Drew University

College of Liberal Arts

A Technical Analysis of Polyphonic Vocal Styles

A Thesis in Music

by

Ian W. Nadler

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Bachelors in Arts

With Specialized Honors in Music

May 2018

ABSTRACT

Polyphonic singing is a lesser-known vocal technique that allows singers to produce two or more pitches simultaneously; in essence, singers can harmonize with themselves. A handful of cultures globally have independently developed polyphonic singing techniques; however, few of these techniques have been studied scientifically. Establishing a scientific understanding of the production and perception of the polyphonic voice is the goal of this thesis. To achieve this, the polyphonic voice will be explored using three scientific disciplines. The first will be physics, in order to evaluate the complex facets of sound, such as the harmonic series, that make polyphonic singing possible. An anatomical and physiological study of the vocal apparatus follows, to identify not only what structures are involved, but how the lungs, throat, and mouth all work in concert to produce these unique polyphonic voices. Third, a neurophysiologic focus on auditory processing will examine auditory perception of the sounding polyphonic voice. Finally, with the scientific foundation established, the explored concepts will be applied to a focal culture, in this case Tuva in central Asia, as a means to contextualize the practice of polyphonic singing in a society. Tuvan throat singing is not only one of the most iconic polyphonic vocal styles worldwide, but in practice maintains one of the most diversified and unique stylistic approaches to its throat-singing voices. By developing the scientific foundation, and applying these fundamentals to a technically diverse cultural example, the hope is that one will then possesses the tools to approach the study of any polyphonic vocal style globally.

TABLE OF CONTENTS

TABLES AND FIGURES

ABBREVIATIONS

1CM	one cavity method
2CM	two cavity method
A1	primary auditory cortex
A2	secondary auditory cortex
AEF	aryepiglottic folds
CN	cranial nerve
CNS	central nervous system
CT	cricothyroid
dB	decibels
Hz	Hertz
K+	potassium
LCA	lateral cricoarytenoid
fMRI	functional magnetic resonance imaging
OA	oblique arytenoid
OCE	otoacoustic emissions
PCA	posterior cricoarytenoid
PET	positron emission tomography
TA	thyroarytenoid
TVA	transverse arytenoid
VTF	vestibular folds (false vocal cords)

INTRODUCTION

Polyphonic singing is actually a general term. In analyzing the Latin roots of the word, “poly” means more than two or many, and “phon” means sound, so essentially polyphonic means many sounds. By then adding singing to the word polyphonic, it then refers to singing techniques where more than one sound is produced. This is not to be confused with vocal polyphony, which is the basis for most choral songs and refers to multiple voices singing different vocal lines harmonizing with one another. The polyphonic voice is performed by a single vocalist (Bannon 2012, 158). This vocal approach has been observed in various cultures of Central Asia such as Tuva, Mongolia, and Tibet, in the South African Xhosa Women, the Dani people of New Guinea, and even in some Western cultures, but each goes about polyphonic singing in a unique way. This can be achieved using a litany of approaches, but there are two overarching categories classifying the polyphonic voice. The first is bitonality, which is simply any polyphonic voice where two or more unique sounds sources are produced, such that the pitches emitted do not have a direct harmonic relationship to one another (Cosi & Tisato 2003, 2003, 8). Examples of this range from growl singing, associated with the timbral quality of Louis Armstrong, to medical conditions such as diplophonia. Diplophonia is not generally produced intentionally, and occurs when an individual has asymmetric vibrations of the vocal cords, as is commonly seen in cases of unilateral vocal cord paralysis, and can even happen to singers who exhibit excessive laryngeal tension

(Edgerton 2015, 101). Essential to the production of the bitonal voice is the ability to volitionally control multiple oscillating structures in the throat in addition to the primary vocal cord sound source. The second category of polyphonic voice is known as diphonia, and occurs when two or more pitches are produced that do have a harmonic relationship to one another (Cosi & Tisato 2003, 9). This is generally thought of as splitting of the voice, since for diphonic voices the multiple pitches generated originate from a single sound source, but due to certain resonant filtering methods like formant manipulation, to be discussed further in CHAPTER 2: The Anatomy and Physiology of Sound, harmonics already present in the fundamental sound source get amplified to audible levels. Diphonic styles are far more prevalent in the current practice of polyphonic voices, the most common of which are Western overtone singing and Tuvan throat singing.

While these two broad categorizations of the polyphonic voice, bitonality and diphonia, were created to make identification easier, it is also critical to note that there are polyphonic vocal styles that can be classified as both bitonal and diphonic, as well as those that cannot officially be grouped into either of these definitions. For instance, kargyraa, which is one of the main styles of Tuvan throat singing to be expanded upon in CHAPTER 4: A Study of Polyphonic Voice in Tuva not only exhibits multiple laryngeal sound sources, but also utilizes formant shaping to amplify a harmonic upper line (Ken-Ichi et al. 2007, 1). Using this technique actually provides the singer with the capability to produce at least three tones simultaneously. Then there are the cases that do not quite fit under either umbrella, where the normal modal singing voice is paired with another phonation technique, such as whistling, or buzzing of the lips.

When polyphonic singing is discussed, the two most common genres are overtone singing and throat singing. Overtone singing is more of a Western musical practice that solely uses the diphonic method of harmonic series manipulation to generate the second tone. On the other hand, throat singing uses the harmonic series, in addition to certain styles that also incorporate both bitonal and diphonic singing. To fully comprehend how these polyphonic vocal techniques work, one must first understand and have an appreciation for the physics of sound and the anatomy of the vocal tract making it all possible, as well as the auditory system allowing humans to perceive this aural phenomenon.

As such the purpose of the first three chapters will be to build a scientific foundation upon which to comprehend the polyphonic voice. CHAPTER 1: The Physics of Sound covers the physics of sound as it pertains to the four fundamental tenets of music: duration, loudness, pitch and timbre. In this section, the main focus is the harmonic series as it defines timbre and how this series facilitates the production of the polyphonic voice. This section also details the unavoidable dissonant clash when utilizing the polyphonic voice in a Western instrumental context due to a combination of differently derived temperaments.

CHAPTER 2: The Anatomy and Physiology of Sound presents an overview of vocal anatomy, categorizing the vocal mechanism into three constitutive parts that fulfill the roles of: 1) supplying power to the voice, 2) providing the sound source, and 3) facilitating resonance for the singer. The physical structures that execute these three tasks are the lungs, throat, and mouth, respectively. Formants are explained and simply correspond to the relative shape of the mouth, or resonating cavity that is integral to the production of any diphonic polyphonic voice. When delving into the specific throat-singing styles exhibited in the Tuvan culture as expressed in CHAPTER 4: A Study of Polyphonic Voice in Tuva referring back to the anatomy presented in CHAPTER 2: The Anatomy and Physiology of Sound may be helpful.

CHAPTER 3: The Neurobiology of Sound is a review of the neurophysiology of the auditory system, presented predominantly as a means to conceptualize the role sensory perception has in influencing the sound a listener ultimately processes cognitively. Highlights of this discussion are the review of the missing fundamental effect, in addition to the concept of combination tones as examples that directly pertain to the polyphonic listening experience. These two cases illustrate the inconsistencies between perception and the original sound stimulus caused by certain aspects of the auditory mechanism where complex tone processing is concerned.

CHAPTER 4: A Study of Polyphonic Voice in Tuva establishes Tuva as the focal culture, as a means to contextualize the use of the polyphonic voice within a society. The choice to analyze the Tuvan culture was made based upon three factors. The Tuvan culture boasts the most well-known polyphonic vocal style in their throat singing. Certain cultural ideals unique to this society likely contributed not only to the early inception of the polyphonic voice in this region as compared to the relatively recent instilment of polyphonic singing in Western society, but also to the diverse development of numerous approaches to this vocal technique conveyed in the different substyles practiced by Tuvan throat singers. Finally, the third distinguishing factor of Tuvan culture is how its approach utilizes both bitonal and diphonic polyphonic techniques, which facilitates an understanding of the production of the voice in this culture and how it can be applied when examining the polyphonic style of nearly any other world culture. While the core of this thesis revolves around the scientific intricacies that facilitate the polyphonic voice, it is worth noting that this culture produces such a vocally demanding technique without an explicit understanding of the underlying scientific anatomical/mechanistic mechanisms.

CHAPTER 1: The Physics of Sound

What is Music?

Limitation in any art form can never truly be attributed to the medium itself. Blame must be accepted by its progenitor. Taking music for example: a popular, while misguided opinion is that since there are only 12 unique pitches in the Western tonal system, there are only so many unique combinations of those pitches possible due to simple probability. Therefore, since such a vast variety of music has already been created, there must not be many new ideas left in music. Anyone even pondering this fallacy does not realize the sheer complexity of the musical art form made possible thanks to the fundamental physics of sound itself. Manipulation of pitch is but one variable a musician possesses to generate unique sonorities. In fact, the sheer volumes of aspects that can be modified about a sound wave are more likely to cause hesitation in a composer, rather than the running out of new ideas. Every unique sound possesses an original sound wave form and defined characteristics. Due to this fact sounds such as the polyphonic voice are possible, and when compared to a piano, sound unique. While the impact of all four musical tenets will be considered as to their role allowing for the possibility of the polyphonic voice, a special focus will be provided to timbre. The harmonic series dictates sound timbre making overtone singing possible.

Music would not exist without the physics of sound. Therefore, understanding the physical properties of sound pertinent to music is crucial to comprehending polyphonic singing as a type of musical production. Sound itself is vibration that manifests as an audible pressure wave, transmitting through a medium, whether it is a gas, which is the case a majority of the time, a liquid, or a solid. In music, sound is broken down into four major tenets: duration, loudness, pitch, and timbre (Dowling & Harwood 1986, 5). Each of these musical classifications, presented in increasing order of importance towards understanding polyphonic singing, describes fundamental aspects of oscillating sound waves.

The first three musical aspects--duration, loudness, and pitch--can be demonstrated when looking at a simple tone. A simple tone is a single sine wave oscillating in harmonic motion. Duration is the total time that a sound wave oscillates. Loudness is the amplitude of an oscillating wave or the height of a wave that is converted into a logarithmic scale of intensity, measured in decibels (dB). Pitch is the frequency at which the wave oscillates, traditionally measured in Hertz (Hz). The fourth musical aspect, timbre, however, is a bit more complicated since there is not a simple specific correlate to an aspect of a sound wave.

Timbre, or tone color, is really a concept that combines the amplitude and frequency of a complex tone. A complex tone is the combination of two or more simple tones (Sundberg 1994, 25). Therefore, a simple tone such as the previously mentioned sine wave tone does not technically have timbre. However, a sine wave tone is not exactly the best musical sound source to us, since it is virtually nonexistent in nature and can really only be generated electronically. Naturally occurring sound, such as the human voice, a clarinet, or chirping of a bird, is innately complex. At the root of it all is the harmonic series, which is the phenomenon responsible for complex tones and timbre itself.

Without the harmonic series, timbre would not exist and all sound would be pure sine wave tones. Also, commonly referred to as the overtone series, the harmonic series is the frequency pattern that any complex sound naturally follows. McCoy details the importance of harmonics in a relatable way as follows:

> Harmonics are extremely important. In that regard, I often tell my pedagogy students that they are like sex. (Do I finally have your attention?) Anyone who has taken a sex-ed course in the last 30 years knows that when you sleep with someone, you also are sleeping with everyone that person has ever slept with. And sometimes you get in trouble not because of the person you are with, but because of someone thrice removed. Similarly, when you sing a pitch, you are not just singing that note; you are singing every harmonic in the series above it. And sometimes we have vocal problems not because of the pitch, but because of the third harmonic. (McCoy 2013, 44)

Calculated mathematically this series starts with the fundamental frequency say F1 (43.65Hz) and follows a general pattern where the 2nd harmonic (an F2) sounds at the octave with a frequency twice the fundamental (87.31Hz), the 3rd (a C3) at the 12th three times the fundamental (130.81Hz), the 4th (an F3) two octaves above the fundamental or times four, and goes on as depicted below (Figure 1).

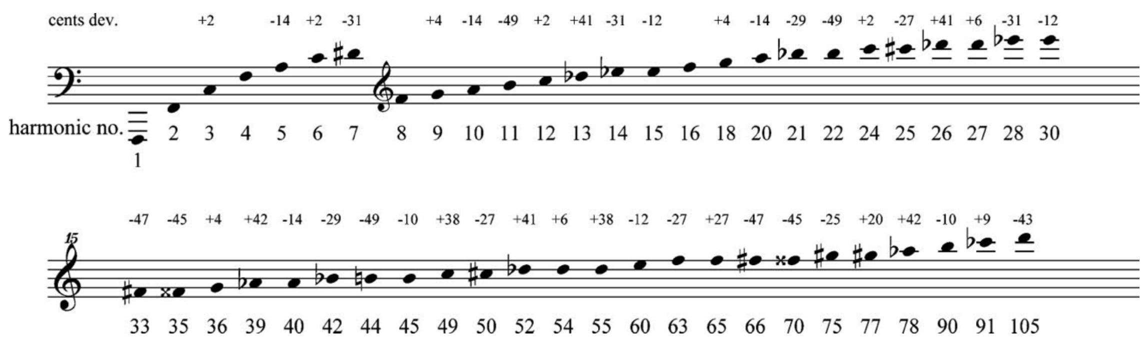

Figure 1: Harmonic Series

Harmonic series musically notated with F1 as the fundamental. The numbers above each harmonic indicate the magnitude and direction of cent difference relevant to equal temperament from which each harmonic sounds from: (Wannamaker 2008, 99)

What is important to note is that all of these higher frequencies sound at the same time within a fundamental pitch but at low amplitudes (Wannamaker 2008, 99). This pattern continues infinitely, where in theory the 512th harmonic would be an F10 that is nine octaves above the fundamental (22350.607Hz); however, not only is that frequency outside the range of human hearing, but the amplitude of that frequency would be so miniscule relative to the fundamental it might as well be considered nonexistent in practice. While infinite, the harmonic series is not without fault, but before we can delve into issues with musical application of overtones, comprehending tuning is crucial.

Tuning is such a major focus in all musical settings that any musician cannot get through a rehearsal without it being brought up at some point. Put simply, tuning is making sure that instruments produce pitches at matching frequencies, such as the cello and a trombone both playing the same fundamental frequency when generating, for example, an A4. When instruments do not play in tune with each other, depending upon the magnitude of that frequency difference, one hears beats if the notes are sustained together (Dowling & Harwood 1986, 36). Those beats slow down the closer the two frequencies get to one another to the point where the beats disappear. Say that same cello player plays a note at 440Hz while the trombone player produces a note of 442Hz. In that case, the beats that are created are due to the constructive interference when the two waves are in phase, briefly creating a larger amplitude in addition to the moment when these two waves are out of phase, causing destructive interferences that generate a lower amplitude. Therefore, notes that are out of tune, as well as notes of certain intervals, sound dissonant to one another, which is not only an issue in the tuning of fundamental frequencies but also carries on into the harmonic content generated by a complex tone. For instance, modifying the previous example but looking at the 2nd harmonic, the cello would create a frequency of 880Hz whereas the trombone's 2nd harmonic would have a frequency of 884Hz, which would clash even more than the fundamental frequencies, creating faster beating but at a lower amplitude since the first harmonic is not as loud as the fundamental. The decreasing intensity of harmonics as they progressively get farther away from the fundamental pitch is known as spectral tilt (McCoy 2013, 44). However, dissonance is not just a result of improper tuning, but a byproduct of the harmonic series.

At a rudimentary level, consonance as well as dissonance in musical terms refer to how well the tones sound together perceptually. However, there is far more to these musical concepts, which can be observed when comparing the harmonic series content of different intervallic relationships, where an interval is the distance between two pitches. Throughout music history, the most consonant intervals--the octave, perfect 5th and perfect 4th--have been the most commonly used, but also almost universally agreed to be in Western society the most aesthetically pleasing to the ear (Goldman 2002, 28). Once again, the harmonic series explains this, in that these intervals share the greatest number of partials in their respective harmonic series. Octaves are the most consonant naturally since the higher note of the interval is twice the frequency of the lower pitch. This is due to the fact that the pattern of the harmonic series is multiples of the fundamental. The higher pitch does not possess any unique overtones relative to the lower tone, so harmonically the upper tone is entirely represented within the lower.

Table 1: Explaining Consonance and Dissonance Using Harmonics

			Harmonics									
	Note	x:y	1	2	3	4	5	6	7	8	9	10
P8	A	1:1	110.00	220.00	330.00	440.00	550.00	660.00	770.00	880.00	990.00	1,100.00
	A	2:1	220.00	440.00	660.00	880.00	1,100.00	1,320.00	1,540.00	1,760.00	1,980.00	2,200.00
P5	A	1:1	110.00	220.00	330.00	440.00	550.00	660.00	770.00	880.00	990.00	1,100.00
	E	3:2	165.00	330.00	495.00	660.00	825.00	990.00	1,155.00	1,320.00	1,485.00	1,650.00
P4	A	1:1	110.00	220.00	330.00	440.00	550.00	660.00	770.00	880.00	990.00	1,100.00
	D	4:3	146.67	293.33	440.00	586.67	733.33	880.00	1,026.67	1,173.33	1,320.00	1,466.67
M6	A	1:1	110.00	220.00	330.00	440.00	550.00	660.00	770.00	880.00	990.00	1,100.00
	F#	5:3	183.33	366.67	550.00	733.33	916.67	1100.00	1283.33	1466.67	1650.00	1833.33
M3	A	1:1	110.00	220.00	330.00	440.00	550.00	660.00	770.00	880.00	990.00	1,100.00
	C#	5:4	137.50	275.00	412.50	550.00	687.50	825.00	962.50	1,100.00	1,237.50	1,375.00
m2	A	1:1	110.00	220.00	330.00	440.00	550.00	660.00	770.00	880.00	990.00	1,100.00
	A#	16:15	117.33	234.67	352.00	469.33	586.67	704.00	821.33	938.67	1056.00	1173.33
TT	A	1:1	110.00	220.00	330.00	440.00	550.00	660.00	770.00	880.00	990.00	1,100.00
	Eb	45:32	154.69	309.38	464.06	618.75	773.44	928.13	1,082.81	1,237.50	1,392.19	1,546.88

Source: Walk That Bass. 2017. "10. The Overtone Series and Dissonance." YouTube. Walk That Bass. https://www.youtube.com/watch?v=0lmS5lQ5MSU.

Note: Harmonic spectrum of different intervals all relative to A=110Hz. In each interval comparison, matching colors correspond with shared frequencies between harmonics produced by fundamental pitches sounding. More color pairings in an interval indicate a greater degree of consonance. x:y= frequency ratio of the interval; Intervals: P8=Octave, P5=Perfect 5th, P4=Perfect 4th, M6= Major 6th, M3=Major 3rd, m2=minor 2nd, TT=Tritone

From Table 1 it is evident that the perfect fifth followed by the perfect fourth respectively share the next most harmonic content in their overtone spectrums, making them very consonant, but less so than the octave. Then, jumping to the other end of the spectrum, there are the minor 2nd and the tritone, which are the most dissonant intervals in all of music. They clash predominantly because neither of these intervals shares a single harmonic in common. Dissonance is especially evident when different tuning systems clash.

Temperaments or tuning systems present an area in which the harmonic series can generate some problems. While numerous tuning systems have been developed throughout the world over time, the two to be focused upon here are just tuning and equal temperament. The entire point of any temperament is to standardize frequencies of musical pitches so that instruments can play in tune with one another. Just tuning is the pattern that the harmonic series follows, and is in general the most authentic to the way in which sound occurs and is processed by our ears (Sundberg 1994, 87). Equal temperament, on the other hand, is what the Western ear is most accustomed to, and is used by modern pianos. There are pros and cons of each of these tuning systems, but generally combining the two does not yield an ideal situation. Mathematically, the two systems are calculated in different ways; therefore, pitches in each of these tuning systems are often out of tune with one another. Just tuning uses the first six harmonics to tune intervals of the naturally occurring harmonics tuning intervals of octaves, fifths, major thirds, and minor thirds. This tuning follows certain ratio patterns as shown in the figure below (Sundberg 1994, 89).

Table 2: Just versus Equal Temperament Tuning

TABLE 1 Interval Comparison in Different Mathematical Tuning Systems[a]

			Just intonation			Equal temperament	
Interval name	Solfeggio	Letter notation	Numerical origin	Frequency ratio	Cents	Frequency ratio	Cents
Unison	DO	C	1:1	1.000	0.0	1.000	0
Minor second		Db	16:15	1.067		1.059	100
		C#	16:15	1.067	111.7	1.059	100
Major second	RE	D	10:9[b]	1.111	182.4	1.122	200
			9:8[c]	1.125	203.9		
Minor third		Eb	6:5	1.200	315.6	1.189	300
		D#	6:5	1.200	315.6	1.189	300
Major third	MI	E	5:4	1.250	386.3	1.260	400
Fourth	FA	F	4:3	1.333	498.1	1.335	500
Tritone		Gb	45:32	1.406	590.2	1.414	600
		F#	64:45	1.422	609.8	1.414	600
Fifth	SO	G	3:2	1.500	702.0	1.498	700
Minor sixth		Ab	8:5	1.600	813.7	1.587	800
		G#	8:5	1.600	813.7	1.587	800
Major sixth	LA	A	5:3	1.667	884.4	1.682	900
Minor seventh			1.750	968.8			
		Bb	16:9[e]	1.777	996.1	1.782	1000
		A#	9:5	1.800	1017.6	1.782	1000
Major seventh	TI	B	15:8	1.875	1088.3	1.888	1180
Octave	DO	C	2:1	2.000	1200.0	2.000	1200

[a] Adapted from Martin, 1962.
[b] Lesser.
[c] Greater.
[d] Harmonic.
[e] Grave.

Source: (Deutsch 1999, 216)

Note: Comparison of frequency ratios between two major tuning systems relevant to the harmonic series. Descending the column denotes increasing intervals.

Equal temperament was designed to break up the octave into 12 even intervals so that each semitone is an equal ratio away from one another. This ratio is $\sqrt[12]{2}$:1, where each half-step is separated by this interval (Roederer 2009, 179). Equal temperament is ideal not only for standardization of intervals, but also for allowing transposition between keys without retuning, which is not possible in just tuning. Table 2 demonstrates where the issue in mixing these two temperaments lies, in that the number values above each harmonic are the cents out of tune relative to the equal temperament system. One can see how this can be an issue when an overtone singing performer who is forced to follow the just tuning system is accompanied by an equally tempered piano, but to an extremely sensitive musical ear the harmonic series could even cause tuning issues internally in a piano alone. This is due to the fact that equally tempered intervals of the piano are not in tune with the naturally occurring harmonics each string generates; however, this is rarely perceptible, even to most musicians. More of an issue arises when one uses certain harmonics melodically together, but this is a topic developed further in Tuvan melodies in CHAPTER 4: A Study of Polyphonic Voice in Tuva.

Where timbre is concerned, it is the resonating body within which the fundamental sounds that determines the amplitude of the harmonics. The relative harmonics amplified are what give different instruments their unique tone colors. When playing an A4 on a clarinet versus, for instance, a French horn, both instruments play the same fundamental, but the relative amplitude of the certain overtones produced differs when sounding that fundamental A4 (Rachele 2013, 23). The cylindrical structure of the clarinet dampens the even-numbered harmonics, whereas the conical shape of the French horn allows for those even harmonics to be amplified, thereby giving the French horn a denser timbre. Presented in Figure 2 is a comparison of the harmonic content found in an assortment of different instruments, all playing the same A4 (440Hz) at a comparable amplitude.

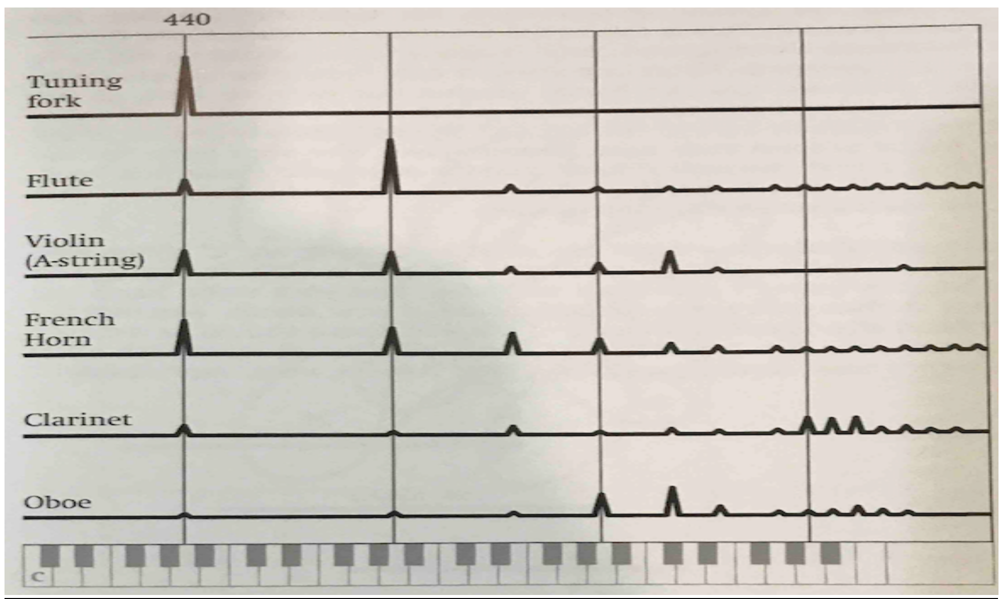

Figure 2: Comparison of Harmonic Content in Different Instruments

Illustration of the relative amplitude and resonances of different harmonics as they sound in certain displayed instruments. Harmonics are all sounding using A=440Hz as the fundamental from: (Rachele 2013, 23)

The voice is no different from these instruments. Each voice gets its unique timbre from the different overtones it produces, but overtone and throat singing take advantage of those generated harmonics within the voice. Singing of any kind, or production of any sound for that matter, is fundamentally polyphonic, excluding sine tones. The harmonic series, while integral to the polyphonic technique, plays a much bigger role in characterizing the sounds that are perceived in nature. From a bird chirping to a throat singer performing kargyraa, each possesses a distinguishing harmonic wave signature. What makes the diphonic polyphonic approach unique is the degree to which the performer amplifies the different partials, and how singers manage their control of relative partial amplitude, both of which can be explained by way of the following discussion of formants in CHAPTER 2: The Anatomy and Physiology of Sound.

CHAPTER 2: The Anatomy and Physiology of Sound

The vocal mechanism can be broken down into three overarching components: the power supply, the sound source, and the resonator. Each has anatomical correlates that together comprise the basic vocal pathway. Manipulations of the vocal tract sound sources are what facilitate bitonal polyphonic styles, while adjustments to the vocal resonators explain diphonic polyphonic vocal approaches. Simply put, the lungs are the power supply, our throat contains the oscillator that serves as the sound source, and the mouth is the resonating body (Tongeren 2002, 11). Our lungs send air up through the trachea during exhalation that serves as the energy source, and that energy gets converted into sound waves when the vocal cords begin vibrating. It is because of their oscillating nature that the vocal cords are the sound source in the vocal tract per the definition, "A sound source refers to an acoustic disturbance within a resonant environment, such as the vibration of the lips against the mouthpiece of a trombone or the oscillation of vocal folds" (Edgerton 2015, 17). Before sound leaves the vocal tract, it goes through somewhat of a filter: the mouth that acts as the resonating cavity for the sound source. As will be discussed later, especially as it pertains to formants, this filter acts in various ways to impact the final output sound as it leaves the vocal tract. While attributing these roles to these general structures may be a gross oversimplification, the concept is being used as a foundation upon which to expand as the anatomical intricacies of the vocal apparatus are explored. Since the lungs are the instigators behind vocal production, this is where this discussion will begin.

Lungs: The Power Supply

On a microscopic level, the intricacies of gas transfer in the respiratory system is critical for an organism's survival; however, the simple diffusion of oxygen and carbon dioxide molecules into and out of the lungs to be used in bodily processes such as aerobic respiration is not directly pertinent toward an understanding of the vocal mechanism. Therefore, this discussion will focus on the more macroscopic level of the respiratory system. The lungs themselves essentially act like a pump, pushing air in and out, and to do so they must expand. In order to achieve this movement, since the lungs have no inherent contractile muscle tissue, they require the aid of external musculature. Of all of the thoracic musculature, the lungs rely predominantly on the diaphragm for this task. The diaphragm separates the abdomen from the thoracic cavity, and during inhalation it works, together with the intercostal muscles located between the ribs, to contract, thereby causing overall expansion of the rib cage (Snell 1986, 73). This thoracic expansion reduces the intrathoracic pressure, allowing the lungs to expand. When the lungs expand, this creates a negative pressure inside the lungs whereby the air pressure inside the lungs is lower than atmospheric pressure. To reach equilibrium, air from the outside environment is drawn into the lungs via the upper portions of the respiratory system, the nose, mouth, and throat (Gray 1974, 355). The opposite process occurs during passive exhalation, except that during this phase, active muscular work by the muscles of inspiration is not required. Expiration occurs as both the diaphragm and external intrathoracic muscles relax, without the aid of any muscles at all so the contraction is due to the elastic recoil of the lungs (Snell 1986, 98). The exception to this is in the case of forced exhalation, where muscles of the abdominal wall and internal intercostal muscles get involved (Figure 3).

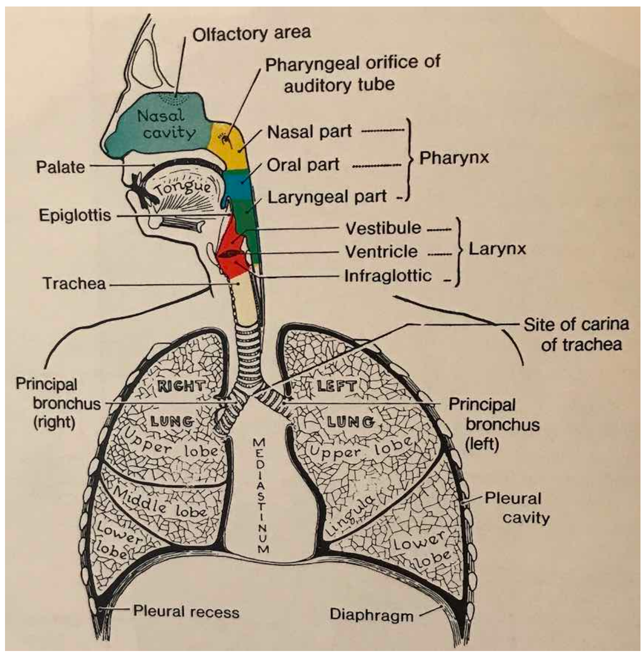

Figure 3: Respiratory Structures

From: (Moore 1985, 62)

When it comes to singing, there is a concept referred to as breath support, which is the mechanism by which vocalists control airflow as they sing. The goal is a consistent stream of air, which is achieved through simple regulation of the diaphragm. In fact, since the diaphragm is not a muscle under volitional control, it requires the help of the abdominal muscles to regulate the rate at which the diaphragm relaxes (Edgerton 2015, 12). Since the diaphragm is connected to the abdominal muscles at the epigastrium, which is a zone of the superior medial abdomen located just inferior to the rib cage, it is the action here of the rectus abdominis that is responsible for the vocal breath support mechanism (Snell 1986, 156). In this way, the abdominal muscles can voluntarily control the breathing process, by not only regulating the relaxation of the diaphragm, but also by using their ability to regulate the pressure in the abdominal cavity that indirectly impacts the pressure levels in the thoracic cavity. Breath support for singing, while crucial to vocal production in general, is even more critical in certain extranormal vocal styles to be explored in CHAPTER 4: A Study of Polyphonic Voice in Tuva, when maintenance of higher than average subglottic pressure is necessary.

As air is pushed up to the larynx through the trachea in the processing of phonation, the level of subglottic pressure is critical. During normal breathing, the vocal cords are open so that there is a free flow of air during inhalation as well as exhalation. However, during sound production the vocal cords contract together to close the glottis, generating a pressure differential in the respiratory system across the vocal cords (Grawunder 2009, 21). As more air is pushed from the lungs during exhalation, the subglottic pressure builds until a sufficient level accrues where the tension of the close vocal cords cannot hold back the air any longer, and the glottis opens (Tongeren 2002, 12). In this moment, the air molecules move to the lower pressure environment to achieve equilibrium, which happens to be up and out of the oral cavity. With the glottis open, the subglottic pressure quickly decreases until the pressure in the supraglottal space is larger, at which point the vocal cords close again until the subglottic pressure builds up again. This entire process is known as the glottal cycle, which follows a pattern described as the mucosal wave, and this phenomenon is responsible for the oscillation of the vocal cords (Edgerton 2015, 154). Also, the glottal cycle occurs over an incredibly minute temporal scale, over milliseconds typically, depending on the frequency of the oscillation.

When it comes to the vocal tract, pressure, while the main proponent impacting amplitude, has a lesser effect on frequency. The impact roughly follows the trend that, with every 0.1 kPa increase in pressure, the frequency of the pitch raises approximately two Hz (Sundberg 1994, 118). Normal speech production requires a lung pressure of about 0.5 kPa, with loud speech requiring as much as 1.5 kPa (117). As more pressure is used to speak more loudly, generally the voice increases in pitch. Therefore, when comparing speech pressure levels to singing levels, there is a noticeable difference in that a singer producing a high frequency at a dynamic level of fortissimo can use anywhere from 3-10 kPa of pressure, depending upon the individual (118). As a result of the singing having a wider range of pressure variation in singing than in speaking, the impact of pressure manipulation on pitch can be far more pronounced, which is why professional singers must go through so much training to develop full control over their vocal instruments. This is also why most non-professional or even semi-professional vocalists struggle to sing higher pitches softly, due to the fact that they compensate for extreme high pitches by greatly increasing the lung pressure used, which does increase the pitch, but at the price of also raising the volume of the tone tremendously. Such compensation, while often aurally assaulting to the listener, is really a necessity in most cases until the singer trains sufficiently to strengthen their laryngeal muscles in order to apply adequately high tension levels on the vocal cords necessary for the top pitches. The major muscles that needs strengthening in this case are the cricothyroid muscles, which will be further discussed in the throat section of this chapter.

Airflow, where singing is concerned, is typically only considered in the egressive direction; however, in certain cases an ingressive flow can be used. Egressive is simply outward airflow while ingressive is inward airflow (Edgerton 2015, 3). While not often utilized in performance, ingressive airflow can actually facilitate all the same vocal registers that typical egressive flow does, although, since it takes a great deal of training to control, a majority of singers do not explore its potential. There are a number of cases for it which it can be used, but the most practical application seems to be in the development of the whistle register, which is the highest vocal register. There are also instances in Inuit throat-singing styles where alternation between egressive and ingressive flow is seen in certain compositions (Edgerton 2015, 4). This combination of airflow in both directions presents the most interesting applications when, as long as the performers are well trained, they can perform without having to pause for inhalation, but rather can continue to phonate as they take air in.

A nasal timbre is immediately recognizable to any listener and is critical to singing in certain languages, especially French. While many subjectively consider this tone color to be moderately irritating, it is interesting to elucidate what anatomically facilitates such a timbre. When passing air through the vocal tract, the nasal cavity can be in one of three states: in use alone, not in use, or used in conjunction with the oral cavity (Edgerton 2015, 57). In the English language, the nasal timbre can be observed when producing the “ng” phoneme, which illustrates the first condition. Generating an “ng” utilizes air flow through the nasal cavity alone due to the back of the tongue touching the soft palate preventing air from traveling through the oral cavity (Figure 3: Respiratory Structures). The second condition is achieved by raising the soft palate, such as when swallowing, so that air is prevented from traveling through the nasal cavity, since the soft palate is forming a seal across the velopharyngeal port and thus air movement is strictly through the oral cavity. The third state, in which air flows freely through both nasal and oral cavity, is the condition most useful for the polyphonic voice. It is the natural state of the vocal tract, that, when singing, is thought of as forward placement (57). This air flow is ideal in certain polyphonic voices due to the fact that the sound source has access to the largest volume of resonating cavities provided by the vocal tract. The sound has the throat, nose, and mouth all to interact with, yielding the largest variety of harmonic resonances from which to choose. Essentially the sound is more harmonically rich due to more resonating spaces amplifying a wider range of the harmonic spectrum. Another benefit of this nasal flow technique is that it creates an anti-resonance around frequencies less than 400Hz (Cosi & Tisato 2003, 7). This can be very useful in that anti-resonance is capable of attenuating lower harmonics that commonly contribute to masking of the higher component harmonics. In addition, nasal flow seems to attenuate harmonics residing around F3, and therefore this method dampens not only the low harmonic, but also certain upper ones, allowing those found in F1 and F2 to be even clearer and more easily audible (7).

Throat: The Sound Source

The major constituents of the throat consist of the trachea, esophagus, larynx, epiglottis and pharynx (Figure 3). The trachea, also known informally as the windpipe since it is responsible for the passage of air between the lungs and the larynx, is connected to the larynx by the cricoid cartilage. The esophagus is the main pathway between the pharynx and the stomach, and is positioned posterior to the trachea. Ascending the throat from the trachea next comes the larynx, which is also known as the voice box, as it is the location of the vocal cords (Snell 1986, 857). By connecting the trachea to the pharynx, the larynx has as its critical function to act as a valve to prevent the aspiration of food during the swallowing process. Important to this role is the epiglottis, because it closes over the larynx during the act of swallowing, thereby diverting food into the esophagus. While part of the larynx, the cartilaginous flap that is the epiglottis divides the pharynx into two regions, the laryngopharynx and the oropharynx. Proceeding upward from the larynx lies the pharynx, which is common to both the respiratory system, as it is connected to the larynx, and the digestive system, as it is connected to the esophagus (Snell 1986, 843). Generally, the pharynx is divided into three sections: the laryngopharynx, oropharynx, and nasopharynx (Gray 1974, 890). The laryngopharynx is superior to the esophagus and larynx and inferior to the epiglottis; the oropharynx is the space behind the oral cavity between the epiglottis and the soft palate; and the nasopharynx lies behind the nasal cavities extending from the soft palate up to the base of the skull. Of all the aforementioned major structures, honing in on the larynx is the most crucial, as it is the major structure pertaining to the vocal apparatus.

Without the larynx, we would not only die from aspirating our food into our lungs, but we would also lack the capacity for spoken language. For sound production, the most critical structures are the vocal cords (VC), also known as the vocal folds. They act as our sound source, manipulating the fundamental frequency and amplitude at which we speak and sing (Figure 4).

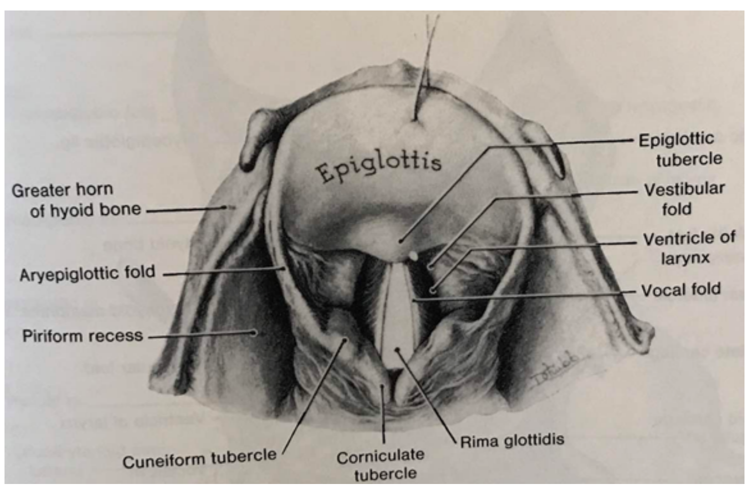

Figure 4: The Larynx

From: (Moore 1985, 1057)

The opening between the two vocal folds is known as the rima glottidis or, simply, the glottis (Moore 1985, 1057). Shaping of this aperture has a major impact upon the mode of phonation. Superior to the vocal cords are two more phonatory structures that come into play in certain polyphonic voices. Immediately superior to the vocal cords are the vestibular folds (VTF) or false cords. This structure is responsible predominantly for laryngeal protection, along with the epiglottis, during swallowing. However, as will be explored in CHAPTER 4: A Study of Polyphonic Voice in Tuva, these false folds also serve as a secondary sound source in certain styles like the Tuvan kargyraa voice (Grawunder 2009, 37). The third laryngeal contributor to sound are the aryepiglottic folds (AEF) that can be found bound between the epiglottis and arytenoid cartilage. Vibration of the AEFs is commonly seen in growl type singing, as can be heard within the vocal stylings of Louis Armstrong. In the average singing voice, neither of these two additional phonatory structures contributes to the source sound, but their activation is what contributes to a number of polyphonic vocal styles. Where amplitude of the sound source is concerned, a manipulation of subglottic pressure is the determining factor as it is this pressure that transfers oscillating power.

The musculature of the larynx is quite complex, and this discussion will focus only on those muscles relevant specifically to vocal production, referred to as the intrinsic muscles of the larynx. Where the vocal cords are concerned, there are five major muscles responsible for movement, which includes the position, length, and tension on the vocal cords. These are the cricothyroid (CT), thyroarytenoid (TA), lateral cricoarytenoid

(LCA), transverse arytenoid (TVA), and the posterior cricoarytenoid (PCA) (Snell 1986, 862). This list can be subdivided into two groups. The first consists of those that impact fundamental frequency, the CT and the TA. The second group consists of those affecting the glottis aperture, the LCA, TVA, and PCA, thus impacting the mode of phonation (Figure 5).

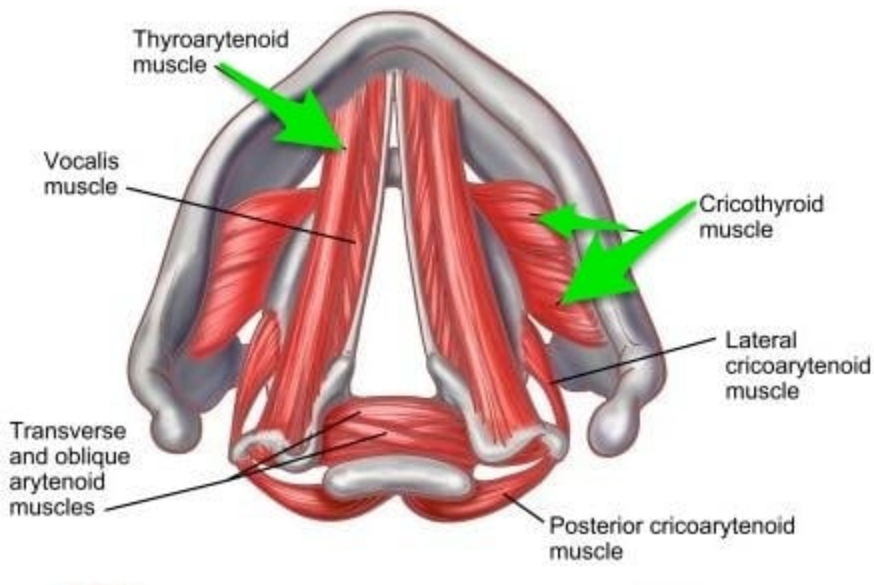

Figure 5: Muscles of the Larynx

From: Karmyn. 2016. "Your Vocal Cords." KT Vocal Studio. ktvocalstudio.com/vocal-cords/.

Inevitably, two main variables affect the frequency of an oscillator such as our vocal cords: tension and length. The first group of muscles manipulates frequency by changing the tension and length of the vocal cords. When long and tight with sufficient subglottal pressure applied to initiate phonation, the cords oscillate at a higher frequency.

Therefore, in order to generate a lower pitched fundamental, the cords must be short and loose. Connecting the cricoid cartilage to the thyroid cartilage is the CT, which is the lone tensor muscle in the larynx (Moore 1985, 1061). This is the muscle responsible for increasing the length and tension of the VC by pulling the thyroid cartilage forward at a tilt, thus raising the pitch. To elicit the opposite effect on pitch the larynx uses the TA muscle. The TA connects the thyroid cartilage to the arytenoid cartilage, also serving as the general body of the VCs themselves, with a thin mucous membrane covering the muscle. The fibers of the TA that act as the body of the VC are more specifically known as the vocalis muscle, while serving as part of the TA muscle (Snell 1986, 862). The TA lowers tension and shortens the VC by bringing the arytenoid cartilage forward towards the thyroid cartilage, creating a lower frequency (Figure 6).

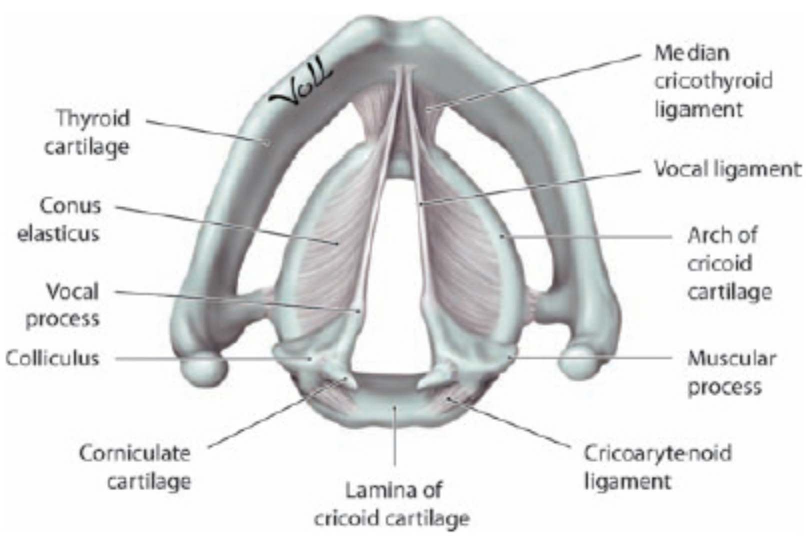

Figure 6: Cartilage of the Larynx

From: Kendall, Katherine. 2016. “Laryngeal Anatomy.” Ento Key. entokey.com/laryngeal-anatomy/.

Modes of phonation make a huge impact upon the created vocal timbres as well as register of the generated sound. The main modes of phonation vocally are classified as pressed, flow, or breathy, which are determined by the percentage of adduction (i.e., moving together) of the vocal folds (Edgerton 2015, 155). A greater adduction percentage of the VC is associated with smaller glottal opening, which is the case for pressed phonation. Flow phonation is attributed to a moderate adduction percentage while breathy phonation has the lowest value. The second group of muscles is responsible for manipulating adduction percentage of the VCs. Beginning with the LCA, this muscle attaches the lateral arch of the cricoid cartilage to the muscular process of the arytenoid cartilage (Moore 1985, 1061). Action of LCA muscle yields adduction of the VC by pulling the muscular process of the arytenoids cartilage forward, resulting in the rotation of the arytenoids. The TVA also yields adduction of the vocal cords, but by a different motion. Connecting the arytenoid cartilages, the TVA works by bringing the arytenoids together causing closure in the posterior section of the glottis. Finally, the PCA works antagonistically to the LCA as the mechanism for abduction (i.e., moving apart) of the VCs (Snell 1986, 863). The PCA muscle connects the posterior cricoid cartilage to the muscular process of the arytenoids and pulls that process backward, resulting in rotation of the arytenoids moving the larynx laterally, causing the abduction of the VCs. So, with an understanding of the roles of each of these muscular groupings, the anatomy can be applied to how vocal registers are produced.

Singers understand the different vocal registers and what facilitates different frequency ranges of the voice, but now the anatomy as discussed can be applied to practice in registration. Depending upon whom one asks, different vocal registers go by different names that often refer to the same part of the voice. The lowest frequency vocal register is termed the vocal fry (Kent & Ball 2000, 5). In this register the VCs are at their lowest tension level due to action of the TA, resulting in a closed glottis due to the adduction of the LCA and TVA. With this loosely closed glottis, glottal pulses are infrequent as air makes its way with low frequency through the VCs, creating the cracking timbre associated with this register, which is why vocal fry is also known as creak voice. One caveat of the vocal fry as will be explored in CHAPTER 4: A Study of Polyphonic Voice in Tuva is that in many cases the aryepiglottic folds accompanies VC oscillation.

The modal or chest register falls next highest in the vocal range, and is the register associated with a person's typical singing voice. With this register, there is consistent oscillation of the VC with a moderate amount of tension and a moderate level of adduction of the cords. Due to the moderately low level of tension found in this register, a higher activation of the TA is still observed, but there is also a balancing act where some opposing CT activation occurs. It is also noteworthy that, as the pitch is raised, muscular activation shifts more in the favor of the CT with less activation of the TA muscle. A similar case can be made for the muscular workings responsible for the moderate level of adduction seen in the VCs in the modal voice. It is once again an equilibrium of muscular activation where the LCA and TVA activities are greatest in the lower range, but PCA activation takes over as the frequency gets higher, abducting the glottis more.

The third register is the falsetto or head voice, which is even higher than modal voice, featuring glottal abduction, in which there is incomplete closure of the glottis during each wave of the vocal folds (Edgerton 2015, 155). Muscular activation in this register follows the previously presented trend in which the CT activation increases as pitch increases and PCA activation also increases when the opening to the glottis is the largest out of all the registers thus far. There also exists a fourth vocal register known as whistle voice, which, although little is known with certainty about what exactly occurs physiologically, is the highest point of the vocal range. Often most easily produced using an ingressive airflow, this register got its name due to the high flute-like whistle timbre produced (Holmes-Bendixen 2013, 34). It is believed that this range features a consistently open glottis along with vibrating VCs, but specific structural movement is often unclear, since little can be visualized using a laryngoscope due to constriction of the superior portion of the larynx by the epiglottis.

Mouth: The Resonator

The mouth serves two main purposes in the body by acting as a branch of both the digestive and respiratory systems. First and foremost, it serves at the initial station for processing of food in the digestive system (Snell 1986, 838). Second to that is its role in the respiratory system, as an alternate pathway for air flow where the nasal cavity is the primary route. As a byproduct of the mouth's role in the respiratory system, which is the main system within which the vocal tract resides, the mouth serves a critical role in speech production. In fact, without the mouth phonetic speech would be impossible for our species, for reasons that will be expanded upon in the formant discussion later in this chapter. The gross anatomy of the mouth will help in defining structures pertinent to the manipulation of formants.

In humans, the mouth extends from the lips all the way to the oropharyngeal isthmus, which is the posterior opening between the oropharynx and the mouth. The mouth as a whole can be divided into two main cavities. The first is termed the vestibule and the second the oral cavity proper. Dividing the two cavities is the alveolar ridge that contains the teeth: the vestibule is the space located between the teeth and gums internally, and the lips and cheeks externally (Snell 1986, 840) (Figure 7).

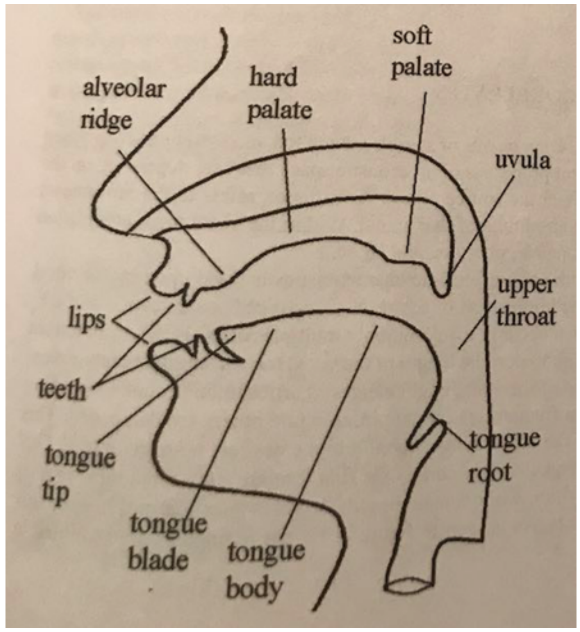

Figure 7: Structures of the Mouth

From: (Edgerton 2015, 157).

The oral cavity proper is located between the teeth anteriorly and the oropharyngeal isthmus posteriorly. Note that the roof is formed anteriorly by the hard palate and posteriorly by the soft palate, while the floor is formed by the mylohyoid muscle with the front two-thirds of the tongue resting on top, further defining the borders of the oral cavity proper. Descending from the middle of the soft palate is the uvula. Articulators, also known as speech organs, are the main oral structures pertinent to speech production. While not all of the articulators are specifically located in the oral cavity, the majority are, including the lips, tongue, teeth, soft palate, hard palate, alveolar ridge, and uvula. Additionally, the previously discussed glottis is also classified as an articulator. The manipulating of each of these structures and their relationships to one another are what facilitates the wide variety of phonemes humans are capable of producing. Specific maps of locations within the mouth, termed lingua-palatal maps as seen in Figure 8 (Edgerton 2015, 52), are used by some vocalists in order to characterize the orientation of the tongue in the mouth in relation to the other articulating structures of the mouth. This allows for clarity in instruction for a performer to generate a certain shaping of the resonating cavity often associated with certain vowel and consonant placements.

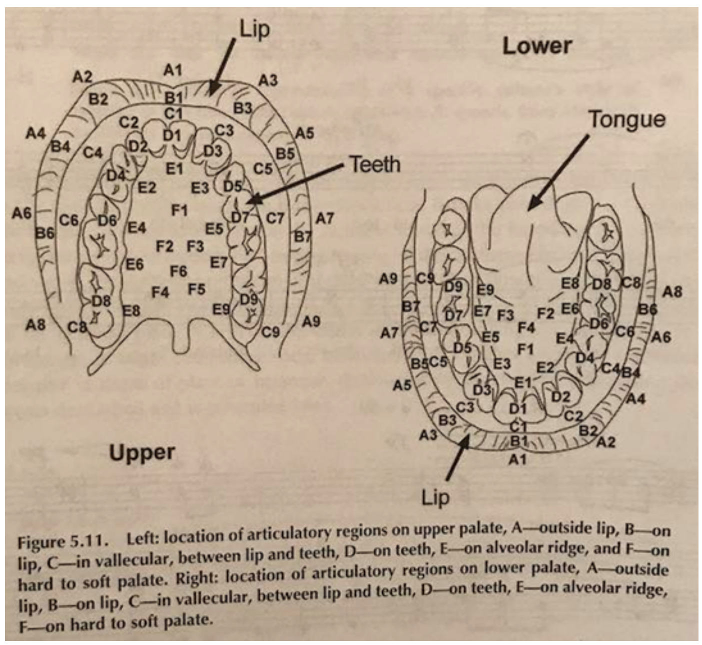

Figure 8: Lingua-Palatal Map

A labeled map to define specific articulator positioning for the production of certain formants within the oral cavity from: (Edgerton 2015, 52)

Formants

Depending on context, the meaning of the word "formant" can vary slightly. The most encompassing definition is, "A broad resonance region that enhances the harmonics lying in a fixed frequency range…," a definition which provides a technical perspective (Roederer 2009, 134). As such, formants act as a filter for a resonating body, amplifying certain harmonics of a fundamental frequency. Discrepancies, however, do not really lie in that definition precisely, but rather in what resonator is used. Some choose to limit their scope by only using the vocal tract as a resonator where formants are concerned, saying that no other instruments exhibit this phenomenon (Sundberg 1994, 121), which in my opinion is quite limited. While I will focus on the application of formant in the vocal tract, for the purpose of this paper in reality any environment where sound interacts with a physical structure possesses formants, whether it be a brass instrument or even a room in a building. In fact, double reed instruments such as the English horn and bassoon have a defined formant at 450 and 1100 Hz respectively (Roederer 2009, 147). Aside from human egocentrism, there are a few simple explanations as to why there has been so much research focus into the categorization of the formants of the vocal tract. The first is due to the high degree of control over the resonators in our vocal tract that can be mastered with training. The second is that one of the major defining behaviors of the human species making us unique is speech, and while a majority of the population lives unaware, formants, especially where vowels are concerned, facilitate the characteristic phonemes of language. However, before delving into formants and their application to the voice, it is crucial to dispel some common misconception regarding formants especially with regard to their relationship to harmonics.

In practice, harmonics and formants are extensively interrelated, but this can lead to confusion. Most important to note is that formants do not have a direct aural manifestation in the same way that harmonics do, which is key to differentiating formants and harmonics. In fact, formants do not even really exist, per se, without a sound source and a resonating body. Essentially formants are a filtering tool for amplification of harmonics whose existences are inferred by spectral analysis of, say, a voice possessing specific harmonics that are amplified (McCoy 2013, 45). Therefore, the harmonics are always there in a sound, but by using a certain formant manipulation, a louder, say 3rd harmonic, can be achieved from the fundamental. A great way to think of this is that it is as if "harmonics are the paint; formants are the artist's brush that helps create the masterpiece" (45). An understanding of what a formant is and is not will allow us to explore the previous query into the formant's role in human speech.

Formants play a major role in language production. A larynx is a necessity for speech, but what exactly do vocal folds do to allow us to produce different words? Breaking down the components reveals that the larynx really only provides the source sound, the pitch and amplitude. If one were to take the larynx of Andrea Bocelli and transplant it into one's own throat, the resulting voice would most likely sound identical timbrally to how the recipient's voice sounded before the transplant. If not within the vocal tract, the sound generated by the larynx is described as a buzzing and nothing more, where one individual's larynx sounds relatively similar to any other human's (McCoy 2013, 43). Putting the larynx back within its native environment, where a generated source sound can interact with the supraglottic space, nasopharyngeal cavity, and oral cavity, is how humans are able to generate the unique timbre of individual speech. The vocal tract is a complex resonator in which numerous alcoves exist, allowing the sound to interact with each of these resonating spaces, whether they are located under the tongue or even around the lingual tonsils. What the articulators of the vocal tract do is manipulate the shape of those resonating cavities to essentially generate formants. From this point forward, the term formant will refer to a resonance of the vocal tract specifically. Where formants are concerned vocally, a specific focus on vowel formants provides the clearest picture.

Technically there are an infinite number of formants in the vocal tract, but where vowel speech is concerned, the first two are the most pertinent, with the third formant also playing a role. In this discussion of vowels, pure vowels [i e a o u] will be used to compare frequencies of vowels in different formant placements. With the first formant (F1), the frequency range of vowels in an average male voice generally ranges from between 150 to 900 Hz. For the second formant (F2), the range is between 500 and 3000 Hz, and the third (F3) is from 1500 to 4500Hz (Sundberg 1994, 122). In each of these formant frequency ranges, the pure vowels have a specific frequency at which the orientation of the articulators in the vocal tract facilitates the purest vowels. Essentially the tongue, lips, teeth, jaw, palates, and other structures all assume certain positions to achieve each vowel sound, and those positions are associated with changed resonant spaces in the vocal cavity where each certain harmonic is amplified. For F1, the pitch centers for each vowel outline a first inverted open position C-major triad, and they are as follows: [i]= E_4 [e]= C_5 [a] = G_5 [o]= C_5 [u]=E_4. For F2, aligning the pattern in a reverse order as follows: [u]= E_5 [o]= G_5 [a]= D_6 [e]= B_6 [i]= D_7 presents a useful mnemonic of "**E**very **G**ood **D**ad **B**uys **D**iapers" (McCoy 2013, 45). While these are accurate average approximations of the tonal centers for each vowel in F1 and F2, they are just that: averages. Depending on the individual vocal tract anatomy, these pitch centers can vary as much as a major 3rd away from average in either direction (45). For short stature individuals who, in turn, tend to have a shorter vocal tract proportionally, their formants are generally higher in pitch, and the opposite is true for a taller person. To see the articulator positions for each of these vowels in F1 and F2 refer to Figure 9.

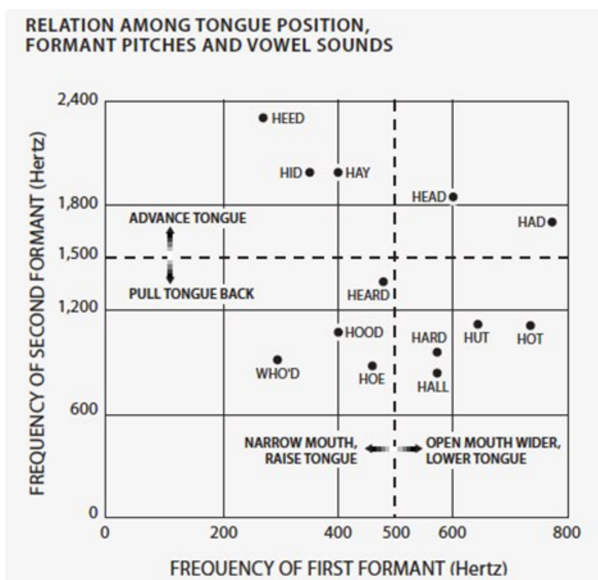

Figure 9: Articular Positions

Mapped F1 and F2 by frequency using the different harmonic frequencies generated by the different English vowel placements from: (Levin & Edgerton 1999, 83)

However, if one of the articulators is misaligned, such as the root of the tongue being too far anterior when attempting to generate the vowel sound [u], the vowel generated will have more [i] character due the vocal tract being oriented in such a way that the second formant frequency has been raised. A case such as this results in a vowel quality that, while it may be desired in some cases, would lack purity if the intention was the generation of an [u] vowel. In order to understand this concept, it is necessary to delve into the physics of tube resonators such as our vocal tract.

While imperfect, the vocal tract is in essence a tube resonator with one closed and one open end. The vocal folds would be considered the closed end, whereas the opening is at the lip aperture. While not entirely accurate, making a few assumptions, where the vocal tract is uniform in cross section rather than tapering along the length, and is roughly 17.5 centimeters long, makes calculating the predicted formant resonances much simpler (Levin & Edgerton 1999, 83). The first assumption is made in order to have resonant frequencies be strictly determined by the length of the resonating tube, although rounding up to 17.5 from the average vocal tract length of about 17 centimeters in the average male is more ideal mathematically, considering the speed of sound at body temperature is roughly 350 m/sec. After completing the calculations, the resonance peak of F1 corresponds to 500Hz, F2 corresponds to 1500Hz and F3 corresponds to 2500Hz, which confirms the experimental ranges previously provided for these first three formants (Sundberg 1994, 120). Next, one must consider the oscillation pattern of air molecules and pressure at each resonance, and how changes in each of these aspects impacts the other as well as frequency in the formant. In such a resonance system, understanding first what pressure nodes and antinodes are, and then what constriction at these points does to the resonating frequency, sheds further light as to the nature of these formants. The pressure node is the point at which the pressure in the system remains constant during which the air molecules must travel their largest distance, whereas at pressure antinodes, the pressure fluctuates to its greatest degree while the air molecules remain stagnant (Levin & Edgerton 1999, 83). Consider also that constriction at a pressure node causes a decrease in frequency, whereas constriction at a pressure antinode raises the frequency. The nodes and antinodes of F1 and F2 are illustrated below (Figure 10).

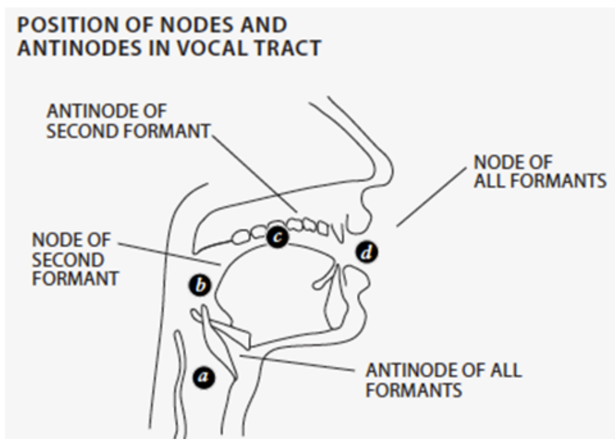

Figure 10: Nodes of the Vocal Track

Sagittal depiction of the mouth cavity into the throat demarking the locations of the nodes and antinodes of F1 and F2 from: (Levin & Edgerton 1999, 83)

Lowered frequency is seen at a pressure node since the molecules of air needing more time to flow through a smaller aperture, whereas pressure remains constant on either side of the opening; therefore, the frequency of the wave slows down with the result being a lower pitch. This can be seen in the second formant from the vowel [i] to the vowel [u], as the root of the tongue constricts the oropharyngeal isthmus, which is the pressure node of the F2. In the inverse of that vowel comparison, looking at [u] to [i], there is constriction at the antinode of F2 for [i]; therefore, by raising the front of the tongue to about the middle of the hard palate, this causes the frequency of F2 to rise (83). At the antinode, an increase in frequency is observed due to pressure fluctuation in a smaller volume, increasing the density of the air so the sound wave can travel faster with molecular air motion held relatively constant.

Formants are not strictly limited to vowel formants. Another well-known formant exists that, although it does not possess a numerical demarcation like F1 or F2, serves a critical function for the professional singer. Known as the singer's formant, when analyzed using a spectrograph, this formant actually behaves like a combination of the spectral peaks located around the frequencies corresponding with formant three (F3), formant four (F4), and formant five (F5) (Sundberg 1994, 125). These spectral peaks fall within a frequency range between 2000 to 3000Hz, which is where the benefit of this formant comes in for singers. It is quite common for singers, especially those performing opera, to be required to sing over an orchestral accompaniment. A group of 20 or more instruments is a lot of sound to compete with, so whether knowingly or not, most trained opera singers have learned a specific vocal approach that allows them to project over an orchestra by using the singer's formant. Physiologically, these singers are widening the laryngopharynx and are also slightly constricting the supraglottic space, most likely by squeezing the arytenoids together (Lee et al. 2018, 92). This is similar to the squeezed voice technique utilized in throat-singing techniques, further described in CHAPTER 4: A Study of Polyphonic Voice in Tuva. This results in is the increased resonance of high harmonics in the 2000 to 3000Hz range. Considering that the majority of the frequencies generated by the instruments in an orchestra, with few exceptions, fall roughly between 100Hz and 2000Hz, the amplified harmonics that the singer's formant facilitates are at a frequency level so high that they do not have as much to compete with (Figure 11).

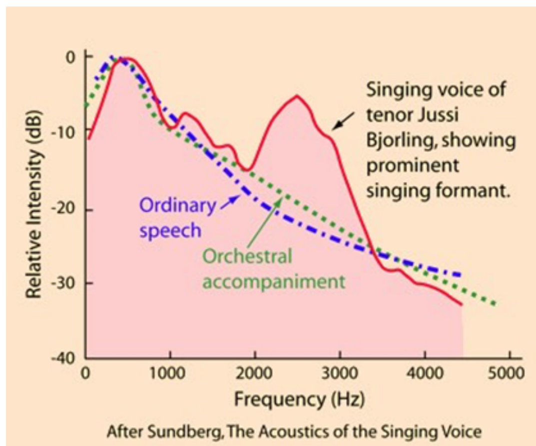

Figure 11: The Singer's Formant

Comparison of the loudness throughout the harmonic content of a singer as compared to either orchestral accompaniment or ordinary speech. The higher relative intensity in the 2000-3500Hz frequency range is an indication of the reinforced harmonics generated due to the singer's formant from: Sundberg, Johan. 1977. "The Singing Formant." Sundberg's Singing Formant. hyperphysics.phy-astr.gsu.edu/hbase/Music/singfor.html.

Therefore, a singer can be more easily heard above the orchestra, and these high overtones are what give singers who use this technique a "shiny" timbral quality (Sundberg 1994, 125).

As helpful as formants can be for singing, they are also responsible for certain vocal occurrences of concern. Soprano vocalists actually run into an issue because of vowel formants. Word and text clarity are crucial when singing so that the listener can understand what is being communicated. There is no denying that consonant clarity is crucial to get diction across; however, vowel clarity is equally, if not more, important. As mentioned, F1 and F2 are critical in vowel purity, but when sopranos get to the top of their range, these formants actually work against them. The problem begins to arise around a fundamental F5 at about (700Hz) due to the nature of the resonance frequencies of F1 and F2 (Sundberg 1994, 127). Vowels sung above this point begin to suffer clarity deficits, especially as concerns intelligible vowels, as the pitch increases. This is because each vowel sound's integrity is dependent upon the amplified harmonics that each of these formants facilitates. Looking back at Figure 9**,** it is clear that vowels like [u] and [o] in the second formant have resonance frequencies around 650 to 750 Hz. It is the case that the fundamental frequency sung around or higher than this range simply does not possess the harmonics that are necessary to characterize the [u] and [o] vowel sound. This leads to muddy-sounding vowels, which only becomes more of a problem as the fundamental gets higher, impacting more vowels as the fundamental surpasses the vowel resonance frequency in F2. In cases in which singers are asked by the composer to sing in this range, the composer often takes into account this phenomenon by either assigning text for which intelligibility is less important, or, in a choral setting, having other vocal parts compensate for the sopranos' diction. It is notable that, while it mainly impacts sopranos and altos, high tenors also face this problem.

Formant application to vocal harmonics is paramount toward understanding the polyphonic voice. Reinforced harmonics are the means by which most polyphonic vocal techniques are possible. Whether in Western overtone singing, throat singing, or countless other traditions, formant manipulation is required to achieve amplification of the desired harmonic. Edgerton describes four different methods by which certain articulator manipulation in the vocal tract facilitates harmonic amplification (Edgerton 2015, 62). Which method is used by a specific polyphonic performer depends upon style, the culture influences, individual preference, or some combination of the three. Method one (M1) focuses on changing the aperture of the mouth at the lips, so that the harmonic being isolated increases as the opening gets larger. Method two (M2) places the tip of the tongue at the alveolar ridge, behind the front teeth, and then, as the mid-tongue moves up towards the hard palate, the harmonic increases. Method three (M3) is easiest to comprehend as a transition between the vowels [o] to [i], where the root of the tongue is positioned posteriorly constricting the oropharynx as well as the oropharyngeal isthmus at low harmonics, and then the entire tongue advances anteriorly with the tip touching the hard palate for higher harmonics. Finally, method 4 (M4) keeps the tip of the tongue constant at the transition between the hard and soft palates and increases the isolated harmonic as the root of the tongue moves away from the wall of the pharynx (Figure 12).

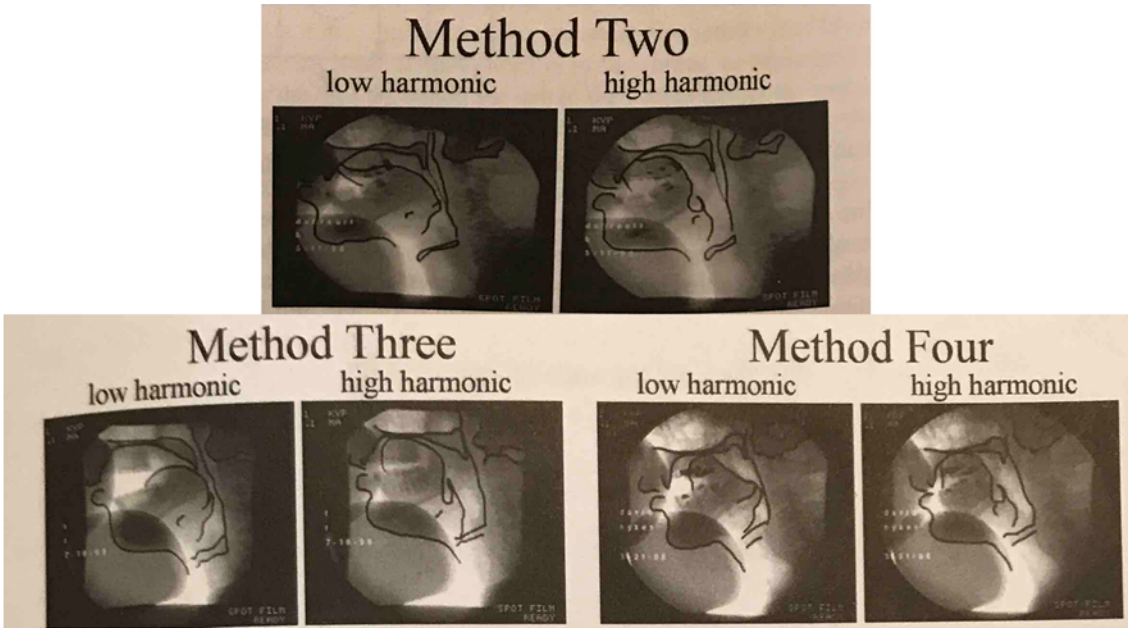

Figure 12: Tongue Position Methods

Different methodological placements of vocal tract articulator situated for harmonic amplification. Method one has been excluded due to an error present in the source, but is detailed within the text from: (Edgerton 2015, 62).

Each of these methods depicted has its own merit. Whereas some generate higher amplitude harmonics at the cost of dampening the fundamental frequency, such as M2, others sacrifice clarity of the harmonic to maintain the integrity of the fundamental, such as M3. Specifically, M2 and M3 are used as the preferred approaches in certain Tuvan throat-singing styles, to be further explored in CHAPTER 4: A Study of Polyphonic Voice in Tuva. These four harmonic reinforcement methods all use internal filtration of the vocal cord sound source; however, examples of using external filters, such as the hands or an instrument, to amplify specific vocal harmonics have also been observed in a variety of global cultures. One well-known example of this is the Australian didgeridoo, while a less popular instance can be heard in the flute/voice songs produced by the Khmu people from Laos (Edgerton 2015, 65).

Whereas Edgerton presents us with these four methods of harmonic reinforcement, there is actually a simpler categorization achieved by classifying how many major cavities are being formed in the mouth. These classifications are denoted as the one- and two-cavity methods. The one-cavity method (1CM) is simply when the unitary nature of the vocal tract maintains its continuity without the tongue subdividing the resonating space, so that the mouth only has one major resonating cavity (Cosi & Tisato 2003, 5). The 1CM in essence is identical to the M1 described by Edgerton. What is useful about the 1CM is that it allows for greater selection of harmonics located in the resonance frequency range of F1. The two-cavity method (2CM) is an approach in which the tongue is used to subdivide the vocal tract at some location. M2, M3, and M4 proposed by Edgerton all fall into this 2CM category. The advantage of the 2CM is that the F2 can be used to reinforce higher harmonics, thus presenting the polyphonic voice with access to any partial from the 6^{th} to 20^{th} harmonic as long as a low enough fundamental is chosen (Tongeren 2002, 20). Tuning in this method is more difficult due to the necessity in some cases to tune the harmonics generated in each resonating cavity to one another.

Part of the challenge in mastering these complex vocal styles can be addressed with a conceptual understanding of how the vocal mechanism is laid out and functions. When attempting to produce, for example, a bitonal polyphonic style, such as the umngqokolo singing style of the South African Xhosa tribe, it helps to visualize the physiological mechanisms inherent in the specific polyphonic style (Dargie 1988, 56). This style of voice is subharmonic, generating a bass-like sonority similar to the Tuvan kargyraa that will be discussed later, and is traditionally sung by the females of the Xhosa so that they sound like men due to the low frequency sounds they produce (Ken-Ichi et al. 2007, 3). In this case, the two different oscillators vibrating are the VC and the AEF, similar to the production of growl singing technique. Therefore, knowing that the AEF are superior to both the VC and the VF, gives a singer a general frame of reference as to where to feel the laryngeal oscillation. This is but one example illustrating how an analysis of the relevant anatomy gives insight into and a greater appreciation for each distinct polyphonic approach. When assessing Tuvan polyphonic singing, the focal culture to be addressed, a similar detailed review is presented. Therefore, the distinguishing changes made within the vocal apparatus, whether manipulating the oscillating sound sources or the resonant cavity, can be identified in each separate throat-singing voice.

CHAPTER 3: The Neurobiology of Sound Perception

Auditory System

The age-old question, "If a tree falls in the middle of the woods and no one is around to hear it, does it make a sound?" is more philosophical than a rhetorical question acting as a means to confront the concept of perception. Witnessed or not, the fundamental laws of physics still apply, so sound would surely occur. However, the perception of a stimulus is highly variable even when one is provided with an identical stimulus. While auditory perception allows appreciation for the polyphonic voice, this sense evolved for other purposes. First and foremost, our auditory system serves to facilitate survival in humans and numerous other species. The auditory system serves the purposes of detecting inter and intra species stimuli as well as detecting sound location in space (Kandel et al. 2013, 683). Like other sensory systems, it allows us to process information about the state of our surrounding environment. It informs us as to where a sound stimulus originates spatially, in addition to allowing us to deduce the source of the sound. For example, in the case of a human hearing a lion roar, our auditory system alerts us to a dangerous predator; we then integrate that input stimulus with other sensory information and cognitive processes to respond accordingly. However, perception does not always remain faithful to the initial stimulus, and certain biological aspects of the auditory system impact human cognition of polyphonic sounds. Specifically, when listening to the polyphonic voice, perception may not always be representative of the initial sound source as will be noted in the discussion of the missing fundamental effect and combination tones.

A general overview of the human auditory system is in order, before considering its interface with music. The auditory system is divided into two main components: the peripheral and the central auditory systems (Purves et al. 2012, 279). These two systems can be further subdivided with the outer, middle, and inner ear making up the peripheral auditory system and the central auditory pathway beginning at the cochlear nucleus, and traveling to a number of different way stations on the path to the auditory cortex. In connecting these two divisions of the auditory system, humans transform sound stimuli into neural activity that gets integrated with other sensory modalities, leading to behaviors such as intraspecies communication (279). This capability of intraspecies communication capability has developed in humans to the point of language, one of the most defining behaviors of our species. Review will begin with the most exterior auditory structure, the outer ear.

Outer Ear

Referred to as either the outer or external ear, it is the only portion of the auditory system visible on the outside of the body. The two main structures that make up this portion of the ear are the pinna and external acoustic meatus. The pinna, also known as the auricle, is the outer cartilaginous portion seen bilaterally on either side of the head, which is connected at the concha to the external acoustic meatus or auditory canal (Figure 13) (Moore 1985, 964).

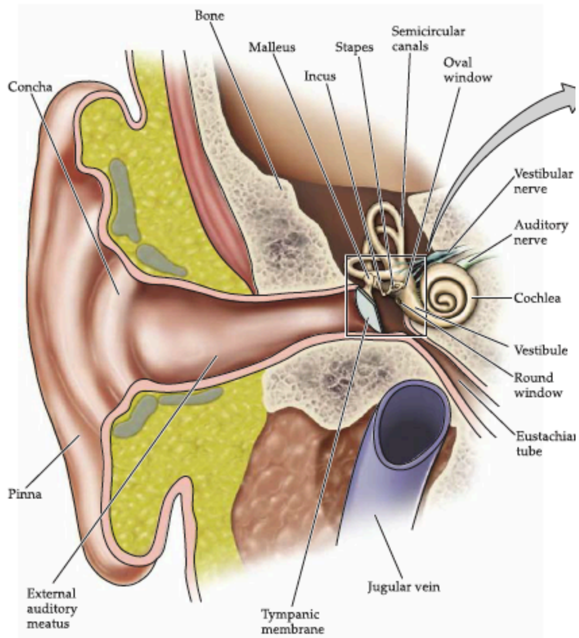

Figure 13: Auditory Structures

From: (Purves et al. 2012, 282)

While the shape of the pinna seems random, it is quite the opposite, as it acts like a funnel with its connection to acoustic meatus, aiding in the selective amplification of sound pressure in air. Specifically, it boosts sounds in the 3000 Hz range by a 30-100 fold margin onto the tympanic membrane (Purves et al. 2012, 281). The human auditory frequency range generally is from 20 to 20,000 Hz on average. But the external ear anatomy developed to amplify sounds around that frequency mainly for the purpose of interspecies communication, which in humans tends to be within a range of 100 Hz to 4000 Hz.

Middle Ear

Sound pressure waves travel through different mediums in the ear. Thus they originate as air vibrations and remain that way until the inner ear, where they are transmitted into the liquid within the cochlea. When sound is transmitted from a medium of low impedance such as the air to a medium or higher impedance such as liquid, a majority of the acoustic energy is lost due to reflection (Purves et al. 2012, 282). Therefore, the middle ear's main duty is to account for this change in impedance, and it does so in two ways. The major structures contributing to these processes are the tympanic membrane, fenestra vestibuli (or oval window), and the ossicles. The tympanic membrane serves as the border between the outer and middle ear, whereas the oval window is a site of connection between the middle and inner ear. Connecting these two structures are the ossicles, which are the three tiny bones within the middle ear: the malleus, incus, and stapes (Figure 14). The first method of impedance matching is a result of the tympanic membrane having a much larger area, roughly 17 times larger, than the oval window and the fact that sound pressure force acts on a smaller area as it moves between these two structures amplifies the pressure (Snell 1986, 835).

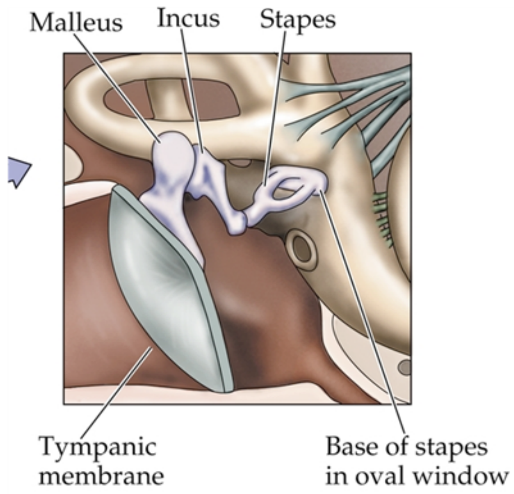

Figure 14: The Middle Ear

From: (Purves et al. 2012, 282)

The size ratio of these bones presents the second method of impedance matching since these bones act to increase the amplitude of the sound due to the 1.3:1 leverage increase demonstrated (Snell 1986, 835). The sound pressure resonating on the tympanic membrane, when conveyed to the inner ear and accounting for each of these mechanical advantages, increases 200 fold in the middle ear, thereby overcoming that impedance differential (Purves et al. 2012, 282). The middle ear is not only capable of amplifying a sound; it also has the means to dampen a sound using the tensor tympani and stapedius muscles. The tensor tympani, which is attached to the malleus, is innervated by the trigeminal nerve or cranial nerve V (CN V), and constriction of this muscle pulls the malleus medially, causing tension on the tympanic membrane, thus limiting the amplitude at which the membrane can oscillate (Moore 1985, 971). Attached to the stapes, the stapedius is innervated by CN VII, and when contracted, it tightens the angular ligament of this bone, also reducing the range oscillation. Contraction of these muscles occurs as a protective measure in response to very loud noises, or even during personal vocalizations, so as to prevent hearing loss in the inner ear.

Furthering our consideration of impedance can help to explain the human ability to hear sound through touch. This can be demonstrated when a vibrating tuning fork is placed against one's temple, or when we listen to the sound of our own voices. What we are assessing is the concept of bone conduction. In the case of the tuning fork, if it is vibrating at an inaudible frequency, but is still vibrating, the sound can still be "heard" (Sundberg 1994, 43). This is due to the relative impedance similarity between those vibrating bony tissues and the inner ear fluid (Purves et al. 2012, 283). In this way, what is said can still be "heard" without the mechanical advantage supplied by the middle ear. Therefore, when people say, "I can feel the sound in my bones," this is the predominant reason explaining that phenomena. Bone conduction also accounts for why our own speaking voice sounds quite different when heard in a recording. We are used to factoring in the vibrations through our own bony tissues, as well as the externally produced sound into the cognitive conception of our voice, but in a recording, we do not experience any of the internally generated sound.

Inner Ear

While the inner ear also plays a role in maintaining balance via the vestibular system, here we are predominantly concerned with its role in sound processing. The major auditory component of the inner ear is known as the cochlea, which is debatably the most intricate and critical of the external auditory system structures. It is a hollow, snail shell-shaped bone that is roughly 30mm long from the base to apex when unraveled, which is embedded into the temporal bone (Gray 1974, 861). Within this structure there are three liquid-filled compartments that are known as the scala vestibuli, the scala tympani, and the scala media (Kandel et al. 2013, 656). The oval window that connects the middle to inner ear at the basal end of the scala vestibuli. Dividing the scala vestibuli from the scala tympani, almost to the apical end of the cochlea, is the cochlear partition. However, the cochlear partition terminates before the apex of the cochlea, it is at this point known as the helicotrema connecting the scala vestibuli and the scala tympani. Within that cochlear partition is the scala media that contains the primary hearing receptor organ known as the organ of corti (Purves et al. 2012, 286). When sound transmission causes the oval window to oscillate, the compression of that membrane causes the perilymph fluid to displace traveling through the cochlea, that fluid motion is what facilities the organ of Corti’s ability to transduce sound into the electrochemical energy used by neurons (Figure 15).

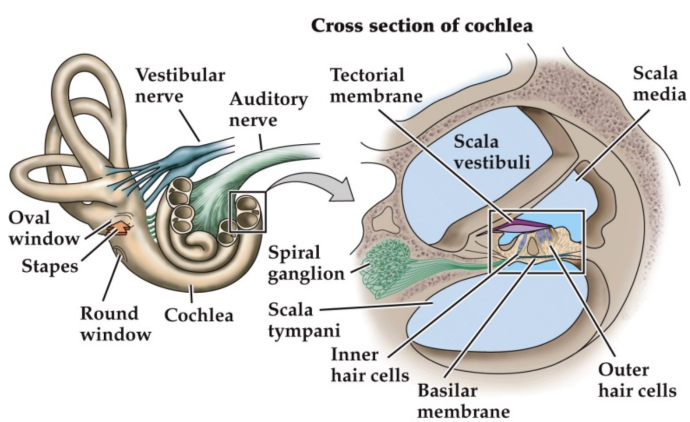

Figure 15: The Inner Ear

From: (Purves et al. 2012, 285)

The inner hair cells and other structures within the organ of Corti to facilitate this transduction mechanism. The organ of Corti itself is a layer of epithelium that extends across the basilar membrane, which divides the scala tympani from the scala media (Kandel et al. 2013, 660). Within the organ of Corti lies one row of inner hair cells and three rows of outer hair cells which extend into the scala media (Figure 16).

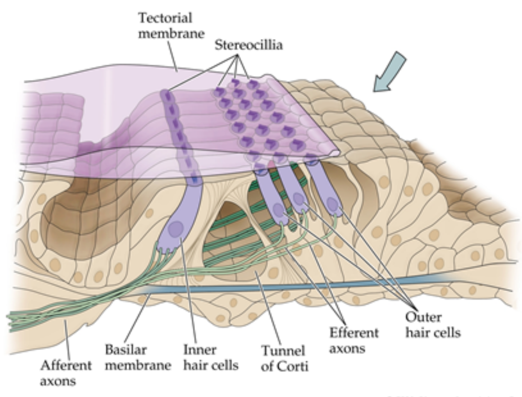

Figure 16: The Organ of Corti

From: (Purves et al. 2012, 285)

The outer hair cells serve to mechanically tune and even amplify sounds in the cochlea, which will be considered below in reference to difference tones. In contrast, the inner hair cells are the sensory receptors of the auditory system due to their transductive role. They make up roughly 90% of the vestibulocochlear projections to the brain. They are epithelial cells that have bundles of stereocilia, known as hair bundles, extending from their apical end into the endolymph of the scala media. The hair bundles are what give these cells their name, and the stereocilia have varying lengths within the hair bundles. The stereocilia of these hair bundles are in contact with the tectorial membrane, which lies above the hair cells also within the scala media. Fluid displacement in the perilymph that occurs when transmitting a sound pressure wave causes movement in the basilar membrane in the form of a traveling wave that propagates from the base of the basilar membrane to the apex (Purves et al. 2012, 286). Movement in the vertical plane of the basilar membrane also causes a shearing motion of the hair bundles, due to the tectorial membrane moving across the top of the stereocilia. As the wave traverses from base to apex, the amplitude of the traveling wave increases, with the velocity decreasing until the basilar membrane reaches its point of maximum displacement.

For each frequency, there is a maximum stimulation region on the basilar membrane (Purves et al. 2012, 286) similar to a resonance region (Figure 17) (Roederer 2009, 31).

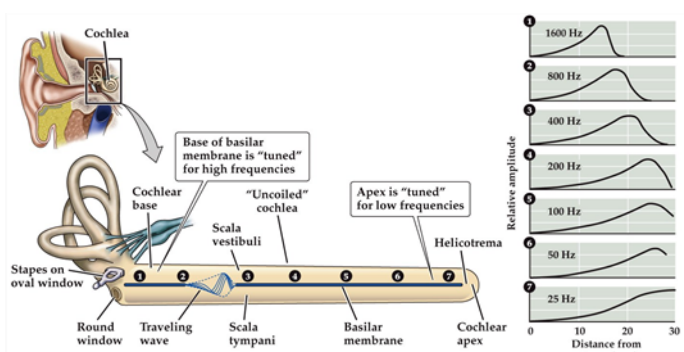

Figure 17: Basilar Membrane Stimulation

Relative frequency tuning curves as conveyed on a linearly rolled out cochlea from: (Purves et al. 2012, 286)

This is where the hair cells most sensitive to that frequency are located. It is because of this property that the inner ear can serve as a mechanical frequency analyzer, able to decompose complex sound waveforms into their components, which is critical in complex tone processing (Purves et al. 2012, 284). The properties of the basilar membrane explain the pattern of spatial frequency tuning of the ear: the basilar membrane is thinner, stiffer where mapping is associated with higher frequency reception mobile at the basal end and is thicker and more flexible at the apical end, the apex of the membrane allowing it to process lower frequencies. This spatial arrangement of frequency is a principle that is maintained throughout nearly all steps of nervous system processing of the sound, an arrangement known as tonotopy (Kolb & Whishaw 2014, 335). While the major contributor to defining tonotopy structurally is the basilar membrane in the cochlea, the stereocilia of the hair cells themselves also have this capability. In general, the smaller stereocilia are found on hair cells encoding for higher frequencies, while longer stereocilia are associated with lower frequencies. Sound pressure is converted into the receptor potentials of these inner hair cells that then get relayed into action potentials (AP) in the cells of the spiral ganglion. To understand this process, taking a closer look at the cellular level of the hair cells themselves and the surrounding chemical environment is necessary.

Since hair cells are transducing a pressure wave into a neuronal signal over such a fast temporal frame in the order of microseconds, direct mechanical gating is a necessity. This is the only transduction channel type fast enough and sensitive enough to facilitate transduction at this speed, since chemical transduction messenger systems are far too slow (Purves et al. 2012, 289). Physical connections between the tip links are what make such transduction possible. At the ends of each stereocilium there are mechanical potassium (K^+) channels that open as the stereocilia move in the direction of the tallest stereocilia in a hair bundle, depolarizing the cell, with hyperpolarization of the cell occurring in the opposite direction (Figure 18**).** All the channels of the smaller stereocilia open in this manner because their tips are physically linked, and as a result, when one moves, all the stereocilia follow suit as one unit (Kandel et al. 2013, 663). However, these tip links, as integral as they are to the rapid transduction mechanism, have a negative trade-off in that they can also result in sound distortion within the auditory mechanism, which will be further explored in the discussion of difference tones.

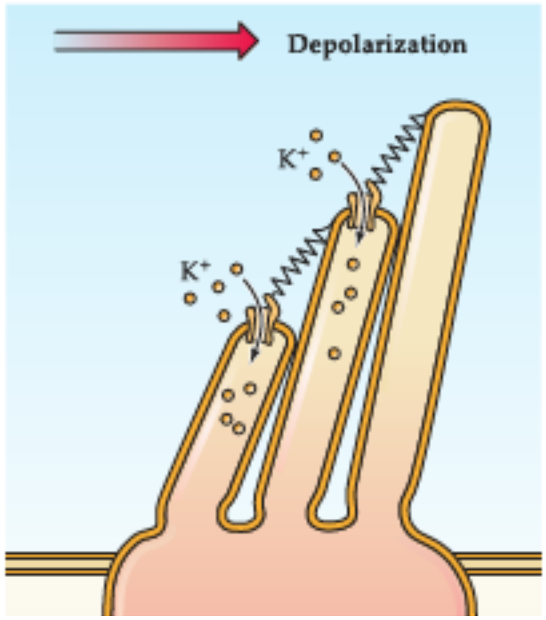

Figure 18: Tip Links

Tip links in motions acting as mechanical gates for ion channels located on the tips of the hair cells stereocilia from: (Purves et al. 2012, 288)

Amplitude of a sound is encoded neurologically in two different ways when transduced from hair cell to the neurons of the spiral ganglion. The first and more simplistic mechanism is the predominant approach used. In a rather intuitive manner, the louder the sound, the more the hair bundles displace, therefore the more mechanical K^+ channels will open in the stereocilia. This in turn leads to a greater depolarization of the hair cell, yielding a greater amount of neurotransmitter release from the hair cell onto the dendrites of the spiral ganglion cells, and resulting in a greater frequency of APs (Fuchs 2005, 7). To understand the second mechanism, it is important to note that, while many hair cells respond to a wide range of frequencies, provided the intensity of the sound stimulus is great enough, there are characteristic frequencies of each hair cell at which they are the most sensitive (Kandel et al. 2013, 671) (Figure 19).

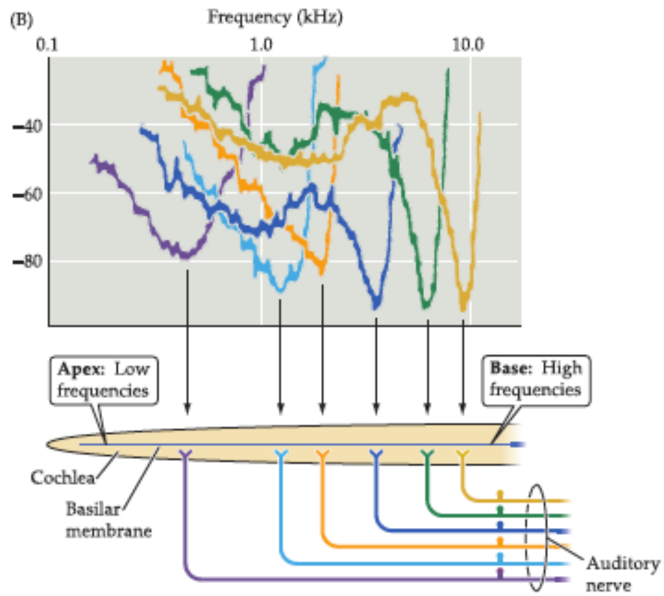

Figure 19: Characteristic Frequencies of Hair Cells

Characteristic frequencies laid out along the cochlea that are defined frequencies at which hair cell activation occurs at the lowest amplitude level from: (Purves et al. 2012, 293)

Therefore, if a certain hair cell's characteristic frequency is 1000 Hz, it will be the first cell to be able to respond to a 1000 Hz pure tone sound even heard at very low intensity levels. To further clarify, let us say that this 1000 Hz tuned hair cell begins to depolarize at an intensity level of 10 dB for a 1000 Hz pure tone, then no other hair cells should be active in response to this stimulus. However, if the intensity of that 1000 Hz pure tone increases to 20 dB, adjacent hair cells, whose characteristic frequencies are around 900 and 1100 kHz may become stimulated. By looking at the ensemble of neurons activated, as a function of these characteristic frequencies of the hair cells, that the second mechanism for intensity becomes clearly understood.

Pertinent CNS Auditory Processing

Transmission of auditory information from the peripheral to the central auditory system occurs at the cochlear nuclei, and follows the path of the hair cells to the spiral ganglion, whose axons form the auditory component of the vestibulocochlear nerve, or cranial nerve VIII (CN VIII) (Squire et al. 2008, 619). As the signal is transmitted to the auditory cortex, various brain structures are involved in processing specific characteristics of the sound stimulus. There are a number of different structures that aid in different aspects of the auditory system; however, this discussion shall only focus on those that serve a significant role in the auditory processing of pitch and complex sound. The general pathway presented in Figure 20 provides an overview of anatomical reference (Purves et al. 2012, 294). The binaural nature of the auditory system adds complexity, since central processing must not only consider the detection of a sound stimulus by the sensory receptors of one ear, but how the streams of information from the two ears integrate with one other.

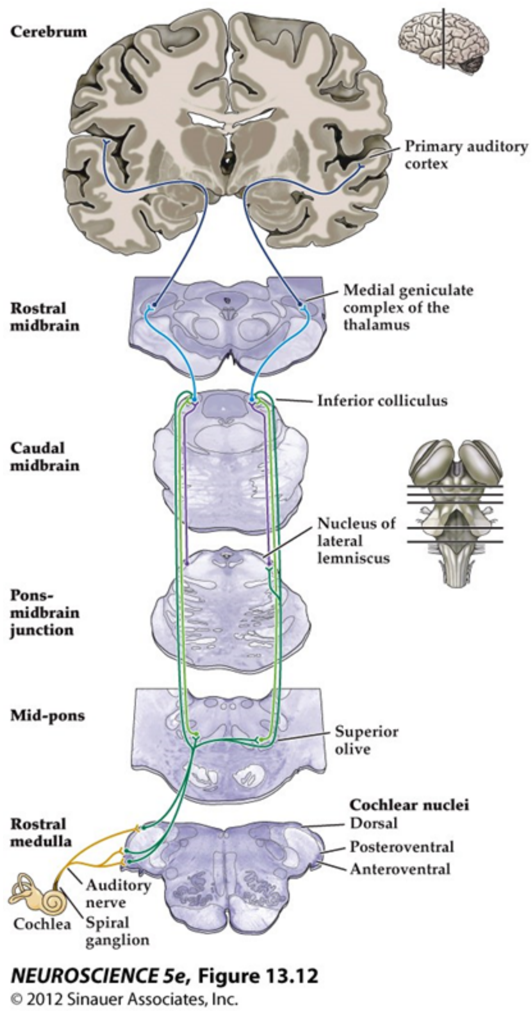

Figure 20: CNS Auditory Pathway

Layout of the neuronal pathway followed in the auditory system detailing each of the brain regions involved from: (Purves et al. 2012, 294)

Auditory processing is performed in both monaural and binaural streams, the complexities of which are beyond the scope of this paper. The integration of these left and right ear streams, which begins at the inferior colliculus in the caudal midbrain, as seen in Figure 20 above, allows humans the ability to begin to conceptualize temporally complex sounds such as speech and music by comparing the signals coming into each ear (Purves et al. 2012, 297). Further integration occurs at the medial geniculate complex in the thalamus, which projects to the primary auditory cortex (AI) and secondary auditory cortex (AII). To further ascertain their roles in complex sound processing, it can be helpful to delve into the roles of AI and AII and compare them to the language processing structures in the CNS: being Wernicke's and Broca's areas.

Missing Fundamental / Virtual Pitches

Contextual processing has a powerful influence on how we perceive sound. It allows us to integrate what we hear not only with our ears, but also with other sensory modalities, through the interconnectivity of the CNS. The primary auditory cortex or AI, is the predominant destination for higher order processing of auditory information. AI is mainly responsible for frequency discrimination using strict tonotopic organization (Kolb & Whishaw 2014, 332). The second auditory cortex also known as AII or the belt region, is less tonotopically organized, serving a more associative function particularly with regards to processing intraspecies communication. Interconnected not only with AI, but also to the language comprehension region, is Wernicke's area, as well as the language motor production region, Broca's area. Human language processing is extensive in and of itself; however, its connection to musical processing requires further understanding of cortex lateralization. First, it must be noted that attributing all aspects of a certain processing circuit to a single region within one hemisphere of the cortex is a gross oversimplification. The notion of being a left- or right-brained individual is a fallacy;

however, certain processes, such as language communication, do follow a left hemisphere lateralization, as shown in positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies of right-handed individuals. In such individuals, regions such as Wernicke's area, Broca's area, and the left AII are far more active when performing language processing-related tasks (Squire et al. 2008, 633). This is not to say that symmetrical structures on the right are not active during these tasks, but there is demonstrably less activity compared to those in the left hemisphere.

In contrast to language, however, the processing of musical sounds, seems to be lateralized in right hemisphere (Purves et al. 2012, 300). Specifically, AII of the right hemisphere seems to play a major role in pitch perception of tones that share overlapping spectral content. As such, it is reasonable to consider this region as a major contributor, in conjunction with AI, to missing fundamental effect. The missing fundamental effect, also known as phantom or virtual pitch, is an occurrence unnoticed to most, in which we perceive sound frequencies that do not exist in the stimulating sound source (Roederer 2009, 51). This is the case when two or more tones that are related harmonically are presented simultaneously. When this happens, it causes us to process the fundamental frequency that fits into the harmonics content presented. For example, if the pitches C4 and G4 are played simultaneously for a subject, the subject would also hear the pitch C3, even though C3 was not played. This is occurring because C4 and G4 are the 2nd and 3rd harmonics whose fundamental is C3. Musical application of these virtual pitches is sparse, predominantly due to the lack of awareness of this biological phenomenon. When it comes to production of the polyphonic voice, utilizing these virtual pitches provides the performer with the ability to produce additional pitches that in actuality are not even present in the sonic environment. The creation of virtual pitches is actually frequently used by common household electronic speakers. Many stereo speakers, especially those of a smaller size or that lack a subwoofer, do not possess the capability to produce certain very low frequencies (Roederer 2009, 46). In order to create the illusion that the listener is hearing frequencies in that low range, for instance F1 (43.65 Hz), the speakers will instead emit an F2 (87.3 Hz) and a C3 (130.96 Hz), thus tricking the listener's nervous system into registering a nonexistent F1. Such processing can be conceptualized as the brain's way of finding the greatest common denominator of the auditory stimuli, and it has been presented as a means to simplify cognition of pitch. This missing fundamental phenomenon illustrates how central processing is not a simple manifestation of auditory input, but context can bias later cognitive processing.

Difference Tones / Combination Tones

Combination tones are acoustic occurrences first discovered by the composer Giuseppe Tartini in the mid-18th century. It is interesting to note how long ago this aural oddity was discovered, and yet there is still a lack of consensus as to its precise origin. Known by various names such as combination tones, tartini tones, difference tones, etc., these only occur when two or more tones are presented to a listener simultaneously (Dowling & Harwood 1986, 36). When an individual is presented with two tones of roughly the same high sound-pressure intensity but that differ in frequency within a reasonable range, and where the frequency of tone one (f_1) is no more than an octave different from tone two (f_2), the individual will experience frequencies which are not existent in the two tones but which are known as combination tones (f_c). The frequencies of the pitches can be calculated mathematically using the formula:

$$f_c = xf_1 \pm yf_2$$

in that x and y are simply positive integers (Roeder 44). Combination tones exist due to a concept known as nonlinear distortion. This all comes into play due to the nonlinear transduction of sound in the ear. However, simply put, we hear these distortion products because the combination of f_1 and f_2 actually cause the basilar membrane of the cochlea, to vibrate at these f_c tones. For reasons that are not fully understood, there are certain distortion products that produce combination tones of greater intensity in the ear (Roederer 2009, 45).

Specifically, difference tones are often the easiest to identify, although it is critical that the two frequencies be at least 20-30 Hz apart for the difference tone to be perceived as a tone, due to the range of human hearing. The most commonly used difference tone is found when f_1 resides roughly a perfect fifth above f_2. For example, if f_1= 783.99 Hz or a G5 and f_2= 523.25 Hz or a C5, then the resultant difference tone is f_c= 260.74 Hz, or roughly a C4. In fact, this specific difference tone pattern is so common that it is used in a number of polyphonic instrumental and vocal techniques. When Tartini first noticed these difference tones acoustically, he applied the concept to his own violin playing by using double stops to play the two initial frequencies, a practice which can be utilized by any string player (Roederer 2009, 46). Pipe organs also take advantage of this unique acoustic phenomenon as a means to build a cost- and space-efficient instrument. In order to produce extreme low bass frequencies, a pipe organ requires the use of huge 32’, 64’, or even 128’ pipes, which are not only extremely costly to produce, but also require ceilings up to 128’ tall. However, thanks to the use of difference tones, the same low bass pitches can be produced by sounding two smaller pipes simultaneously in the same 3:2 frequency ratio as before, generating that difference bass frequency, all without having to buy huge, costly organ stops. Even brass instruments can utilize these difference tones by playing one pitch using their normal lip buzzing embouchure while simultaneously singing a P5 through the instrument, giving these instrumentalists access to extremely low sub-bass frequencies. This is also the same approach that certain singers are able to use to sing difference tones without the use of external filters, such as instruments, in a difficult polyphonic style (Edgerton 2015, 105). To perform these tones, Tartini singing requires the combination of glottal phonation and lip buzzing in the same previously indicated interval relationship. One can hear this vocal approach in Bobby McFerrin’s composition “Drive,” in which he utilized it to produce three pitches at the same time, creating a sound mimetic of a motorcycle engine (Trujillo 2007). Of course, any f_c can occur if it can be calculated using the above formula, but cases such as summation tones, which are the opposite of difference tones, are far more difficult to hear.

Before exploring what causes these nonlinear distortions within our auditory system, let's first clarify what exactly distortion is. At its most basic form, distortion is any modification made to a waveform of a signal as it goes through an amplifier. Distortion is frequently broken down as either linear or nonlinear, with the main difference located in the output signal's frequency. With linear distortion, the change from input to output through the amplifying system yields the same input frequencies, but generally at a variable intensity level (Deutsch 1999, 7). Nonlinear distortion, on the other hand, yields a change in intensity, while also resulting in additional frequencies in the output signal that were not present in the frequencies' input into the system. In electronic audio processing, which is where principles of distortion are most heavily used, these nonlinear distortion products tend to either be harmonic or intermodulatory in their patterns of presentation, meaning that they follow either the harmonic series or the previously discussed combination tone formula, respectively (Roederer 2009, 46). For electronic synthesis, intermodulatory distortion products can occur using a number of devices, but the most pertinent example to our auditory system is the use of compressors. Processing of an audio signal using compression serves the purpose of either amplifying a quiet sound or reducing the intensity of a loud sound, which is in essence the process occurring in our inner ear.

The level of sensitivity that the ear is able to achieve over various manipulations of sound pressure is noteworthy, conferring the ability to detect frequencies ranging from 20 to 20,000 Hz, or intensities from 0 to 120 dB, while the transduction apparatus should be limiting both of these capabilities. Even though the mechanotransduction apparatus limits the frequency with which the receptor potential can fluctuate sinusoidally at around 4000 Hz, humans can still hear frequencies above this level thanks to the labeled line tuning of frequency where pitch is attributed to specific location of hair cells on the cochlea, and as this is transduced into the CNS, tonotopy is maintained (Purves et al. 2012, 291) (Figure 21).

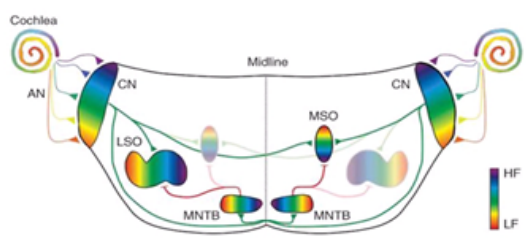

Figure 21: Tonotopy in the CNS

Conveyance of the conservation of tonotopy throughout the auditory pathway. MNTB= Medial Nucleus of the Trapezoid Body; MSO= Medial Superior Olive; LSO= Lateral Superior Olive; AN= Auditory Nerve; CN= Cochlear Nucleus; HF= High Frequency; LF= Low Frequency from: Kandler, Karl, et al. 2009. "Tonotopic Reorganization of Developing Auditory Brainstem Circuits." *Nature Neuroscience* 12, no. 6 (October): 711–717.

Where intensity is concerned, the dynamic range of a single inner hair cell is dependent upon the AP firing-rate capability of that afferent cochlear neuron in the spiral ganglion. That is, one inner hair cell is only capable of conveying roughly a 40 dB change in the input-tone intensity at their characteristic frequency (Yates et al. 1990, 206). How then are humans capable of hearing over an intensity range larger than the cells can signal? The answer is where the outer hair cells come into play. The outer hair cells act as the amplifiers of the inner ears through a compressor-like behavior. Outer hair cells possess a unique ability due to an anion transporter motor protein known as prestin located in the lateral membrane (Ashmore 2008, 186). Due to prestin, outer hair cells are capable of electromotility, which allows them to change their shape with the changing receptor potential. Therefore, when a traveling wave propagates across the basilar membrane and reaches the point of maximal displacement, this causes depolarization of the outer hair cells as well as resulting in their contraction longitudinally. Since the outer hair cell is also embedded right above the basilar membrane, and changing the cell shape exhibits a force on that membrane, this contraction yields a greater amplitude displacement of the basilar membrane (Alberti 1988, 61). This is what improves the sensitivity of the ear allowing us to detect very quiet sounds. However, these outer hair cells are also efferently controlled by top-down processing originating from the medial superior olivary complex located in the mid-pons section of the brainstem as seen in Figure 20 (Squire et al. 2008, 620). This results in inhibitory control, releasing acetylcholine under conditions of a high-intensity sound stimulus, thereby inhibiting the amplifying nature of the outer hair cells. Not only is this a protective measure, but this rounds out the compression, since low-intensity sounds are amplified and higher-intensity sounds are not. Therefore, the input range of about 120 dB exhibited by a sound source is compressed to roughly 35dB within the cochlea for processing purposes (Yates et al. 1990, 218). While this electromotility of the outer hair cells is useful for intensity compression in addition to frequency tuning, it is this property along with properties of tip links that causes nonlinear distortion within the cochlea.

Otoacoustic emissions (OAE), which are simply sounds that are generated from the inner ear, are actually a result of this nonlinear distortion phenomenon. They are either spontaneous in nature, in that they occur without any input stimulus, or evoked, as in done in hearing tests for infants (Kandel et al. 2013, 673). Tinnitus, often described as ringing of the ears and associated with cochlear damage, is an example of spontaneous OAE. As for evoked OAE, difference tones are the most common examples of this, and they can be elicited in the same manner as previously explained. Due to the compression performed by the outer hair cells, any nonlinear distortion product can be elicited as an evoked OAE. However, these nonlinear distortion products occur experimentally without outer hair cells as well, so the movement of the outer hair cells cannot be the only factor involved in generating combination tones. It is here where the tip links re-enter the equation. Stereocilia, at the point between open and closed transduction channels, exhibit a less stiff stability known as their gating compliance (Jaramillo et al. 1993, 527). As a result of the decrease in rigidity of the stereocilia during that small range of motion, the tip links connecting the stereocilia can exhibit extra springiness on the hair bundle. That extra springiness facilitates a force exerted by the tip links that are attached to the transduction channels (Purves et al. 2012, 292). This force elicits pitches that do not coincide with the input oscillations associated with the two tones f_1 and f_2. The force exerted by the transduction apparatus yields distortion products in the oscillation of the hair cells that can be calculated using the formula previously presented for combination tones. Evidence of this fact is experimentally supported when the combination tones are no longer present in the two tone stimulated hair cell, if tip links are artificially destroyed, or the transduction ion channels are chemically blocked (Jaramillo et al. 1993, 527). The magnitude of the force elicited by the tip links is substantially inferior to the force instilled by the outer hair cells themselves; therefore, while still significant, the tip links play only a supporting role to the outer hair cells in the generation of combination tones. The nonlinear characteristics of the ear remain areas of active research, so tip links and outer hair cells might not be the only elements involved. In fact, compelling theories also postulate that CNS processing and distortion within the neuronal circuitry could explain the existence of combination tones (Ziebakowski 2012, 574).

Without the auditory system, concepts such as polyphonic singing or even music in general would be of little consequence; however, when considering auditory perception of the polyphonic voice, necessity for this sensory system is inherent. This is especially true since the factors that are innate to auditory transduction and processing that can modulate the input stimulus they receive. As noted in the previous discussions of the missing fundamental effect and of combination tones, it is evident that there is more involved in reception of the polyphonic voice than the sound source produced. Perception of an auditory stimulus is not only subjective based upon the individual perspective of each listener, but also due to the biological layout of the auditory system.

By considering these perceptual stimulus distortions, listeners can account for the bias they instill into the listening experience when a complex sound source is concerned.

CHAPTER 4: A Study of Polyphonic Voice in Tuva

Culture of Tuva

A scientific foundation to conceptualize the polyphonic voice is of interest, but becomes more meaningful with a point of reference to apply the information provided. Therefore, analyzing a culture that puts the polyphonic voice into practice facilitates contextualization of this vocal technique. Of all the cultures that have developed a polyphonic vocal style, the Tuvan culture possesses the most globally renowned techniques. This alone would not be enough to justify focus on this one culture among the other geographically disparate societies that have independently developed this characteristic vocal approach. However, there are two additional factors that validate the choice to delve into the intricacies of Tuvan throat singing. The first pertains to certain cultural ideals unique to the Tuvan people that impacted the relative time frame during which the inception of this musically atypical vocal production occurred. Compared to Western exposure to the technique only 60 years ago, Tuva's involvement occurred many centuries ago, making the timeline discrepancy abundantly clear. Certain attributes of Tuvan culture, such as their approach to pedagogy as well as their spiritualistic ideals, provide probable explanations for this early polyphonic development. The second, and even more important, factor for choosing Tuva is that it includes among its throat-singing styles both bitonal and diphonic polyphonic approaches, which is quite remarkable since the majority of early societies did not develop even one of these techniques. Therefore, the physiological framework through which to conceptualize the intricacies of the vocal mechanisms that differ between each unique voice in Tuva allows us to extrapolate how these approaches can be applied to nearly any other culture's polyphonic style, whether they use a bitonal or diphonic approach.

Even though Tuvan throat singing may be the most well-known throat-singing culture of today, they were not the sole progenitors of throat singing. Throat singing itself originated among the Turk-Mongol people of the Southern Siberian and Central Asian regions (Pegg 1992, 32). Essentially, the Tuvan people were influenced by Mongolian culture, which was instrumental in its inception. The word "Khöömei" in Turkish, which is the basic Tuvan throat-singing style, was originally a Mongolian word, spelled "xöömei," indicating that the Mongolian culture actually introduced throat singing to the Tuvan tribe. This likely occurred as a result of the great degree of cultural mixing throughout ancient history in this region due to the highly nomadic nature of both the Mongolian and Tuvan people whose two cultures were in contact through trade routes and tribal warfare (Tongeren 2002, 84). For centuries thereafter, until the end of the Chinese Manchu Empire in the early 20th century, Tuvans and Mongols were actually unified politically (Pegg 2001, 11). This is yet another contributing factor as to why there is so much overlap culturally between them, especially evident in the early throat-singing traditions of these cultures. The first written record referring to what historians believe now to be throat singing dates to some point during the 16th century (Rachele 2013, 10). While certain specifics surrounding the accepted origins of throat singing may never be answered, this does confirm the roughly four century lead Tuva had on Western polyphonic development. To further illustrate the point, the historical record detailing the introduction of overtone singing into Western society is far more precise and indicates its first use in the 1960s by Karlheinz Stockhausen in his piece "Stimmung."

When considering the polyphonic technique, superficial beauty is easily identifiable in the sonority generated, but there must be more to the Tuvan use of throat singing than aesthetic pleasure. When uninformed Western musicians hear throat singing for the first time, they are often intrigued by the unique quality of the sound produced; however, that is not the sole intention of this technique in Tuvan culture (Barras & Gouiffès 2008, 60). Tuvan throat singing was initially utilized exclusively by the herdsmen and hunters of the tribe. Their reasoning for singing was due to their belief that the throat song pleased their horses, making them move faster (Tongeren 2002, 56). Naturally, in a nomadic culture whose main form of transportation was on horseback, making the animals speed up would have been quite useful if it were actually effective. The modern culture of Tuva allows people of all professions to practice and perform throat singing; it is no longer restricted to the nomads and hunters (56). As a nomadic culture, the people were and are constantly interacting with their external environment and that is where the true purpose for Tuvan throat singing lies.

Animism is a strong spiritual underpinning that has been integral to the belief system of the Tuvan people since before the inception of throat singing centuries ago. The concept of animism entails a belief that natural structures and phenomena have spirits (Levin & Edgerton 1999, 80). As a means to convey these spiritual beliefs, the

Tuvan society found a use for throat singing by using the sonorities produced to mimic nature. They use their singing as a way to connect with nature, almost as a form of sonic meditation. Since it is commonplace in animism to believe that animals, the weather, or even landscapes possess spiritual power, by emulating the sounds found in nature it is thought that humanity can assimilate this raw power by connecting to the spirits. This tendency towards mimicry is substantiated through the development of the different styles of throat singing in Tuva, where a number of the styles are meant to emulate aspects of nature like rushing water or the galloping of horses. It is also common for Tuvan throat-singing performers to seek solitary refuge in steppe or taiga biomes within the Republic of Tuva to perform their singing (Tongeren 2002, 56). In doing this they can commune intimately with the earth through their song. Animistic practice is more of a societal ideal of the Tuvan people, not directly linking to a specific religious observance; however, there have been documented cases of the use of throat singing in shamanistic study. In this practice throat singing is used in ritualistic contexts by shamans, or “kham” as they are called in Tuvan, to interact with different spirits beyond nature spirits (Tongeren 2002, 80). This is quite common due to the high prevalence of shamanism within Central Asian cultures. When shamans are referenced, the idea of spiritual healing often coincides, but the Tuva people do not actually believe that throat singing itself possesses any specific healing power. The singing is merely used as a tool by shamans in conjunction with ritual practice to facilitate spiritual healing or even to combat possession in some cases (Glenfield 2007, iv). They do, however, believe that soft production of khöömei or sygyt can aid in childbirth (Tongeren 2002, 81). While there is little to substantiate this claim, it is ironic that throat singing is used to help women in this way, particularly since, historically, women were not allowed to throat sing. However, when looking at gender roles in the singing practice today, more women are learning the art form.

The historically low female participation in professional throat singing is not due to misogynistic prejudice, but is rather a byproduct of a Tuvan cultural misconception. From a technical standpoint, there is no denying that, when it comes to overtone singing, women are at a biological disadvantage. Men have the advantage, since, anatomically, women tend to have shorter vocal cords and a smaller tube for resonance between their lips and vocal cords, which results in two main issues. The first is that they have a smaller resonating body with which to amplify the sound they produce and, in turn, the harmonics they generate, so their voices are quieter (Tongeren 2002, 30). In addition, women's average pitch is about an octave higher than men's, which limits the number of harmonics they can generate. This is an issue because overtone singing is best produced in frequencies going up to 2000 Hz and women start higher in the frequency range, so they therefore have fewer harmonics from which to choose (30). However, this can be remedied through the practice of kargyraa, which is a technique that allows for women to lower their base frequency by about an octave or more, similar to the range of the average male singer.

It is important to note though that kargyraa can be more difficult for women to learn due to the absence of these lower frequencies in the female voice, but it can be done generating a timbre unique from the male voice production of kargyraa (Cope 2004, 40). The misconception of the people of Tuva is that throat singing causes infertility, which is why there was such a strict taboo against throat singing by women for such a long portion of their history, but more and more women are starting to practice throat singing professionally in modern Tuva (Levin & Edgerton 1999, 82). Because of this broader use, the practice of throat singing in Tuvan culture has developed its own structural characteristics, which speak to the influences of the culture in the musical practice.

Musical structure is one of the most telling attributes indicating certain values of a culture. In countless cultures throughout history, pentatonicism is a recurring scalar pattern seen in musical structure. The pentatonic scale comes in a variety of different forms, but the main constant is that it is made up of only five unique pitches, whereas our Western system is predominantly diatonic using seven unique pitches and possessing a set interval structure. A well-known example of pentatonicism can be noted in the spiritual, *Amazing Grace*. Pentatonicism, while not exclusively Asian, occurs frequently in Asian cultures, which is fitting for this context. Tuvan throat singing is no exception to the pentatonic tonal tendency, which can be heard in the harmonic pitches most frequently used in the melodies they generate. The overtones most commonly isolated together in Tuvan throat-singing melodies are the 6^{th}, 7^{th}, 8^{th},9^{th}, 10^{th}, and 12^{th} harmonics, avoiding the 11^{th} harmonic, making a pentatonic pattern, since the 6^{th} and 12^{th} harmonics are the same pitch one octave apart (Levin 2017). Important to note is that, of those harmonics, the 7th is used quite sparingly in the majority of melodies (Levin & Edgerton 1999, 87).

Western music theory further supports the aesthetically pleasing pentatonic patterns that these throat-singing melodies tend to follow. By analyzing the intervals of each of these harmonics relative to the fundamental, predominantly aurally pleasing consonant intervals are observed. Using Figure 1 as reference for this discussion, for simplicity all intervals will be presented within a single octave, where the 6th harmonic is described as a Perfect 5th (P5) from the fundamental when it would really be a Perfect 19th. Building sequentially from that P5 generated between fundamental and the 6th harmonic, one hears a minor 7th (m7) for the 7th harmonic, an octave/unison for the 8th harmonic, a major 2nd (M2) for the 9th harmonic, a M3 for the 10th harmonic, and another P5 for the 12th harmonic. Aside from the m7, which is created by the 7th harmonic that is the least frequently used in this pentatonic group, each of the previously mentioned intervals are among the most commonly used in modern Western music. While not a direct correlate, the popularity of these intervals in Western culture is predominantly due to the positive emotional associations we have with them. Looking from the opposite perspective of harmonics that are avoided in these Tuvan melodies, such as the interval generated by the 11th harmonic, a tritone, as well as the m7 generated by the 7th harmonic, one can justify why these singers choose to shy away from these tones. The tritone in Western music is often referred to as the devil's interval due to its unpleasant sonority, and a m7 is not much better (Clendinning & Marvin 2005, 107). Taking Western analysis one step further, the instability of the 7th and 11th harmonics can be further supported. As described in CHAPTER 1: The Physics of Sound, the harmonic series does not follow the equal temperament tuning system to which Western ears are accustomed. As a result, the 7th harmonic falls 31 cents flat of the pitch that would make the interval between the fundamental and this harmonic a true m7, and the 11th harmonic falls 49 cents flat of the pitch that would make the interval between the fundamental and this harmonic a true tritone (Figure 1). In the equal temperament system, there are 100 cents between each semitone; therefore, being 50 cents away from a semitone is the most out of tune any pitch can be because it falls a quarter tone away, or halfway between two semitones. Therefore, the 7th and 11th harmonics are so far flat that even an untrained ear can hear how the intervals they create sound mistuned and unpleasant. For perspective, when looking at the preferred harmonics in the pattern, none of them falls more than 14 cents away from their respective semitone, which is a frequency variation that is imperceptible to most untrained ears.

The harmonic series can be a limiting force in the production of the polyphonic voice. Up until this point, this unique vocal phenomenon has been portrayed as an unfettered means for a singer to produce multiple notes simultaneously; however, both the bitonal and the diphonic voices have their shortcomings. Bitonal singing is in fact the more restrictive in that the pitches produced can only create octave intervals, and, in certain cases, a perfect 5th. As for the diphonic approach, the mathematical consistency of the harmonic series is what limits the upper melody. A diphonic singer who keeps the fundamental constant can only generate a melody based on the partials produced by that pitch, although manipulation of the fundamental opens up a whole new realm of possibilities (Levin & Edgerton 1999, 87). Through analyzing these restrictions of the polyphonic voice, a divide between the Tuvan approach and the Western overtone singing style is highlighted.

It is this divide that presents the first prominent cultural distinction that may have increased the probability of Tuvan society to develop polyphonic singing at an earlier time than Western cultures. Solely focusing on diphonic styles, Tuvan throat-singing performers tend to remain static on one fundamental while manipulating the harmonic melody with a greater focus on the timbral differentiation of their songs (Tongeren 2002, 238). When a Western overtone singer produces a song, the focus is more on harmony, assessing the relationship between both fundamental and harmonic as each line moves simultaneously. These different musical tendencies are indicative of intent for the two cultures' use of the polyphonic voice. Such harmonic motion requires not only an intimate knowledge of the harmonic series, but also how the intervals relate as the fundamental shifts, which can be heard in the performances of Western singers like Anne-Maria Hefele (Hefele 2014). Her performances are indicative of the common Western musical ideal found in virtuosic mastery, and in this case the overtone style is the technique over which she has total command. Such a level of mastery would never even be attempted by even the most prolific Tuvan throat singers, not because they are incapable, but simply due to differing cultural perspectives where music is concerned. For Western society, often competition lingers as an unspoken motivator and monetary reward is typically the incentive; being able to exhibit a superior intellect or technical prowess gives the competitor an advantage. However, the technical demands of polyphonic singing are what left Western society unaware of the polyphonic voice for so long. Tuvans developed throat singing as a means to further their spiritual connection to nature through aural mimesis, demonstrating a purpose dependent upon shared social values rather than individual advancement. This is in contrast to some Western musicians where purpose is instead directed at methods that differentiate and/or appeal to a larger target audience. Implementation of overtone singing in popular Western musical genres is quite limited, not only due to shorter life span of popular songs in the West, but also due to a nearly nonexistent target audience. Capitalism, while positing regions of the Western world with unprecedented technological proliferation and prosperity, may have, contributed to limitations of our creative innovation. Westerners frequently convey their virtuosity through complex rhythmic and harmonic musical displays, whereas the Tuvan people are far more interested in the production of rich timbral content in their vocal production as demonstrated by their devotion to throat-singing practice. In addition, in the timbrally complex tones they produce, Tuvans strive for improved mimesis, always trying to make their tones sound more like the natural phenomena that they are attempting to emulate.

Aside from their pentatonic nature, most throat-singing songs have a character similar to folk music in that they are normally not written down, are unique to the culture, speak to tales of nature, and are constantly evolving with time through oral tradition. In fact, due to the lack of rigid form, the freedom of throat singing compositionally facilitates a highly improvisational nature (Tongeren 2002, 67). Performance in many cases often requires original creation on the performer's part since there is not an extensively documented common repertoire. There are also extended pauses of up to 30 seconds in throat singing, which is strange to a Western ear since one might think that these divide the unitary musical melody. However, Tuvan people do not see each phrase as belonging to one overarching musical idea, but rather each phrase is its own original sonic portrait and the phrases can interact with one another. Aside from allowing the performer to catch his or her breath, a major purpose for the pause is for the performer to listen to the ways in which the past phrase is interacting with nature and ambient sound (Levin & Edgerton 1999, 87). The performer then uses that preceding sonority to formulate a response in his or her next phrase.

When it comes to familiarity, throat singing is the norm in Tuvan culture and is so common that it is held in the same regard as monophonic singing is in the West. As such, children are fascinated by it at an early age but quickly get used to its prevalence throughout their cultural life as they move into adulthood (Tongeren 2002, 57). A concept foreign to Western trained musicians, throat-singing teachers in Tuva use a very hands-off, non-detail oriented approach. Clearly it is quite effective since some children are able to produce the khöömei voice by two years of age, but generally the students learn almost exclusively by watching and listening to a teacher sing and receive very little direct technical instruction (70). As the structural form of Tuvan throat singing developed, different unique voices were created as well. A hands-off approach is due to their lack of a formally defined musical teaching style. Music is taught by imitation of an experienced singer where the student gets little formal instruction, and any guidance received is through word of mouth. After a demonstration, the "teacher," who would be more of an elder singer providing an example as a Western student might see it, would likely say "You see," meaning the student should gather all they need just from listening to the master (Tongeren 2002, 70). This further solidifies the role of mimicry in this society as seen not only in the music, but in the pedagogy of throat singing. Looking at the Tuvan word "chylandyk," which are the first sounds made by children when they are attempting to produce the khöömei voice through imitation of an elder singer, the mimicry learning practice can be seen (Tongeren 2002, 69). Still, it is intriguing that students do not really master their voices until their late teens to early twenties, likely due to change in VC length during puberty. For instance, at age 13, children are able to produce the harmonics of sygyt voice, but cannot control the fundamental or the specific harmonic pattern desired.

Contrastingly, musical instruction in the West is highly structured. Transmission takes place using highly ordered written notation, with specific guidelines and restrictions. Such an approach promotes the conventional and discourages the esoteric.

For instance, this is shown through the low prevalence of unique musical concepts such as quarter tones or multiple melodies moving in parallel motion. Being taught what to do often allows one to overlook how else something can be done. Expanding further, the process of discovering how to do something allows one to stumble into accidental innovation. All of this is not to provide the misconception that imitative teaching has negative consequences, but rather to convey how the Tuvan approach to musical pedagogy yielded the ideal conditions for stumbling upon the production of the polyphonic voice. In fact, there are tradeoffs to each pedagogical method, the traditional versus the serendipitous. Where, for instance, the logically ordered structure established in Western society is time efficient, this approach limits wasted moments and prevents iterative failures that the trial-and-error approach Tuvan pedagogy relies upon utilizes. Perhaps Tuva's less definitive oral method may allow for greater creative opportunity without the restrictions dictated in the West, possibly factoring into the early inception of throat singing. The trial-and-error processes Tuvans must experience to find an approach that allows them to produce the distinctive throat voices has its advantages. Not only does the larger allotment of time devoted to practicing allow them to hone their voice at earlier ages; it has also likely facilitated further creative innovation that has led to the many unique throat-singing substyles.

When viewed through a Western lens even the concept of musicality in Tuvan culture is foreign in comparison to the Western conception of the idea. In fact, Tuvan culture does not even have a word for music that equates to the Western term, but rather refers to the idea of music as the sounds of their world that they are offering back to the world itself (Levin 2006). Western society is often constrained by our need to understand. There is no denying the freedom that unfaltering faith can grant an individual or even a society collectively. By not having to constantly question every phenomenon or precept of the physical world, the mind is free to create and take action, rather than to hypothesize and explain. It is this concept that most likely played an integral role in the early development of throat singing in Tuva. In the West, a drive to define and categorize every new and unfamiliar observance can restrict the innovative process. Categorization may serve many benefits facilitating communication and clarity; however, as demonstrated in Tuvan society, the more time spent doing and less time asking how, allowed the development of multiple approaches to the polyphonic voice. The caveat is that the nomadic herders did not really have a scientific conception of what they were doing physiologically to generate these characteristic timbres. However, in finding their own path, they not only generated their own techniques along the way, but also formed their own approaches to even the most basic khöömei voice that were most suited to their individual voices. In addition to not having a word for music, there are no set words for drone or overtone like there are in Western culture, which, while explicable due to the cultural distinctions, is still surprising due to the fact that these musical constructs are fundamental to the production of throat-singing techniques (Tongeren 2002, 71). Tuvans developed an understanding of these ideas by using practice of the techniques. Each performer throughout their experience comes to a personal perspective of what these sonorities mean to him or her. Generally, they are using their musical performances as a means to connect spiritually with nature. Creative skill comes predominantly from intuition and experience, not a detail-orientated anatomical understanding of what they are producing from their throats. This is partially the reason why research into understanding their unique polyphonic voices is not extensively documented to date. However, looking into the physiology that facilitates the Tuvan throat-singing voices can provide Westerners with a means to appreciate and approach this unique aural phenomenon.

Anyone has the capability to produce these sounds anatomically: it just takes practice and buildup of the musculature. Practice of throat singing is comparable to monophonic singing, for which it takes years to hone mastery of these techniques, building power and endurance. Strength of the throat-singing voice peaks from age 20-30 and the power declines thereafter (Tongeren 2002, 70). As such, young adult performers are normally held in high esteem as the best practitioners. Master throat singers are called “khöömeizhi” and, to receive that title, one must have a certain set of skills. One must be able to improvise new throat-singing melodies on the spot, have a handle on multiple different throat voices, be in peak physical vocal condition, and be able to generate one's own melodies and musical themes. Master throat singers have impeccable endurance which they use to hold a stable drone with little sign of wavering or fatigue; they are also almost always capable of playing one of their traditional string instruments to accompany themselves as they sing. The doshpuluur or the igil are each common string accompanying instruments; however, the igil is challenging to master and few professionals can both sing khöömei and play this instrument simultaneously, leaving only the true virtuosos capable of conquering that feat.

Tuvan Voices

While there are numerous unique Tuvan throat-singing styles, three main voices make up the fundamental core of the average Tuvan throat-singing performer. They are khöömei, sygyt, and kargyraa (Tongeren 2002, 64). As with most polyphonic singing, the fundamental is maintained in the voice as a pedal tone while harmonic manipulation generates the melody (31). Khöömei, meaning pharynx or throat appropriately, is the main style that is the easiest to learn, similar to Western overtone singing, and is sung in the midrange of the voice, generating the lower overtones in the F1 resonance of the harmonic series. Sygyt is sung in the upper part of the vocal range, generating high overtones often described as flute- or whistle-like, and is meant to emulate the sound of the wind, which is a fitting name since sygyt translates to flute. The third style, kargyraa or "hoarse voice" in Tuvan, is the most technically unique form, generated in a completely different manner than the previous two voices. It produces a low subharmonic fundamental frequency meant to ground the singer to the earth, generating a harmonic rich tone, producing lower harmonics.

Technical production of these core voices seems easy when watching a professional, but it can take years to master. It is also uncommon for throat-singing professionals to attain expert control of all the main voices and substyles. Typically, they focus their training on one or two throat styles in particular. Khöömei, while also being the main throat-singing voice, is often used as a blanket term for all Tuvan throat singing. In order to produce any of these voices it requires the use of khorekteer, which is the Tuvan equivalent of chest or modal voice. Similar to the word khöömei, khorekteer is commonly used to refer to any throat-singing voice as well (Glenfield 2003, 33). The main aspect that separates throat singing from Western overtone singing is the Tuvan use of the squeezed voice technique that allows for airflow control (Cope 2004, 36). Aside from giving throat singing its unique timbre, the main purpose of the squeezed voice is to amplify the harmonics by limiting the backflow of the sound from reverberating into the throat. The khöömei voice uses similar formant manipulations, creating varying resonant positions in the oral cavity, comparable to Western overtone singing. At the same time, however, squeezed voice technique is applied, which is why this is the most simplistic of the main voices since it requires the least amount of specific laryngeal or articulator control (36). Although this is comparable to the other throat-singing voices, mastering the khöömei voice is still incredibly difficult. For this style, it is mainly the manipulation of the base or back of the tongue in conjunction with changing the degree of opening made by the lips that determines the harmonic frequency generated as conveyed in Edgerton's M3 (Figure 12: Tongue Position Methods).

The difference, when listening to the khöömei style compared to Western overtone techniques, is that the squeezed voice technique provides an easily discernible timbre, but what physiological action that facilitates it is more difficult to ascertain.

Squeezed voice mainly comes from two main actions. The first is constriction of the laryngeal vestibule, also known as the supraglottic space, which is the space from the false folds to the laryngeal inlet. The second is increased adduction of the glottis (Grawunder 2009, 47). Musculature responsible for such laryngeal positioning works synergistically to achieve each of these tasks. Many of the muscles involved in this process are also active in the swallowing process, so it can be helpful to conceptualize this to begin assessing placement. Constriction of the supraglottic space is achieved through the action of the oblique arytenoid muscle (OA), which acts to approximate the arytenoids to one another as well as draw them forward ventrally towards the epiglottis (Figure 5: Muscles of the Larynx) (Snell 1986, 861). In the process of swallowing, this is the muscle that controls the constriction of the laryngeal inlet, serving as the second sphincter in the larynx aside from the vocal cords. More specifically, in the supraglottic constriction observed when using squeezed voice, the aryepiglottic fiber that is a part of the OA is predominantly active (Dmitriev 1983, 196). This all causes AEF tension as well as dorsal closure of the laryngeal vestibule, yielding constriction of the supraglottic space in the anterior to posterior plane. In addition, the thyroepiglottic muscle fibers, which are part of the TA and attach between the thyroid and epiglottic cartilages, constrict the VCs pulling the arytenoids towards the thyroid cartilage, yielding the same dorsal closure, but only farther down in the larynx. Finally, the increased adduction of the glottis is achieved through action of the LCA keeping the aperture tight. While not done to the same degree as when used in throat-singing voices, this squeezed voice is also what is responsible for the singer's formant. This yields a boost for harmonics in the 2000 to 4000 Hz range due to the resonating cavity of the larynx being manipulated (Cosi & Tisato 2003, 7). It is important to note that squeezed voice is predominantly a laryngeal constriction, rather than the misattributed pharyngeal constriction that many non-specialists misinterpret. In fact, if anything there is increased opening of the laryngopharynx during implementation of this technique. To maintain the steady fundamental tone, along with initiating phonation of the vocal cords when they exhibit such a high adduction percentage, this pressed voice technique also requires a great deal of subglottic pressure (11). Therefore, it is the job of the singer's support mechanism, specifically the abdominal muscles around the epigastrium working with the diaphragm, to maintain adequate pressure. To recreate this action, it can be helpful to compare activation of the support mechanism to the feeling of bearing down when lifting a very heavy object.

Of the remaining two styles, there is much debate over which is more difficult, but each presents its own challenge. Sygyt uses the same squeeze voice technique as khöömei, while using the tongue to filter out the lower overtones and isolate the high piercing harmonics. This is easiest achieved using either "l" or "r" sounds to position the tongue properly. Specifically, it is done by placing the tongue blade against the upper alveolar ridge, right above the front teeth, pressing the sides of the tongue against the upper gums relative to the molars on each side, and then using the rest of the tongue to block the opening between the mouth and the throat, leaving a small opening on either side of the mouth so that very limited airflow is allowed to pass (Levin & Edgerton 1999, 82). This approach is described by Edgerton M2, mentioned in CHAPTER 2: The Anatomy and Physiology of Sound (Figure 12**)**. The opening can also be left in the center of the tongue; however, most of the most proficient sygyt singers prefer to have the sound emanate out of one side of their mouths, using the hard surface of their teeth to enhance the bright intensity of the whistle-like tone (Cosi & Tisato 2003, 11). By so doing, the resonating space in which the harmonic is generated is made smaller, allowing for the higher harmonics to sound, and changing the position of the middle part of the tongue allows for different harmonics to be elicited. The reason that these whistle-like harmonics occur is due to formant resonance created by the movement of the tongue. Specifically, the squeezed voice is working to dampen the fundamental, while the tongue filter system, given this placement, shapes the resonant cavity such that F1 harmonics are attenuated and F2 harmonics get amplified. Then the raising of the center of the tongue up to the hard palate increases the frequency of the amplified harmonic due to compression at the F2 antinode (Levin & Edgerton 1999, 83). The most skilled sygyt singers can come close to eliminating the fundamental frequency altogether, essentially isolating the whistle harmonics.

Kargyraa is often considered to be the most characteristic style, most easily recognizable by the Western ear. If a Westerner has heard of Tuvan throat singing at all, typically the voice most commonly associated with the overall genre tends to be kargyraa, due to its highly distinguishable subharmonic timbre. As mentioned previously, the fundamental in kargyraa throat singing is created in a completely different way when compared to the other two voices. The source sound generation is bitonal in nature in that there are multiple oscillating structures phonating simultaneously. The three main oscillators in the throat as previously mentioned are the VC, VTF, and the AEF, but what is so unique about kargyraa is that it can be performed using any combination of these three structures (Figure 4**).** By far the most common form of kargyraa combines phonation of the VC-VTF together. When first approaching learning this unique style, it is often easiest to think of the sound source placement as moving vocal fry phonation slightly above the vocal cords. Here the subharmonic fundamental is generated by getting both the ventricular and vocal folds vibrating concurrently using a constant glottal push, whereas in khöömei and sygyt the vocal cords were the only sound source (Ken-Ichi et al. 2007, 2). The VTFs actually generate a pitch that is one octave lower than the vibrating frequency of the VCs (1). This is due to the mucosal wave cycle of the VTFs following a rate that is half that of the VCs, meaning that the VCs open and close twice as often as the VTFs during the production of the kargyraa. Therefore, this octave is generated due to the nature of the octave interval following a 1:2 frequency ratio (Grawunder 2009, 53). Once this base phonation is achieved, the formant's position can be manipulated in the mouth to generate different overtones, adding in the diphonic aspect of kargyraa such that at least three tones can be produced simultaneously. Resonant filtering in kargyraa is similar to Edgerton M3, except that the tip of the tongue remains resting behind the lower teeth instead of at the raised midpalatal position (Figure

12: Tongue Position Methods). To generate the lower harmonics, the base is positioned proximal to the rear wall of the oropharynx, and then moves rostral to amplify mid-range harmonics. During the production of the highest harmonics, a rising of the epiglottis anteriorly is observed, such that the epiglottis blocks the vallecula, which is the space between the tongue root and epiglottis (Levin & Edgerton 1999, 86). When performing kargyraa, instrumental accompaniment is uncommon since this style features a large audible frequency range as a result of the multiple low frequency source tones. Fundamentals sounding in kargyraa are of lower frequency, so there is a denser complex of harmonics in the audible range to choose from, and therefore the sonority generated does not need instrumental supplementation (Tongeren 2002, 65). An additional side effect of the dense sonority is the increase in intensity in comparison to the other Tuvan styles since there are more sound source waves generated.

Alternate phonatory combinations are also available in the production of the kargyraa voice. Aside from the aforementioned VC-VTF form, kargyraa can also use both VC-AEF as the oscillators, which is the next most common approach (Grawunder 2009, 81). When it comes to using this style of kargyraa, using the AEFs as the secondary sound source instead of the VTFs is the only change. In using the AEFs, the second pitch generated is also one octave lower than the VC frequency, following the same pattern as VTF phonation. The main difference between these two approaches to kargyraa comes in their timbres. The VC-VTF method generates a much deeper resonance in comparison to the VC-AEF method's higher and richer timbre. Considering the anatomical orientation of each of the VTFs and the AEFs relative to the VCs, this timbral pattern is explained. A lower resonance in the VC-VTF approach is due to the more proximal nature of the VTFs to the VCs lower in the vocal tract. Consequently, the higher timbre of the VC-AEF method can be explained as a result of the more distal nature of the AEFs and the VCs in the larynx, such that the AEFs reside higher up in the vocal tract. Kargyraa has two main substyles, denoted as dag-kargyraazy and khovu-kargyraazy. Dag-kargyraazy, which translates to "mountain" in English, uses extended long notes and creates a deeper timbre with an unstrained voice that is meant to represent height of the mountains (Tongeren 2002, 65). Khovu-kargyraazy, which means "steppes" in English, presents with a higher timbral sonority as a result of greater laryngeal strain conveying the vast openness of the steppes that are abundant in the Tuvan landscape. While there is little conclusive research attributing specific anatomical origins to these individual kargyraa substyles, the distinct timbral quality of each makes it seem logical to hypothesize that dag-kargyraazy uses VC-VTF phonation, while khovu-kargyraazy uses the VC-AEF approach.

Even though it is used infrequently in practice, the kargyraa voice can be generated using all three laryngeal sound sources. This unnamed style combining VC-VTF-AEF still sounds like a rather standard form of kargyraa to the untrained ear; however, the overall sound intensity of this voice is greater and more timbrally rich than any other throat-singing style when performed by a throat-singing expert. The combination of all three of these source oscillators not only generates three simultaneous fundamental frequencies, it can also utilize harmonic filtering methods to amplify a harmonic, thus creating four simultaneous pitches. However, there is a redundancy in the subharmonic pitches generated by both the AEFs and VTFs resounding at the same octave lower than the fundamental, so that the wave summation yields an overwhelming subharmonic pitch. Since the standard kargyraa voice does not utilize the squeezed voice technique, adding it to kargyraa singing creates a khöömei-kargyraa style; however, this is difficult to master due to added laryngeal constriction impeding upon vibration of the VTFs (Grawunder 2009, 66). The ornamental styles that follow add unique timbral color to these main voices.

Some musicologists argue that there are two more core voices called borbangnadyr and ezenggileer, but in actuality these are more ornamental techniques that can be used in conjunction with one of the other three main voices, generally khöömei or sygyt, to produce unique musical effects (66). Borbangnadyr is used to mimic the sound of rushing water, and ezenggileer is meant to emulate the sound of a cantering horse, demonstrating how throat singing is meant to further a connection with nature. Borbangnadyr originates from a verb in Tuvan meaning “to roll over” like water that rolls over stones in a brook, and is produced using lip protrusion and the initiation of a clonus of the masseter muscle in the mandible (Grawunder 2009, 61). This creates the rapid brook oscillating timbral trill associated with this technique. Ideally the lip trilling motion oscillates at a rate of roughly 10-12 cycles per second. Ezenggileer means “stirrups” and represents the rhythmic nature of the rider’s boots on horseback (Levin &

Ûrevna 2011, 66). It is easiest to conceptualize approaching this technique through the use of velar consonants. Specifically, in English “g” and “ng” are the best examples, and ezenggileer is achieved through repeated articulation of these consonants. At each consonant articulation, the back of the tongue touches the soft palate, creating a slow passed trilling effect (Grawunder 2009, 61). Using “ng” as the articulator would allow this to come naturally, but often this technique also utilizes rising and lowering of the soft palate, causing a fluctuation between nasal resonance coupled with oral resonance, and strictly oral resonance.

Aside from the main three throat-singing voices, there are endless sub-techniques due to different regional technical approaches, and partially due to many famous throat singers generating their own styles of throat singing to differentiate themselves, thus making their timbre unique and desirable (Pegg 2017). For starters, there are the aforementioned substyles of kargyraa, dag-kargyraa and xov- kargyraa, which each create highly contrasting sonorities in the low frequency range. Additionally, a sub-technique of the sygyt voice, known as “chylangyt,” is meant to mimic the sound of a snake, which is fitting since the root of the word “chylan” means snake (Tongeren 2002, 69). This sub-technique is not to be confused with the aforementioned “chylandyk,” which is used to define the first khöömei voice of a young child. Chylangyt actually combines the technically demanding aspects of both kargyraa and sygyt, making it among the most difficult substyles to master. The bitonal voice sources using the VC-VTF vibration found in kargyraa are filtered by the M2 harmonic filtering method utilized in sygyt (Grawunder 2009, 66). Dumcuktaar, meaning "to nose," is yet another substyle in the Tuvan repertoire. Its transliteration is highly appropriate, since it is produced using only nasal phonation. As such, when singing dumcuktaar, the lips are closed and the oral cavity maintains a constricted orientation (Grawunder 2009, 61). Humming is the closest equivalent found in the West and manipulation of the harmonics produced is achieved through constriction of the oropharynx by the tongue or soft palatal movement. There is even another substyle of kargyraa called kanzat kargyraa that is unique to the singer Albert Kuvezin, who created the technique (Yat-Kha 2009). While he features this original style in his traditional Tuvan/rock fusion band called Yat-Kha, it unknown how he produces this contrabass style of kargyraa. One more substyle of kargyraa, named after its inventor, Vladimir Oidupaa is called oidupaa kargyraa (Cole 2017, 1). This voice, while not only known to Oidupaa, is still rather exclusive, utilizing an overall higher vocal resonance range in conjunction with a heavy nasal resonant component. These are but a small number of the named substyles observed in Tuvan throat-singing practice.

Tuva has proven to be an informative culture where the contextualization of the polyphonic voice is concerned. By analyzing how each unique voice is produced, awareness as to how polyphonic techniques developed in cultures throughout the world can be ascertained. In addition, it has been made clear how different technical approaches observed in the voices of Tuva can be applied in the consideration of countless other cultural polyphonic styles. Their spiritual purpose, their oral teaching style, and their freeing lack of a definitive framework either compositionally or theoretically for musical conventions have all contributed toward the early development of the polyphonic voice in Tuva. These factors have also clearly facilitated an environment in which the polyphonic voice can thrive and even innovate as indicated through the multiple substyles within the throat-singing repertoire, even to the point at which throat-singing masters are often associated with certain new timbral voices they have created for themselves.

CONCLUSION

How versus do. This dichotomy is pervasive not just in scientific study, but it is integral to human consciousness itself. Humanity's innate craving for understanding can only be satiated through action in the same way. Culture is highly indicative as to one's reliance upon one of these two concepts. In Western society, the "how" seems to dominate, whereas in Asian societies such as Tuva they often are satisfied with the "do." These tendencies have been alluded to throughout this discussion of polyphonic singing. It is especially apparent in CHAPTER 4: A Study of Polyphonic Voice in Tuva of how polyphonic voice developed so early historically in the Tuvan culture, compared to the more recent initiation of overtone singing to a limited subsection of Western musical practice. The major cultural distinctions that likely contributed to these occurrences are the Tuvan mimetic spiritual purpose for singing, their oral trial-and-error heavy pedagogical approach, as well as their freedom from questioning and categorization. However, the most noteworthy proof of our Western bias was not an example explicitly provided, but rather the overall structure and topics covered in this thesis as a whole: a scientific study done specifically as an overview of a polyphonic culture that itself cares little about how they produce these rich, timbrally entrancing sonorities, but instead only that they are doing it. Their drive is mainly instilled by their application of throat singing to their mimesis of the natural as seen through their spiritual animism. By contrast, often in Western societies there is more concern shown with regard to questions and how things work. Here that concern is taken into a culture different than our own, as a framework is presented to conceptualize the mechanistic intricacies of polyphonic vocal styles, specifically in Tuva.

Polyphonic singing has developed in a number of different cultures worldwide, so that by focusing on its inception in the Tuvan culture through a technical analysis of their throat-singing techniques, purpose and approach for polyphonic singing globally can be extrapolated. In order to approach each throat-singing style it was imperative to first develop a more Western-centric scientific understanding as was posited in the first three chapters, establishing concepts to delve into a culture that does not truly comprehend the complexity of this idiosyncratic vocal style they have developed. First presented was the importance of the fundamental physics of sound that pertain to music, focusing on the implication the harmonic series has in dictating the timbral characteristics of polyphonic styles. Next was a discussion of vocal anatomy and physiology of the laryngeal mechanism within the throat in conjunction with the formant position of the mouth, as they combine to impact the distinction between bitonal and diphonic polyphonic vocal approaches. Moving from there, the neurophysiological mechanisms highlighted were intended to explain not only the processing of sound, but more specifically how humans encode complex sounds like those produced in polyphonic singing, revealing the role perception plays in potential distortion of a stimulus sound source. As a means to put everything into context, Tuva became the focal culture upon which to incorporate each of these applicable scientific fundamentals. Thanks to their advanced techniques, grasping this culture's approach through these disciplinary lenses provides an encompassing perspective from which to compare, and even attempt to understand, various other cultures' own polyphonic voices.

BIBLIOGRAPHY

Alberti, Peter W., and Robert J. Ruben. 1988. *Otologic Medicine and Surgery*. London: Churchill Livingstone, 1988.

Ashmore, Jonathan. 2008. “Cochlear Outer Hair Cell Motility.” *Physiological Reviews* 88, no. 1: 173–210.

Bannon, Nicholas. 2012. *Music, Language and Human Evolution*. Oxford: Oxford University Press.

Barras, Marie-Cécile, and Anne-Marie Gouiffès. 2008. “The Reception of Overtone Singing by Uninformed Listeners.” *Journal of Interdisciplinary Music Studies* 2, no. 1/2: 59-70.

Clendinning, Jane Piper, and Elizabeth West Marvin. 2005. *The Musician's Guide to Theory and Analysis*. New York: W.W. Norton.

Cole, Krystle. 2017. “Throat Singing.” Neurosoup. Accessed January 13, 2018. www.neurosoup.com/throat-singing/.

Cope, Jonathan. 2004. *How to ‘Khöömei’ and Other Overtone Singing Styles*. Wild Wind & Sound for Health.

Cosi, Piero, and Graziano Tisato. 2003. “On the Magic of Overtone Singing.” *Unipress*: 1–18.

Dargie, David. 1988. *Xhosa Music: its Techniques and Instruments, with a Collection of Songs*. Cape Town: David Philip.

Deutsch, Diana. 1999. *The Psychology of Music*. Cambridge: Academic Press.

Dmitriev, L.b., et al. 1983. "Functioning of the Voice Mechanism in Double-Voice Touvinian Singing." *Folia Phoniatrica et Logopaedica* 35, no. 5: 193–197.

Dowling, W. Jay, and Dane L. Harwood. 1986. *Music Cognition*. Cambridge: Academic Press.

Edgerton, Michael Edward. 2015. *The 21st-Century Voice: Contemporary and Traditional Extra-Normal Voice*. Lanham: Rowman & Littlefield.

Fuchs, Paul Albert. 2005. "Time and Intensity Coding at the Hair Cells Ribbon Synapse." *The Journal of Physiology* 566, no. 1:7–12.

Glenfield, Alexander James. 2007. "Embodying Numinous Sounds, Exchanging Numinous Symbols: 'New Age' Overtone-Singing Rituals in Tuva." PhD diss., Yale University.

_______. 2003. "The Pearl of Tuva: Authenticity and Tuvan Khorekteer (Throat Singing)." *Canadian Journal for Traditional Music* 30: 32–46.

Goldman, Jonathan. 2002. *Healing Sounds; The Power of Harmonics*. Healing Arts Press.

Grawunder, Sven. 2009. *On the Physiology of Voice Production in South-Siberian Throat Singing: Analysis of Acoustic and Electrophysiological Evidences*. Berlin: Frank & Timme.

Gray, Henry, et al. 1974. *Anatomy, Descriptive and Surgical.* 1901 ed. Baltimore: Bounty Brooks.

Hefele, Anna-Maria. 2014. "Polyphonic Overtone Singing - Anna-Maria Hefele." YouTube. Accessed March 11, 2018. www.youtube.com/watch?v=vC9Qh709gas.

Holmes-Bendixen, Allison Ruth, et al. 2013. "The Influence of Whistle Register Phonation Exercises in Conditioning the Second Passaggio of the Female Singing Voice." *University of Iowa*: 1–161.

Jaramillo, F., et al. 1993. "Auditory Illusions and the Single Hair Cell." *Nature* 364, no. 6437: 527–529.

Kandel, Eric R., et al. 2013. *Principles of Neural Science*. New York: McGraw-Hill Medical.

Karmyn. 2016. "Your Vocal Cords." KT Vocal Studio. Accessed March 7, 2018. https://ktvocalstudio.com/vocal-cords/.

Kendall, Katherine. 2016. "Laryngeal Anatomy." Ento Key. Accessed March 7, 2018. entokey.com/laryngeal-anatomy/.

Ken-Ichi, Sakakibara, et al. 2007. "Observation of Subharmonic Voices." *19th International Congress on Acoustics Madrid* (September): 1-5.

Kent, Raymond D., and Martin J. Ball. 2000. *Voice Quality Measurement*. Norwich: Singular Publishing Group.

Kolb, Bryan, and Ian Q. Whishaw. 2014. *An Introduction to Brain and Behavior*. Worth Publishers.

Lee, Sang-Hyuk, et al. 2018. “The Singers Formant and Speakers Ring Resonance: A Long-Term Average Spectrum Analysis.” *Clinical and Experimental Otorhinolaryngology* 1, no. 2: 92–96.

Levin, Theodore. 2006. “The Art of Tuva Throat Singing.” On Point, WBUR Boston's NPR. Accessed February 25, 2018. onpoint.legacy.wbur.org/2006/01/13/the-art-of-tuva-throat-singing.

Levin, Theodore. 2017. "Tuvan Music." Grove Music Online, Oxford University Press. Accessed December 14, 2017. <http://www.oxfordmusiconline.com/subscriber/article/grove/music/51667>.

Levin, Theodore, and Michael Edgerton. 1999. “The Throat Singers of Tuva.” *Scientific American* (September): 80-87.

Levin, Theodore, and Suzukej Valentina Ûrevna. 2011. *Where Rivers and Mountains Sing: Sound, Music and Nomadism in Tuva and Beyond*. Bloomington: Indiana University Press.

McCoy, Scott. 2013. “Formantology.” *Journal of Singing* 70, no. 1 (October): 43–48.

Moore, Keith L. 1985. *Clinically Oriented Anatomy*. Philadelphia: Williams & Wilkins.

Pegg, Carole. 1992. “Mongolian Conceptualizations of Overtone Singing (Xöömii).” *British Journal of Ethnomusicology* 1: 31-54.

_____. 2001. *Mongolian Music, Dance, and Oral Narrative: Performing Diverse Identities*. Seattle: University of Washington Press.

_____. 2017. "Overtone-singing." Grove Music Online, Oxford University Press. Accessed December 14, 2018. <http://www.oxfordmusiconline.com/subscriber/article/grove/music/49849>.

Purves, Dale, et al. 2012. *Neuroscience*. Sunderland: Sinauer Associates Inc.

Rachele, Rollin. 2013. *Overtone Singing Study Guide*. Abundant Sun.

Roederer, Juan G. 2009. *The Physics and Psychophysics of Music: An Introduction*. New York: Springer US.

Snell, Richard S. 1986. *Clinical Anatomy for Medical Students*. Boston: Little Brown and Company.

Squire, Larry R. 2008. *Fundamental Neuroscience*. Cambridge: Elsevier/Academic Press.

Sundberg, Johan. 1994. *The Science of Musical Sounds*. Cambridge: Academic Press.

________. 1977. "The Singing Formant." Sundberg's Singing Formant. Accessed March 7, 2018. hyperphysics.phy-astr.gsu.edu/hbase/Music/singfor.html.

Tongeren, Mark Van. 2002. *Overtone Singing: Physics and Metaphysics of Harmonics in East and West.* Fusica.

Trujillo, Diego Cerda. 2007. "Bobby McFerrin - Drive." YouTube. Accessed March 7, 2018. www.youtube.com/watch?v=1qOQHB_V2g0.

Walk That Bass. 2017. "10. The Overtone Series and Dissonance." Youtube. Accessed March 7, 2018. https://www.youtube.com/watch?v=0lmS5lQ5MSU.

Wannamaker, Robert. 2008. "The Spectral Music of James Tenney." *Contemporary Music Review* 27: 91-130.

Yates, Graeme K., et al. 1990. "Basilar Membrane Nonlinearity Determines Auditory Nerve Rate-Intensity Functions and Cochlear Dynamic Range." *Hearing Research* 45, no. 3: 203–219.

Yat-Kha. 2009. "The Story of Yat-Kha (a Kind of Memories but Not Yet Actually)." Ethno-Rock-Band Yat-Kha. Accessed March 7, 2018. www.yat-kha.ru/en/biography.

Ziębakowski, Tadeusz. 2012. "Combination Tones in the Model of Central Auditory Processing for Pitch Perception." *Archives of Acoustics* 37, no. 4 (January): 571–582.