of voice quality in singing; however, they may play a part in voice health. Here the visual Lx analysis did not agree with the auditory analysis. R&B showed firm adduction in the Lx visual analysis yet the all samples showed a phonation type of BV or WV.

Skirt elevation in the Lx waveform suggests surface bulging in the TA muscle (Titze, 1990). Skirt elevation was found in the Lx waveforms of Belt and Country at low pitch. Estill (1988) found a very high level of activity in the TA muscles for belting. Skirt elevation in Belt was found to be greater than Pop, and skirt elevation in Classical was found to be greater than Legit at medium frequencies. These results correspond to what is known about TA activity in the various qualities. There was no evidence of skirt elevation in Country, R&B and Jazz medium frequency Lx samples.

. Nasality was heard in Country and Classical. Country showed more nasality than classical. Nasality was not present on the running sample of Classical. Nasality can be used to encourage vocal fold vibration (Morrison, 1986). It is unclear whether nasality has an affect on the vocal fold vibration in these qualities. Nasality is also known to influence the acoustic pattern by acting as a negative filter, thus removing upper partial activity (Fry, 1979: 118); however, Country quality was nasal sounding, but rich in upper partials. It is unclear whether the nasality heard in the auditory analysis is real nasality caused by an open nasal port or false nasality as a result of a pharyngeal constriction. However, there were two samples that sounded constricted, but not nasal. Perhaps the degree of constriction is linked to the perception of nasality. More research is needed to clarify this acoustic phenomenon.

Some of the subject's sensations can be described in phonetic terms. For example, Pop uses the labial setting of horizontal expansion and vertical constriction. Country and Jazz use labial protrusion. Laver's mandibular settings don't seem sufficient to describe the subject's water in the mouth and molar mouth, although they may fit in subcategories under open jaw position. The subject's description of cheeks in Classical singing suggests engaging of the zygomatic muscle and the levator (Quadratus) labii superior muscle to lift the cheeks. This posture creates space around the maxilla bone (zygomatic process), which is a well-documented singing posture in the pedagogical literature (Miller, 1986). Raising the cheeks is usually accompanied by a relaxing of the lower half of the face in order to create a vertical space in the oral cavity and an arching of the velum. The subject describes this posture as "ick face". Laver's terminology does not seem to have a way to describe this posture or the voice quality associated with it.

4.5 Summary

The original data show that the glottal source and vocal tract are configured in a specific way for each singing quality. This produces a specific blueprint. The following summaries show the [i] for each of the seven singing qualities studied in the thesis.

4.5.1 Speech

The pitch of the speech sample was 161 Hz. This pitch is slightly lower than the average frequency for female speech (235 Hz). The speech sample showed no F1/P1 formant tracking. The location of the first formant (F1) fell between first and second partials (P1 and P2). The energy in P2 was 78.0% of P1. Partial energies around the formants were: F1 (100.0%), F2 (10.0%), F3 (7.0%), F4-F5 (2.0%). F1 and F2 showed considerable spread F1 (305 Hz), F2 (2683 Hz). F3 (3246 Hz) was near F3 for the average female. These results are consistent with related work. There was a gap of generally low energy between 3246 Hz and 5339 Hz; therefore singer's formant energy was not present. Upper partial energy was also weak. The strongest formant in the upper partial range 5000 Hz – 10000 Hz was at 8280 Hz. The vowel was a pure [i] with no nasality noted in the auditory analysis. The phonation type was slightly whispery voice, and as a result, the inharmonic activity was high. The laryngeal position however, was slightly low, and there was no pharyngeal constriction noted in the auditory analysis.

The laryngograph showed a large speed quotient and large slope. The results indicate that the vocal folds were contacting very fast and taking a longer time to open. The visual analysis showed a slight skew in the Lx waveform with a triangular shape indicating a slap and roll mechanism for the vocal fold closure. The Lx waveform had a triangular shape in most samples, which may indicate that this quality used Testo Voice phonation. The contact quotient was moderately low (42%), which would indicate the adduction was moderate. There was a difference between initial slope (0-.25) and slope (.25-.75), which suggested zippering in the horizontal plane.

4.5.2 Pop

The mean pitch for the Pop samples was 349 Hz. The Pop samples showed great strength in P1 and generally low energy in P2 (4/5 cases). F2 was a little higher than expected. The LPC showed the mean F2 frequency as 2750 Hz. There was a medium amount of energy in the vowel formant. The harmonic energy showed that the partial nearest to F2 was 4.2 % of P1, and the summed energy was medium. Perhaps the moderate energy in the vowel formant increases intelligibility for this quality. F3 was an approximately the same as for speech (F3 Spk – 3246 Hz; F3 Pop – 3147 Hz) and energy for F3 was similar to speech (6.4%), but F4 and F5 were very low in energy F4(0.6%) and F5 (0.6%). The low energy in the upper formants F4 and F5 supports Schutte and Miller's (1993) theory that Pop has a less rich spectrum due to the falsetto vocal fold adjustment (TV?). Summed energy plot showed singer's formant energy was variable. There was a wide band of energy between 6246 Hz and 9360 Hz. The partials were barely visible. The energy in this area ranged from 1.1%-5.6%. This energy was relatively high for these upper frequencies. Mostly inharmonic activity was recorded in this range. The auditory analysis showed that the tongue was forward. There was nasality in the running samples. The phonation type was mostly modal voice with an assortment of various other types: whispery voice, whispery creaky voice, harsh whispery voice, harsh creaky voice, and breathy voice. There was a medium amount of inharmonic activity; however, it was less than her spoken samples (Pop - 3.2, Spk - 4). Auditory analysis found a laryngeal position of mostly neutral; however, raised larynx was recorded on the two higher steady state pitches E4 and B4. Also, there was some pharyngeal constriction noted on the auditory analysis; more was heard on the sustained vowel samples than in the running samples.

The LGG for Pop showed a generally low speed quotient that seemed to increase with frequency. The slope was less steep than speech, but the visual analysis still found a slight skew to most Lx samples. This skew indicates a slap and roll mechanism for the vocal folds; however with a slower contacting phase than speech. The Lx waveform had a triangular shape in most samples, which may indicate that this quality uses a Testo vocal fold adjustment. The contact quotient was moderate (approximately 50%), which would indicate the adduction was moderate. There was little difference between initial slope (0-.25) and slope (.25-.75), suggesting parallel closure of the vocal folds in the horizontal plane. Pop quality showed one outlier on the graph slope vs. initial slope (Figure 22) indicating the initial slope was greater than the slope.

4.5.3 Country

The mean pitch for the Country samples was 270 Hz. Country had high energy in P2 (21% of P1), but the fundamental tracking of F1 with P1 was very good. The high energy in P2 may contribute to the brighter quality. F2 placement was similar to speech (Speech [i] 2683 Hz, Country [i] 2631 Hz). However, F2 was found to vary with pitch.

The auditory analysis noted laryngeal movement, but it wasn't linked to pitch. The laryngeal position varied from slightly low to slightly high in the auditory analysis, although generally it was neutral to high. The subject commented she was trying to obtain a high position. There was some vowel modification noted on the auditory analysis of higher pitches, where the vowel went from [i] to [e], which could also affect the F2 location. The energy in F2 was medium (partial energy 8.0% of P1; summed energy medium). The LPC showed mean formant peaks at 3473 Hz and 4378 Hz. These were considered to be the merged peaks of F3-F5. The energy in this frequency range was high; the partial energies from the DFT were F3 (8.8%), F4 (7.8%), and F5 (4.0%). The summed energy was also high. The predominant mean upper formants were at 6257 Hz (3.4%), 7410 Hz (2.7%) and 9146 Hz (2.3%). The formant at 9146 Hz may represent a formant cluster. This is relatively high energy for these upper frequencies. While the energy in 3000-5000 Hz range may make this quality perceptually louder, the strength of the upper partials can make a rougher timbre than some of the other qualities. There was nasality in most of the samples. The subject also noted nasality in the quality, and the sensation of a forward tongue. The auditory analysis noted that modal voice was the most common phonation type in this quality; however, there was also evidence of harsh voice and whispery voice. One of the samples (C11) containing harsh voice was repeated on the subject's request. Inharmonic activity was low in this quality (2.2 out of 5). Pharyngeal constriction varied from no constriction to slight constriction. The subject indicated pharyngeal constriction and labial protrusion were characteristics of this quality.

The LGG for Country showed a low-moderate speed quotient that seemed to increase with frequency. The slope was variable. The visual analysis found the Lx wave fairly wide with very little skew indicating less "slap" of the vocal folds in their closing phase, and more roll open/roll closed. This should decrease the upper partials in the source wave; however this quality showed rich upper partials. The Lx waveform had a triangular shape in most samples, which may indicate that this quality uses a Testo vocal fold adjustment. The contact quotient was moderate (approx 45%), which would indicate the adduction was moderate. There was not much difference in the initial slope (0-.25) to the mid slope (.25-.75), which suggested parallel closure of the vocal folds in the horizontal plane. There was a large degree of nasality in this quality.

4.5.4 Jazz

The mean frequency for Jazz samples was 288 Hz. Jazz showed good tracking of P1 with F1, and generally (3/4 cases) a weak P2 (mean 6%). This may indicate a finely tuned F1 formant. The mean F2 placement was a little lower than speech (Jazz – 2525 Hz; Speech – 2683 Hz). The energy in the F2 formant was medium in strength (5.3 %). Mean formant locations were also found at 2883 Hz 3803 Hz and 5396 Hz. F3 was a bit low (Jazz 2883 Hz; Speech 3246 Hz). This may be due to a lower larynx or labial protrusion. Energy around the singer's formant (F3,F4,F5) higher and summed energy in SF was low-moderate; however, there was more energy than for speech (Jazz – F4 (3.0%); Speech – F4 (2.0%). The mean F5 is in a similar location to speech, but has very low energy (Jazz – 5396 Hz (0.3%); Speech – 5339 Hz (2.0%)). Energy in the upper frequencies was also very low. The vowel remained pure throughout the samples, however there was nasality noted in the auditory analysis for the running samples. The subject noted a low tongue position that was not observed in the auditory analysis. The phonation type was mostly BV or WV. HV was heard on the upper pitch - B4. There was a medium amount of inharmonic activity (2.5) and a generally low laryngeal position noted in the auditory analysis, which is consistent with the subject's description of a low larynx and labial protrusion. There was no pharyngeal constriction heard in this quality.

The LGG for Jazz showed a moderate speed quotient. The slope was variable. The visual analysis found the Lx wave showed moderate peak width with very little skew indicating less "slap" of the vocal folds in their closing phase, and more roll open/roll closed. The Lx waveform had a triangular shape in most samples, which may indicate that this quality uses a Testo vocal fold adjustment. The contact quotient was moderate (approx 40%), which would indicate the adduction was moderate. There little difference between initial slope (0-.25) and slope (.25-.75), which suggested parallel closure of the vocal folds in the horizontal plane. The triangularity and small degree of skew indicate that there may be fewer upper partials produced in the source wave. The Jazz LGG data was similar to the Country LGG data yet the acoustic output was very different for these qualities. The difference in the acoustic output was attributed to the vocal tract configuration. The auditory analysis heard consistently no pharyngeal constriction as part of this quality. High partials introduced at the source, may removed by the resonant properties of the vocal tract.

4.5.5 R&B

The mean frequency for R&B samples was 298 Hz. R&B showed good tracking of P1 with F1, and generally (4/5 cases) a weak P2 (mean 6% for those 4 samples. The LPC showed the mean F2 frequency as 2152 Hz. This low F2 value could be due to a lower laryngeal position. The energy in this formant was medium (4.0% in mean of 4/5 cases). Mean formants were also found at 3006 Hz and 3961 Hz. F3 was a little lower than for speech (R&B – 3006 Hz; Speech – 3246 Hz) and had higher mean energy than F2 (7.6%). Singer's formant (F3,F4,F5) showed moderate energy in F4 (3.2%) and moderate summed energy, however above F4 the energy was low F5 (1.4%) and F4 (0.4%). Upper partial energy was very low (1.0%). This is audibly the darkest quality. This may be due to the weak P2, the lower F2 and F3, the weak upper partial activity. The auditory analysis showed vowel modification from [i] to [e] with ascending pitch in sustained vowels samples, and no nasality. The phonation type was either whispery voice or breathy voice dependent on the sample. There was more incidence of whispery voice on the sustained vowels, whereas in the running samples there was more breathy voice. The inharmonic activity was shown as high (4) in the spectrograms. The laryngeal position was generally low, although on the sustained vowels samples the larynx rose with the ascending pitch. The auditory analysis agrees with the subject's description of the laryngeal movement from low to neutral.

The LGG for R&B showed a low-moderate speed quotient. The slope was relatively low with one outlier. The visual analysis found the Lx wave showed moderate peak width with very little skew indicating less "slap" of the vocal folds in their closing phase, and more roll open/roll closed. The Lx waveform had a triangular shape with no skirt elevation, which may indicate that this quality uses a Testo vocal fold adjustment. The contact quotient was lower than for speech (30-40% with one outlier), which would indicate weak adduction. There was little difference between initial slope (0-.25) and mid slope (.25-.75), which suggested parallel closure of the vocal folds in the horizontal plane. The weak adduction, and triangularity of the waveform indicates that the vibratory pattern may not introduce many high partials. BV and WV suggest weak adduction and were present in all R&B samples.

4.5.6 Classical

In Classical quality there was good tracking of P1 with F1. P2 energy was generally weak (5% for 3/4 cases). This indicates a narrow F1 bandwidth. The mean F2 was 2187 Hz, and the mean F3 was 3172 Hz. F2 is low compared to speech; however, F3 is similar to speech. The auditory analysis showed the steady state vowels move from [i] to [e] with ascending pitch. The [e] vowel had a lower second formant. A low F2 was attributed to the vowel modification or a lower larynx. Energy in F2 was medium (4.5%) in the associated partial from the DFT data; however the summed energy plot showed high energy. There seemed to be a number of partials with substantial energy between the fundamental and 3000 Hz. The energy level then dropped dramatically: F1 [P1 (100.0%), P2 (5.0%)], F2 (4.5%), F3 (3.5%), F4 (2.3%), F5 (0.8%) and F6 (0.3%). Very little energy was shown in the DFT above F7. Even formant locations above F4 are probably not significant because the energy was so low. There was nasality heard in some samples in the auditory analysis. Nasality is not advocated in Classical vocal technique. The phonation type was mostly modal voice. Slight BV, HV and WV were noted; however, the inharmonic activity was generally low (1.6). The laryngeal position was low, pharyngeal constriction was minimal. The subject noted that for this quality she used "fish face" and a raised arched velum to create length in the oral cavity.

The LGG for Classical showed a low-moderate speed quotient. The slope was steep. The visual analysis found the Lx wave showed moderate peak width with a slight skew indicating a slap of the vocal folds in their closing phase, and roll open. The Lx waveform had a triangular shape with little skirt elevation. The contact quotient was lower than speech (35-41%), which would indicate weak-moderate adduction. There was a difference between initial slope (0-.25) and slope (.25-.75), which suggested a zippering of the vocal folds in the horizontal plane.

4.5.7 Legit

Legit quality showed good tracking of P1 with F1 and a slightly higher energy in P2 than with Classical quality (mean 6.0% in 5/7 cases). The lower pitches had much more energy in P2 than corresponding pitches in Classical quality. This is also consistent with the pedagogy, which requires Legit to be more speech-like in the lower register. Mean F2 was located similar to Speech F2 (Legit – 2571 Hz; Speech – 2683 Hz). F2 energy was 8.7% which was the second highest vowel formant energy recorded. Belt had the highest vowel formant energy. Summed energy was medium. There seemed to be less energy between formant peaks in Q1 (P2-3000Hz) in Legit than in the Classical quality. The energy in the higher formants is attenuated, but not quite as severely as with Classical: F3 (6.0%), F4 (2.6%). F5 (1.7%) and F6 (0.9%). Therefore, the formant placements may be important for this quality. Formant placements were found at 3232

Hz and 4037 Hz. These were considered merged formants and further investigation found partial activity at 3132 Hz, 3687 Hz, 4368 Hz and 5742 Hz. F3 is in a similar location to speech. (F3 Legit – 3132 Hz; Speech – 3246 Hz) The singer's formant (F3,F4,F5) seems approximately where predicted. F5 was missing in speech but there was some energy in the partial at 4390 Hz (2.0%). Legit showed energy in a similar spot at 4368 Hz (1.7%). F6 was a little higher than speech and energy slightly lower (Legit – 5742 Hz (0.9%); Speech – 5339 Hz (2.0%)). Legit showed formants between 5000 Hz and 10000 Hz at 6337 Hz, 7441 Hz and 8434 Hz with variable energy ranging from (1.1-4.8%). The subject commented that Legit was more like Pop on the bottom and more like Classical on the top. This study didn't investigate the higher pitches for each quality so it is difficult to know if this pedagogical comment holds true; however, this research seems to indicate that the lower range is similar to Pop in the upper partial activity. The vowel showed slight variations from [i] to [I] and nasality was noted in the sustained vowel samples, but not in the running samples. The subject noted that her tongue felt forward, but auditory analysis indicated her tongue moved back. The phonation type was whispery voice on the sustained vowels and breathy voice in the running samples. The inharmonic activity was medium (2.7). The larynx was neutral to low on the lowest pitch of the sustained vowel (B3), and slightly raised on the higher pitches (E4 and B4). In the running samples, the auditory analysis showed a slightly low to very low larynx even though the subject felt it as neutral. There was generally no pharyngeal constriction in this quality, except on the upper two notes of the sustained vowels (E4 and B4), which showed a medium amount of constriction. The two upper pitches also showed a raised larynx. The subject noted that the labial posture was horizontal expansion and vertical constriction (smile).

The LGG for Legit showed a moderate speed quotient. The slope was low-moderate. The visual analysis found that the Lx wave showed narrow peak width with a slight skew indicating less slap of the vocal folds in their closing phase, and more roll open/roll closed. The Lx waveform had a triangular shape with no skirt elevation, which may indicate that this quality uses a Testo vocal fold adjustment. The contact quotient was lower than speech (20-35% with one outlier), which would indicate weak adduction. There was some difference between initial slope (0-.25) and slope (.25-.75), which suggested parallel closure of the vocal folds in the horizontal plane in 5/7 samples, zippering in one case and one outlier that had an initial slope greater than the slope.

4.5.8 Belt

The mean frequency for the Belt samples was 286 Hz. There was good formant tracking of P1 with F1 in the Belt quality, however energy levels of P2 were considerably stronger than the other qualities. The mean energy in P2 was 27.0%. Bandwidths seem considerably wider than for the other qualities. F2 was low (2325 Hz). The larynx was raised in the majority of cases (5/7). The auditory analysis, however, didn't note any constriction. Summed energy was also high for Q1 and Q2. The high energy in Q2 was one of the major differences between this quality and the others. The DFT numerical results showed energy in F3 (16.0%), F4 (15.8%), and F5 (7.1%). Energy in F6 dropped rapidly to 0.8%. Approximate formant placements were 3000 Hz, 3600 Hz and 4250 Hz. F3 is a little lower than for speech (Belt – 3075 Hz; Speech – 3246 Hz). The singer's formant (F3,F4,F5) is in the expected location around 3500 Hz and F5 is near that for speech (Belt – 4248 Hz; Speech 4390 Hz). F5 moved downward toward the energy in the singer's formant. Belt shows some upper partial energy between 6500 Hz and 9000 Hz. There were formants at approximately 6500 Hz and 7500 Hz and a clustering of formants around 9000 Hz. The energy around 9000 Hz was medium strength (1.0%-4.3%). The auditory analysis showed little vowel modification with ascending pitch, but the [i] vowel was consistently modified slightly. The auditory analysis found the tongue position back and low. There was nasality in the running samples. Mostly modal voice phonation was used with some evidence of BV, WV and HV. Generally there was very little inharmonic activity (2.1). The laryngeal position was slightly high. Pharyngeal constriction varied from none to slight constriction. The subject noted sensations of wide larynx and mouth, "molar mouth" and firm lips while producing this quality.

The LGG for Belt showed a high speed quotient. The slope was low-moderate. The visual analysis found the Lx wave showed moderate peak width with a moderate skew. The low slope value seems to indicate a more roll open/roll closed mechanism; however, the skew indicates a large angle of convergence. In other qualities a steep slope was found in conjunction with the skew of the Lx waveform. Samples with ramping showed less skew and more triangularity. Instead, the Belt waveform was more rectangular with evidence of skirt elevation on the lower pitches. Higher pitches showed a more triangular shape with no skirt elevation. These results support the idea of register change in Belting. The contact quotient was higher than speech (40% +), which would indicate firm adduction. There was little difference between initial slope (0-.25) and slope (.25-.75), which suggested parallel closure of the vocal folds in the horizontal plane in 7/9 samples, slight zippering in two cases. This seems to indicate that the zippering found in all samples of the Classical quality may play a role in the voice quality. There also was more surface bulging, which may be due to the increase in TA.

4.6 Conclusions

The results showed that the subject used her voice consistently to produce seven different singing qualities and speech for the vowel [i]. A series of analyses quantified parameters that contribute to our perception and form a blueprint of each voice quality. The results of laryngographic and acoustic analyses also provide important clues to the physiology of voice production in different voice qualities. The process of observing and examining voice quality in singing and speech generated hypotheses that can be tested by further research.

The exploratory research in voice quality introduces ideas about the relationship of acoustical events to the perceived quality. For example, Formant tracking of F1 with P1 suggests an increase the volume. Increased energy in the second partial suggests the perception of a brighter quality (Belt, Country). The strength of F2 corresponds to vowel definition and intelligibility (Belt, Legit). The strength of the singer's formant contributes to the loudness of the quality, and the brassiness of the voice quality (Belt, Country). The absence of strength in the upper partials contributes to the purity of the quality (Classical, R&B). The lowering of formants darkens the quality (Classical, R&B). The raising of formants brightens the quality (Pop, Legit). The strength of the upper partials and inharmonic energy adds a hazy quality to the sound (Legit and Pop).

The LGG analyses provided clues to the mechanism at the source. The vocal fold shape was suggested to be the primary determining factor in voice quality. For example, brassier qualities tended to have squarer medial vocal fold edges. (Belt) Adduction and the slap and roll open seem to be less important to voice quality; but zippering in the horizontal plane may be important.

Another very important aspect of this thesis was the discussion surrounding terminology. This section raised questions about the flexibility of linguistic taxonomy for the study of the singing voice. However, the linguistic terminology is still much clearer than many of the alternatives and is constantly being updated. Singers tend to be very clear about sensation and sound, but less clear about physiological linkages to the acoustic events. Linguists even have a developed a language (IPA) which includes these linkages. Each symbol in the IPA is defined each speech sound very accurately supporting the acoustic to physiological linkages with years of research. Furthermore, linguists have begun to tackle voice quality issues with the same type of systematic labeling and definitions. Since singing is a human vocalization, I suggest that linguistic terminology be expanded to better describe the singing voice. The development of a common inclusive language for the study of voice would help all those who study human vocalizations communicate across the disciplines.

However, voice quality diversity in musical styles presents the performer and pedagogue with many challenges. This thesis may help to clarify some pedagogical practices of non-classical qualities vs. classical quality. For example, a voice teacher of Belt quality may prescribes an extreme forward placement and a feeling of pressure behind the nose to a student. This high degree of resonant sensation in the facial area and sinuses is slightly different than in classical voice, where a soprano may not feel as much frontal sensation, especially at higher fundamental frequencies. The results of the thesis show suggest singer's ring is stronger for the Belt voice, than the Classical voice. The thesis results are consistent with the existing pedagogical practice. On the other hand, a Classical voice teacher may demand a longitudinal stretch in the pharynx some call the "internal stretch" or the "beginning of a yawn" sensation. The results of the thesis also are consistent with this pedagogical practice. The results of the thesis suggest narrow formant bandwidths are very important to classical quality. Conversely the results of the thesis suggest that relaxing the pharyngeal walls in Belt and Country may contribute to the perception of these two voice qualities. The thesis results showed that the first formant for all soprano voice qualities studied tracked the fundamental frequency. Sundberg (1987) showed that the fundamental frequency and first formant lined up if the singer gradually opened her mouth as the pitch ascended. This posture shortened the vocal tract. The thesis results suggest that all sopranos, regardless of quality, need to track the fundamental, by making the vocal tract shorter.

At the vocal fold level, this study also shows a predominance of the rectangular shape of the vocal folds for Belt quality, but the migration to more wedge shaped vocal folds at higher pitches. These results are consistent with current Belt pedagogy, which advocates mixing registration as pitch ascends. The thesis results also show laryngeal position is important in voice quality. Sundberg (1987) related the breathing style to the laryngeal position. Low breathing promotes tracheal pull and lowers the larynx. The thesis results showed that many non-classical voice qualities require a higher laryngeal position. The thesis results support using diaphragmatic or low breath in Classical, R&B

and Jazz; however low breath in Pop, Legit, Belt and Country singing may not be advisable.

However, as much as this thesis has clarified some issues for perception and pedagogy, it has also raised more questions. Appendix B contains some direct physiological information from video fluoroscopy and MRI scans of Belt and Classical that could be a direction for further research. The questions raised in the thesis have been formulated into series of hypotheses for further research:

1) Whistle register in female singers is produced with vocal fold damping
2) Vocal fold damping can occur apart from pharyngeal constriction
3) Vocal fold damping can occur at any pitch
4) Countertenors use damped falsetto
5) Modal Voice, Testo Voice and Falsetto can be perceived as distinct voice qualities
6) Head voice in females above the secondo passaggio uses Falsetto as defined by Laver, 1980
7) The vibratory patterns of square medial edged vocal folds, wedge medial edged vocal folds, and vocal ligament are all perceivable.
8) Pitch changes in creak are a result of air pressure
9) Angle of convergence is dependent on vocal fold plane
10) Zippering of the vocal folds reduces upper partial energy
11) Pharyngeal walls are lax in Belt quality
12) Pharyngeal walls are tense in Classical quality
13) Nasality in Country quality is produced with an open velar port
14) Belt singers are more prone to vocal injuries than Classical singer

4.7 Recommendations for further study

This thesis took a limited number of [i] vowel samples from one subject using three acoustic analysis processes: Spectrogram, LPC and DFT. The LPC analysis was questioned earlier in this thesis. It is recommended that the LPC analysis method be tested experimentally in order to confirm its accuracy. Many questions surrounding registration emerged as a result of the survey of terminology. The taxonomy should be expanded and clarified. It is also recommended that the other nine vowel samples be analysed by this process in the running samples for differences in consonant activity and transitions. Running samples may provide a better indicator of different singing qualities since they represent a habitual response including articulations, and other stylistic parameters. This type of research may clarify a number of questions in this thesis. Hypotheses concerning acoustical, physical, and pedagogical correlates may be developed further. Perceptual testing may also aid in deducing which correlates form the important acoustical cues for each quality. The hypotheses should then be tested on a larger sample. A study that defines the linkages between acoustic and physiological events should be done. Lastly, vocal health should be assessed for these qualities.

It is my hope that in time the questions raised in this thesis will be answered and that the pedagogy surrounding non-classical voice qualities can be documented and tested by research in the same way that the classical voice quality has been documented and tested. Furthermore, it is hoped that this thesis contributes towards an inclusive taxonomy that will help us understand each other as we continue to talk about human vocalizations between the disciplines.

References:

AUSTIN, S. (1992). *Subglottal Resonance and Voice Registers.* PhD.Thesis, Iowa City: University of Iowa.

BAKEN, R. AND ORLIKOFF, R. (2000). *Clinical Measurement of Speech and Voice.* Second Edition. San Diego, Ca.: Singular Publishing Group.

BALL, M.J. (1993). *Phonetics for Speech Pathology.* London: Whurr Publishers.

BERK, G. AND GERRATT, B. (1993). Laryngeal Biomechanics: An Overview of Mucosal Wave Mechanics. *Journal of Voice,* 7, 123-128.

BEVAN, R. (1989). *Belting and Chest Voice: Perceptual Differences and Spectral Correlates.* EdD Thesis, New York: Columbia University Teachers College.

BOUNOUS, B. (1997). *Belt Voice: Acoustical Measurements and Esthetic Correlates.* PhD dissertation, Greeley: University of Northern Colorado.

CARLSSON, G. AND SUNDBERG, J. (1992). Formant Frequency Tuning in Singing. *Journal of Voice*, 6, 256-260.

CHILDERS, D. AND LEE, C. (1991). Vocal Quality Factors: Analysis, Synthesis, and Perception. *Journal of the Acoustical Society of America*, 90, 2394-2410.

CLEVELAND, T. (1994). A Closer Look at Formant Tracking. *Journal of Singing*, 50(3), 41-43.

COLLYER, S. (1997). The Classical Teacher and Belting. *Australian Journal of Voice.*

DICKSON, C. (1993). Acoustical Features for Identifying Phonetic Classes and Phonetic Elements, Victoria: STR.

DICKSON, C., ESLING, J. AND SNELL, R. (1994). Real-time Processing of Laryngographic Waveforms for the Evaluation of Phonation Types. *ICSLP 94 Proceedings.* Yokohama: Acoustic Society of Japan.

DICKSON, C., ESLING, J. AND SNELL, R. (1992). Analysis of Phonation Type Using Laryngographic Technique. *ICSLP 92 Proceedings,* 2, 1107-1110. Edmonton: University of Alberta.

ESLING. J. (1998). Laryngoscopic Analysis of Pharyngeal Articulations and Larynx-Height Voice Quality Settings. *Proceedings of the 5th International Conference of Spoken Language Processing,* 2, 543-546. Sydney: ASSTA.

ESLING. J. (1984). Laryngoscopic Study of Phonation type and Laryngeal Configuration. *Journal of the International Phonetic Association,* 14, 56-73.

ESLING. J., CLAYARDS, J., EDMONDSON, J.A., FUYUAN, QUI, AND HARRIS, J. (1998). Quantification of Pharyngeal Articulations Using Measurements From Laryngoscopic Images. *Proceedings of the 5th International Conference of Spoken Language Processing,* 7, 3091-3094. Sydney: ASSTA.

ESLING. J., DICKSON, C. AND WOOLSEY, J. (1991). Automatic Procedure for Laryngographic (Lx) Analysis of Phonation Contrasts. *Actes du XIIeme Congres International des Sciences Phonetiques*, 3, 6-9. Aix-en-Provence: Universite de Provence.

ESLING. J., DICKSON, C., HEAP, L. M. AND SNELL, R. (1994). Analysis of Pitch Dependence of Pharyngeal, Faucal, and Larynx-Height Voice Quality Settings.*ICSLP* 94: 1475-1478. Yokohama: Acoustic Society of Japan.

ESLING. J. AND HARRIS, J. (2003). An Expanded Taxonomy of States of the Glottis. *Proceedings of the 15th ICPhS,* 1049-1052. Barcelona: UAB.

ESTILL, J. (1984). Compulsory Figures for Voice and Belting. Unpublished text book, 71-74.

ESTILL, J. (1988). Belting and Classic voice quality: Some Physiological Differences. *Medical Problems of Performnig Artists*, 3, 37-43.

ESTILL, J., FUJIMURA, O., SAWADA, M. AND BEECHIER, K. (1988). Temporal Pertubations and Voice Qualities. In Fujimura, O. (Ed.) *Vocal Fold Physiology Volume 2, Vocal Physiology Voice Production, Mechanisms and Functions.* 237-252.

ESTILL, J., KOBAYASHI, N., HONDA, K. AND KAKITA, Y. (1990). A Study on Respiratory and Glottal Controls in Six Western Singing Qualities: Airflow and Intensity Measurement of Professional Singing. ICSLP '90 Proceedings, 169-172. Kobe: Acoustical Society of Japan.

ESTILL, J., FUJIMURA, O., ERICKSON, D., ZHANG, T. AND BEECHLER, K. (1994). Vocal Tract Contributions to Voice Qualities. *Proceedings of the Stockholm Music Acoustics Conference July 28-August 1, 1993,* 161-165. Taberg: Royal Swedish Academy of Music.

FABRE, P. (1958). Etude comparee des glottogrammes et des phonetogrammes de la voix humaine, *Annales d'Oto-Laryngology*. 75, 767-775.

FANT, G. (1960). *Acoustic Theory of Speech Production.* The Hague: Mouton, 44 and 51.

FISHER, K., SCHERER, R., SWANK, P., GIDDONS, C., AND PATTEN, D. (1995). Electroglottographic tracking of phonatory response to Botox. *NCVS Status and Progress Report*, 8, 97-111.

FRY, D.B. (1979). *The Physics of Speech*. Cambridge: Cambridge University Press.

HEAP, L.M. (1998). *A Phonetic Model of the Human Pharynx*. M.Sc. Thesis, Victoria: University of Victoria.

HERBERT-CAESARI, E. (1936). *The Science and Sensations of Vocal Tone*. New York: Crescendo Publishing.

HIRANO, M., VENNARD, W., AND OHALA J. (1970). Regulation of register, pitch and intensity of voice. *Folia Phoniatrica*, 22, 1-20.

HOLLIEN, H. (1972). Three Major Vocal Registers: a Proposal. *Proceedings of the 7th ICPhS*, 320-331. The Hague: Mouton.

HOLLIEN, H., DAMSTE, H., AND MURRY, T. (1969). Vocal fold length during vocal fry phonation. *Folia Phoniatrica*, 21, 257-65.

HONDA, K., HIRAI, H., ESTILL, J. AND TOHKURA,Y. (1994). Contribution of Vocal Tract Shape to Voice Quality: MRI Data and Articulatory Modeling. In O. Fujimura and I. Hirano (Eds.), *Vocal Fold Physiology: Voice Quality Control*, 8(2), 23-38. San Diego, Ca: Singular Publishing Group, Inc.1995.

HOUBEN, G., BUEKERS, R., AND KINGMA, H. (1992). Characterization of the Electroglottographic Waveform: A Primary Study to Investigate Vocal Fold Functioning. *Folia Phoniatrica*, 44, 269-281.

KAY ELEMETRICS CORPORATION (1991). *Computerized Speech Laboratory Operations Manual*. New Jersey: Kay Elemetrics.

KEIDAR, A., HURTIG, R. AND TITZE, I. (1987). The Perceptual Nature of Vocal Register Change. *Journal of Voice*, 1(3), 223-233.

KENT, R. AND BALL, M. (2000), *Voice Quality Measurement*. San Diego, Ca: Singular Publishing Group, Inc.

KMUCHA, S., YANAGISAWA, E. AND ESTILL, J. (1990). Endolaryngeal Changes During High-Intensity Phonation Videolaryngoscopic Observations. *Journal of Voice*, 4: 346-354.

LADEFOGED, P. (1993). *A Course in Phonetics*. Fort Worth: Harcourt Brace Jovanovich College Publishers.

LADEFOGED, P. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

LAVER, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

LAWRENCE, V. (1979). Laryngological Observations on Belting. *Journal of Research in Singing* V2/1:26-28.

LECLUSE, F., BROCAAR, M. AND VERSCHURRE, J. (1975). The Laryngography and its relation to glottal activity. *Folia Phoniatrica* 27, 215-224.

MACCURTAIN, F. AND FOURCIN, A. (1982). Applications of the Laryngograph waveform display. In Lawrence, V. (Ed.), *Transcripts from the Tenth Symposium on the Care of the Professional Voice,* II, 51-57.

MARKEL, J.D. AND GRAY, A. H. JR. (1976). *Linear Prediction of Speech*. New York: Springer-Verlag.

MAULTSBY, P. K. (1981). *Afro-American Religious Music: a study in musical diversity*. Boston, MA: The Hymn Society in the United States and Canada.

MAULTSBY, P. K. (1990). Africanisms in African-American Music. In Holloway, J. E. (Ed.), *Africanisms in American Culture.* Bloomington: Indiana University Press. 185-210.

MCKINNEY, J.C. (1994). *The Diagnosis and Correction of Vocal Faults*, Nashville, Tennessee: Genevox Music Group.

MILES, B. AND HOLLIEN, H. (1990). Whither Belting? *Journal of Voice*, 4(1), 64-70.

MILLER, R. (1986). *The Structure of Singing.* New York: Schirmer.

MILLER, R. (1994). Spectral Components of Five Cardinal Vowels in the Soprano Singing Voice Considered by Means of "Sequential Vowel Diagonal". *Journal of Singing*, 50(4), 5-7,60.

MILLER, R. (1995). What about the "Pure Vowel" and "Vowel Modification"? *Journal of Singing*, 52(1), 37-40.

MILLER, R. (1996). *On the Art of Singing*. New York: Oxford University Press.

MOORE, G.P. (1971). *Organic Voice Disorders.* Englewood Cliffs: Prentice-Hall.

MORRISON, M. (1986). *Video Highlights from the Vancouver Voice Clinic.* NewYork: Voice Foundation.

MURRY, T. (1988). Vocal Tract Parameters Associated with Voice Quality and Preference. *Journal of Singing*, 2:111-117.

ORLIKOFF, R. (1991). Assessment of the Dynamics of Vocal Fold Contact From the Electroglottogram: Data From Normal Male Subjects. *Journal of Speech and Hearing Research*, 34, 1066-1072, October 1991.

ORLIKOFF, R. (1995). Vocal Stability and Vocal Tract Configuration: An Acoustic and Laryngographic Investigation. *Journal of Voice,* 9, 173-181.

ORLIKOFF, R., BAKEN, R., AND KRAUS, D. (1997). Acoustic and Physiological Characteristics of Inspiratory Phonation. *Journal of Acoustical Society of America,* 102, 1838-1845.

PARSONS, T. (1986). *Voice and Speech Processing.* New York: McGraw-Hill Book Company.

POPEIL, L. (1996). The Total Singer Voiceworks Method. Sherman Oaks: Voiceworks.

ROBINSON, C., BOUNOUS, B. AND BAILEY, R. (1994). Vocal Beauty: A study proposing its acoustical definition and relevant causes in classical baritones and female belt singers. *Journal of Singing*, 51(1), 19-30.

ROSSING, T. (1990). *The Science of Sound*. Massachusetts: Addison-Wessley Publishing Company.

ROTHENBERG, M. (1981). Some relations between glottal airflow and vocal fold contact area. *ASHA Reports* , 11, 88-99.

ROTHENBERG, M., AND MAHSHIE, J.J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research*, 31, 338-351.

SCHERER, R.C., DRUKER, D. AND TITZE, I. (1988). Laryngography and Direct Measurement of Vocal Fold Contact Area. In Fujimura, O. (Ed.), *Vocal Physiology: Voice Production. Mechanisms and Functions*, 279-291. San Diego, California: Raven Press Ltd.

SCHERER, R.C., GOULD, W.J. AND TITZE, I., MEYERS. A.D., AND SATALOFF, R.T. (1988). Preliminary Evaluation of Selected Acoustic and Glottographic Measurements for Clinical Phonotory Function Analysis, *Journal of Voice*, 2, 230-244.

SCHUTTE, H. AND MILLER, D. (1993). Belting and Pop, Nonclassical Approaches to the Female Middle Voice: Some Preliminary Considerations. *Journal of Voice,* 7(2), 142-150.

STONE, R. E., CLEVELAND, T. F. AND SUNDBERG, J. (1999). Formant Frequencies in Country Singers' Speech and Singing. *Journal of Voice,* 13(2), 161-167.

SULLIVAN, J. (1985). *The Phenomena of the Belt/Pop Voice.* Denver CO: Logos Ltd.

SULLIVAN, J. (1989). How to Teach the Belt/Pop Voice. *Journal of Research In Singing and Applied Vocal Pedagogy*, 13, 41-58.

SUNDBERG, J. (1974). Articulatory interpretation of the "singing formant". *JASA* 55, 838-44.

SUNDBERG, J. (1977). The Acoustics of the Singing Voice. excerpted from *Scientific American, The Physics of Music*, March 1977, California: W.H. Freeman and Company.

SUNDBERG, J. (1987). *The Science of the Singing Voice.* Illinois: Northern Illinois University Press.

SUNDBERG, J., CLEVELAND, T.F. AND STONE, R.E. AND IWARSSON, J. (1999). Voice Source Characteristics in Six Premier Country Singers. *Journal of Voice*, 13 (2), 168-183.

SUNDBERG, J., GRAMMING, P. AND LOVETRI, J. (1993). Comparisons of Pharynx, Source, Formant, and Pressure Characteristics in Operatic and Musical Theatre Singing. *Journal of Voice*, 7 (4), 301-310.

TITZE, I. (1983). The Importance of Vocal Fold Loading in Maintaining Vocal Fold Oscillation. *Proceedings from the Stockholm Musical Acoustics Conference, Royal Swedish Academy of Music*, SMAC83,1, 46(2), 61-72.

TITZE, I. (1988). A Framework for the Study of Vocal Registers. *Journal of Voice*, 2(3), 183-194.

TITZE, I. (1989). A Four-Parameter Model of the Glottis and Vocal Fold Contact Area. *Speech Communication* 8, 191-201.

TITZE, I. (1990). Interpretation of the Laryngographic Signal. *Journal of Voice*, 4(1), 1-9.

TITZE, I. (1994). *Principles of Voice Production*. New Jersey: Prentice Hall.

TITZE, I. (1994). Toward Standards in Acoustic Analysis of Voice. *Journal of Voice*, 8(1), 1-7.

TITZE, I. (1995). Speaking Vowels Versus Singing Vowels. *Journal of Singing*, 52(1), 41-42.

TITZE, I. (1995). What's in a Voice. *New Scientist*, September 23, 1995: 38-42.

TITZE, I. (1997). Acoustic interactions of the voice source with the lower vocal tract. *Journal of the Acoustical Society of America*, 101(4):April 1997:2234-2243

VALLENCIEN, B., AND FAULHABER, J. (1967). What to think of Glottography. *Folia Phonetiatrica,* 19, 39-44.

VAN MICHEL, C., PFISTER, K.A. AND LUCHSINGER, R. (1970). Electroglottographie et cinématographie laryngée ultra-rapide: Comparaison des résultats. *Folia Phoniatrica*, 22, 81-91.

VENNARD, W. (1949). *The Mechanism and the Technic*. New York: Fischer.

VILKMAN, P., ALKU, P. AND LAUKKANEN, A. (1995). Vocal Fold Collision Mass as a Differentiator Between Registers in the Low-Pitch Range. *Journal of Voice*, 9, 66-73.

WALKER, R. (1992). Formants in singers: stylistic and cultural perspective. *British Voice Association*, 1, 19-46.

WECHSLER. E. (1977). A Laryngographic Study of Voice Disorders. *British Journal of Disorders of Communication*, 12, 9-22.

YANAGISAWA, E. AND ESTILL, J. (1989). The Contribution of Aryepiglottic Constriction to "Ringing" Voice Quality. *Journal of Voice*, 3, 342-350.

YANAGISAWA, E., ESTILL, J., MAMBRINO, L., AND TALKIN, D. (1991). Supraglottic Contributions to Pitch Raising. *Annals of Otology, Rhinology and Laryngology*, 100:10-30.

CD-ROM Contents

Appendix A: Measurements and Results

List of samples

LPC order # decision

Numerical Data: Spectrogram and LPC

Box Plot Analysis of LPC Numerical Data

P1 vs. P2: Spectrogram, DFT numerical and LPC/DFT overlay comparisons

Vowel Formant and Singer's Formant Comparison: Spectrogram, LPC, and DFT

Graph of Average Formant Placement (Spectrogram)

Scatterplots of LPC formant data

Dot plots comparing Spectrogram and LPC data: median and mean

Summed Energy Plot and DFT numerical Results

Inharmonic Activity

Lisa's sensations

Auditory Analysis

Graph: LGG Frequency vs. Speed Quotient

Graph: LGG Frequency vs. Contact Quotient

Appendix B: Spectrograms, FFT, LPC, LGG, X-ray Fluoroscopy, MRI

Belt, Country, Jazz, Legit, Classical, Pop, R&B and Speech

- spectrograms
- LPC: long and short (numerical data in Appendix A)
- FFT: with/without preemphasis, numerical data and stem plots
- LPC/DFT overlay
- CSL LGG numerical data
- CSL graphical LGG data
- LGG Slope and Frequency calculations
- Representative Lx waveforms used for slope calculations and visual analysis
- X-ray and MRI

Appendix C

List of Samples

Speech

SPK 7: Speech Quality-Heed Had Hawed Hayed Who'd

Belt Quality

B1	[i] vowel low pitch B3
B11	[i] vowel medium pitch E4
B21	[i] vowel Medium high pitch B4
B31	[i] [ae] [e] vowels Medium high pitch B4 retake
B40	running sample "Starting here starting now everything's coming up roses"
B40a	"here"
B40b	"every"
B41	running sample "There are guys just meant for some kissing and I mean to kiss me a few"
B41b	"mean"
B45	running sample "There are guys just meant for some kissing and I mean to kiss me a few" higher in pitch
B45b	"mean"

Country

C1	[i] vowel low pitch B4
C11	[i] vowel medium pitch E4
C21	[i] vowel Medium high pitch B5
C31	[i] vowel Medium high pitch B5 retake
C37	running sample "Home home on the range where the deer and the buh..antelope play"
C37b	"deer"

Classical Quality

O1	[i] vowel low pitch B4
O11	[i] vowel low pitch B4 retake with more chest sound
O21	[i] vowel medium pitch E4
O31	[i] vowel Medium high pitch B5
O44	running sample "You make the listening shores rebound"
O44b	trimmed "rebound"

Jazz

J1	[i] vowel low pitch B4
J11	[i] vowel medium pitch E4
J21	[i] vowel Medium high pitch B5
J33	running sample “Got the moon above me and no one to love me”
J33b	“me”
J33b2	“me”
J38	running sample “I’ve got daisies in green pastures. I got my man who could ask for anything more
J38b	“daisies”
J38b2	“green”

Legit

L1	[i] vowel low pitch B4
L2	[i] vowel low pitch B4
L12	[i] vowel medium pitch E4
L22	[i] vowel Medium high pitch B5
L34	running sample “with the songs they have sung for a thousand years”
L34b	“years”
L35	running sample “do a deer a female deer, re a drop of golden sun”
L35b	“deer”
L35b2	“deer”

Pop

P1	[i] vowel low pitch B4
P12	[i] vowel medium pitch E4
P22	[i] vowel Medium high pitch B5
P33	running sample “So get up and get away to McDonalds we do it all for you”
P33b	“we”
P39	running sample “or ask the grinning bobcat why he grinned”
P39b	“he”

R&B

R1	[i] vowel low pitch B4
R11	[i] vowel medium pitch E4
R21	[i] vowel Medium high pitch B5
R31	running sample “Amazing grace how sweet”
R31b	“sweet”
R37	running sample “You can reach me by trainway”
R37b	“reach”