THE SHAPE OF A–E–I–O–U

A Case Study of Vowel Height, Vocal Synthesis, and Placement in Collegiate A Cappella Singers

Arun Venugopal
Christian DeKnatel

NORTHEASTERN UNIVERSITY

MUST 4611 – Spring 2020
Professor Anthony Paul De Ritis
PHYS 5318 – Spring 2020
Professor Bryan Spring

Submitted April 14th, 2020

Abstract

The human voice is the most widely accessible musical instrument, and with its ubiquity comes a dynamic range of timbres, tones, and techniques. Through manipulation of various qualities of the voice, a range of combinations and sonic characteristics can be achieved. Singers execute performances through technical control of their vocal folds, vocal tract, and placement of the tongue, which alter qualities such as pitch, vowel height, and type of vocal synthesis. This experiment analyzes the acoustical phenomena that occur when singers change placements on each vowel, and how placement affects the frequency spectrum depending on vocal register and type of synthesis. We seek to apply our analyses about harmonic qualities to music composition, arranging, and direction by determining which choices and combinations of vowels blend or clash across voice parts. Collegiate vocalists of different voice parts were recorded singing five pure vowels – oo, oh, ah, eh, and ee – in combinations of three pitches, three registers of their voice, and three vowel height placements. Each audio sample was processed in MATLAB and analyzed in MS Excel; vowels with the highest standard error of mean values are analyzed to determine causes of their uncertainty. In terms of spectral density, the maximum, minimum, and median cases were plotted together for each register and technique to understand the range of harmonic content for each vocalist. In addition to takeaways for each voice part, our general findings include that descriptions of vowel height placement (colloquially bright, neutral, and dark) are not perfectly congruous to descriptions of sound as it pertains to high harmonic content. "Ah" was generally found to be the spectrally brightest sound, while "oo" sat on the opposite end of the spectrum as the least spectrally dense. Power was not correlated to placement or technique, though it was found that the high belt is the most powerful vocal technique for all ensemble members with the greatest range in power. A plot of vocal samples on a graph of spectral density v. centroid illustrates the validity in partitioning specific harmonic responsibilities between members within the ensemble. Finally, the discussion concludes with the postulation of a new visual musical notation called Cube Charts that blends three variables of vowel nuance into one concise system that can be scaled and applied to professional vocal environments or educational choral settings.

Keywords: Physics, Music, Singing, FFT, Acoustics, Resonance, Vowels, Vocal Synthesis, Placement, Vowel Height, Vocal Register, Arranging, Composition, A Cappella, Collegiate, Spectral Density, Spectrogram

Contact:
Arun Venugopal – venugopal.a@husky.neu.edu
Christian DeKnatel – deknatel.c@husky.neu.edu

Contents

Preface

When life changes course outside of one's control, the best option is to focus on what one can control. Such is the mantra of this paper. The COVID-19 pandemic disrupted the latter half of Northeastern University's Spring 2020 semester, and with it, prematurely ended our ability to complete the research. Our task was to record six vocalists – from low to high: bass, baritone, tenor, alto, mezzo-soprano (referred to as mezzo in this paper), and soprano. The campus closure on March 12th, 2020 forced our project to hone in on the four vocalists that had been recorded in time; and forced us to forgo recording our alto and soprano. As a result, our research paper has evolved. It was, and remains, a detailed study of how certain vowel techniques may quantitatively be compared; however, it now additionally discusses a novel method of musical notation based on observations from our data, and an outline for areas of further research based on limitations on the scope of this project. It is our hope that while it may not be a complete overview of singing techniques, this paper shines a light on the topic of vowels in a new, unique lens, and ushers in more complete physical and musical research on the observations and postulations discussed through this preliminary study.

I. INTRODUCTION

i. *How do scientists understand the voice?*

Artistically, the human voice is the only instrument directly connected to the soul. Scientifically, the voice is a part of a system consisting of the vocal folds, vocal tract, resonators, and the tongue. At the junction of the two lies this paper, what can simply be described as a study of vowels. This is an investigation of vowels in combinations with pitches, modes of synthesis, and vowel height – five vowels will be sung on three different pitches in three vocal modes of synthesis and three vowel heights. The objective is to analyze and compare vowels for different singers through Fast Fourier Transforms and spectrograms, and apply our observations to musical arrangement and composition in vocal music; specifically, in contemporary a cappella ensembles.

Modes of vocal synthesis are related to the vocal folds, the infolded mucous membranes stretched over the larynx that vibrate and manipulate the airflow from the lungs. Studies on the human voice have included "Articulatory synthesis of words in six voice qualities using a modified two-mass model of the vocal folds" by Birkholz, Kroger, and Neushaefer-Rube (2011), which models the vocal folds as a two-mass system with components inclined towards each other instead of parallel. The paper discusses types of vocal synthesis in the context of pop music singers. Three forms of synthesis discussed in that study will be implemented in this experiment through the singing styles of chest voice, belting, and head voice. "Chest voice" and "head voice" refer to forms of resonance, rather than actual specific resonators; there are, instead, more standardized terms that are tailored to the anatomical functioning of the vocal folds. The former "chest voice" utilizes the **modal voice**, which “is characterized by a regular and periodic vibration pattern, where both the ligamental and cartilaginous part of the vocal folds vibrate as a single unit” (Birkholz et. al, 2011, p.4). The latter “head voice” utilizes **falsetto**, which “is characterized by regular vocal fold vibrations at noticeably higher frequencies than modal voice. The glottis often remains slightly open resulting in an audible friction noise component” (2011, p.5). The third singing style is belting. The musical technique for belting is often described as a “loud mix” of chest and head voice, with a large quantity of breath pushed through the vocal tract to support the tone. Considering these aspects, we will scientifically approach belting as a technical combination, or mix, of modal voice and falsetto, and take into account **pressed voice**, which “is characterized by a higher medial compression than modal voice… and in consequence by a higher degree of vocal fold adduction” (2011, p.4).

Vowel height is related to the vocal tract, the space within the mouth that filters the air coming through the vocal folds. It refers to whether the tongue is positioned high, medial, or low within the mouth, and is musically described as dark, neutral, or bright, respectively.

Pitch is the result of how quickly the vocal folds vibrate at a fundamental frequency $f0$. Timbre describes partials that are produced at different frequencies as a result of the vocal tract’s shape.

In research prior to this, the focus has been on choral singers, vocal modeling, resonance strategies, and analyses of vowel production in relation to pitch. Harris and Mehaffey’s 2016 paper “Pitch, dynamics, and vowel tuning in choral voice: Utilizing resonance strategies to train stylistic variance for choral singers” examines spectral analyses to understand the impact certain harmonics have on absolute timbre. It also explains that the vocal tract can be trained to achieve predictable acoustic environments (Harris & Mehaffy, 2016). This inspires the basis for our definition of a “technically-trained” singer: a vocalist who has the ability to consistently create specific acoustic environments using the components of vocal synthesis. The technically-trained singers chosen for this study are able to differentiate and produce the different modes of vocal synthesis and vowel heights with consistency and ease (see Appendix B for full singer bios).

Chuang and Wang’s 1976 study “Influence of vowel height, intensity, and temporal order on pitch perception” indicated that there is a high correlation between pitch perception and vowel height (Chuang & Wang, 1976). There is, however, a general lack of knowledge related to how modes of vocal synthesis and vowel height acoustically contrast in various singing registers within one person’s voice, and how techniques compare between singers with different vocal ranges.

ii. *How do musicians understand the voice?*

The most common categorization in choral and vocal music is abbreviated to SATB: soprano (high female), alto (low female), tenor (high male), and bass (low male). In contemporary mixed-gender a cappella groups, arrangements are often broken into SMATBB, with possible upper and lower divisis per voice part. Women are usually split into soprano (high), mezzo (middle), and alto (low) ranges, and men are usually split into tenor (high), baritone (middle), and bass (low) ranges. These categories do not include singers with unique or extreme vocal ranges. Examples include males who sing in ordinarily-female registers, i.e. countertenors, or females who sing in ordinarily-male registers, i.e. female tenors. In general circumstances, the individual's vocal range, not gender, is the ultimate determining factor in what category they fit in; our understanding, ability to test, and nuance of gender are discussed further in Section VI.iii.

The goal for this study is to understand how contemporary collegiate a cappella singers can influence arrangements through their choice of vowel, vowel height, and type of vocal synthesis. By studying the physical spectrums of the vocalists, we can compare these nuances in vowels both across voice parts and within different registers of one voice. These findings will ideally aid music directors who wish to give more nuanced instruction of an arrangement; arrangers who wish to refine how they assign vowels to singers; and singers who wish to blend better with their fellow musicians.

II. BACKGROUND

i. *Vowels and the Vocal Tract*

Vowels can be characterized in different ways. Depending on the shape of the lips, vowels can be "rounded" or "unrounded." The shape of the oral cavity allows for three classes of vowels related to "tongue backness:" front, central, and back. Front vowels feature the tongue placed far forward in the mouth, in the pre-palatal region. Central vowels hold the tongue in the medio-palatal region. Back vowels keep the tongue in the post-palatal region. Tongue backness and lip roundedness are physically separate from vowel height (Singwise, 2020).

For this experiment, the quality that will be examined is "brightness," a colloquial term referring to the timbre created by the height of a vowel. Brightness is defined by the placement of the tongue and jaw, and is also understood to be related to the relative intensity of the first formant. Multiple terms refer to the same concept from different angles. Dark vowels, also known as high, tall, and open vowels, are characterized by high tongue placement, tall shaping of the vocal tract, and a relatively open jaw. Bright vowels, a.k.a. low, spread, and closed vowels, are associated with low tongue placement, a horizontally-spread shape of the vocal tract, and a relatively closed jaw. Neutral vowels can be described as mid, close-mid, and open-mid, depending on whether it is exactly halfway between open and closed or tends toward one of them.

In vowel production, "A E I O U" as one would normally speak are not considered "pure" vowels – they are usually spoken as diphthongs in daily life, where *A* is "ay-ee", *I* is "ah-ee", and *U* is "ee-oo" ("Singing Better Vowels," 2020). For the purpose of this experiment, we will look at the five primary vowels:

Table 1: Vowels Selected for Testing

Vowel	**Primary Vowel**	**IPA Symbol**	**Example of Pronunciation**	**Vowel Description**
A	Ah	**ɑ**:	Father	Open back unrounded
E	Eh	**ɛ**	Met	Close-mid front unrounded
I	Ee	**i**	Meet	Close front unrounded
O	Oh	**o**	Home	Close-mid back rounded
U	Oo	**u**:	Blue	Close back rounded

It is worth noting that brightness is separate from resonance, although both can sometimes be referred to as placement. In the case of the former, "placement" refers to the positioning of the tongue, jaw, and the vocal tract, and is what will be analyzed in this experiment. In the latter, the term describes where sound is focused through combined usage of three "resonators," which are spaces where sound resonates: the oral cavity, the nasal cavity, and the labial cavity. This form of resonance placement will not be directly measured by this experiment, but can influence results depending on the style of singing the subject usually performs in. Since the vocalists in this study all primarily perform contemporary popular a cappella in a selective auditioned group that mainly accepts singers that naturally resonate in their oral cavity, it is assumed resonance placement will be similar between singers, and thus an insignificant role in the data acquisition and analysis.

ii. *Choosing singers and methodology*

In this experiment, the vocalists are college-aged (19-23), popular music singers who all participate in the Northeastern University-based a cappella group *The Nor'easters*. These singers were selected for this experiment due to their advanced technical training and vocal styles that reflect contemporary popular styles. The singers were native English speakers, had a New England regional dialect, and were confident in singing contemporary American music.

In creating the methodology, the goal was to be time-efficient yet thorough when recording the vocalists, and reduce the amount of strain that was imposed on their voices. We wanted to test modes of synthesis on different pitches. On low pitches, chest voice is the predominant way to sing; therefore, modal voice was the only synthesis mode tested on the low note. The middle note was selected with testing both chest voice and belting in mind, as it allowed us to compare the synthesis modes of modal voice and the mix. The high note was selected to test belting and head voice, and provided the opportunity to compare the mix and falsetto. Starting with a modal voice on the low note, the singers sang in bright, neutral, and dark placements. The vowels were sung from back-most to front-most, one placement at a time, in the order: oo, oh, ah, eh, ee. Once the three placements were sung, then the next step was to sing in the modal voice on the middle note. The process for the three placements repeated, and then moved on to belting; then onto belting on the high note; and finally falsetto on the high note. This organization allowed for sessions to be completed within forty-five minutes to one hour, well before the singers became vocally fatigued.

The process began with the singers warming up their voices across their entire range, so that they could then accurately complete a range test to determine their lowest, middle, and highest notes. Then, to reduce the strain of singing extremely low or high, the "low" note was an interval of a fifth above the lowest extremity; the "high" note was a fifth below the highest extremity; the middle note was halfway between these.

III. MATERIALS

i. *Software*

An assortment of hardware and software was utilized through the experiment. The digital audio workstation Avid ProTools 12.8 was used for recording and editing the samples that were collected. The Sonic Visualizer family of applications are open-source programs developed by the Centre for Digital Music at Queen Mary, University of London that produce detailed visualizations and analyses of audio recordings; Sonic Lineup was used specifically for spectral comparison (Appendix E). A Macintosh computer running MacOS Sierra V10.12.6 was the machine that processed and stored the data from each session.

Once data was collected and organized, it was moved to another computer running MacOS Mojave 10.14.2. Audio samples were trimmed into clips (0.5s) in Logic Pro X, an Apple-produced Digital Audio Workstation (DAW). A program was coded in MATLAB_R2020a to further trim clips to median slices (0.1s) and produce Fast Fourier Transforms (FFT) of these slices recursively. Microsoft Excel for Mac (Version 16.35) was used for data normalization, analysis, and visualization. Visualizations were compounded in Adobe Photoshop CC 2017.

ii. *Hardware*

Since the goal was to precisely determine how the spectrums of each vowel are represented in the frequency domain, selecting a microphone that did not alter or "color" the sound was of great importance. An array of microphones was available from the Media Studios at Northeastern University. After researching the options, the Electro-Voice RE20 was chosen. Its even frequency response shown in Figure 1 made it stand out.

It also features a unique cardioid design that picks up sound evenly both on- and off-axis and mitigates the proximity effect of bass. This means if the singer accidentally moved to the side or too close by accident, there was little negative effect on the quality of the recording.

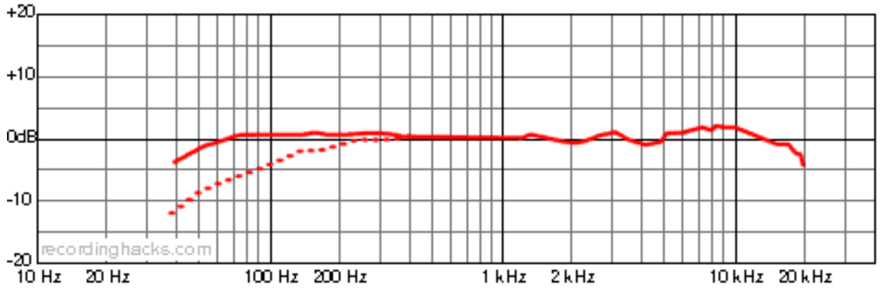


Figure 1. Frequency Response of Electro-Voice RE20 Microphone

The microphone was placed on a stand and isolated by an sE Electronics SPACE vocal shield. The vocal shield protected the microphone from unwanted noise caused by the singer's voice reflecting off of the walls. The preamplifier was an API-512C. It was connected through a Switchcraft Patchbay to a Universal Audio Apollo 16 interface, which acted as the Digital Audio Codec for the computer. The function of these devices is to: (1) boost the microphone's weak electrical fluctuations into a significant, processable signal and (2) send the analog signal into the computer as digital values. The signal flow is shown in Figure 2.

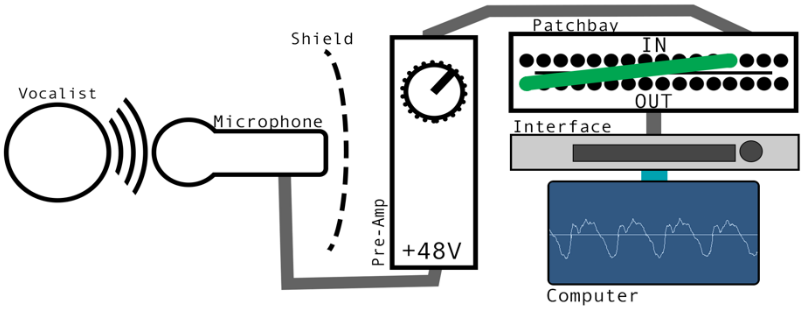


Figure 2. Signal Flow of Vocalist Recording setup. XLR cables (Grey) connect microphone, pre-amp, patchbay and interface. Patch cable (Green) link in/out channels. Thunderbolt (blue) connects computer.

iii. *People*

Quantitative descriptions for each singer are displayed in Table 2, with full bios available in the appendix. A4 is equal to 440Hz. From this reference point, the expected frequency for each note is found in equal temperament[1] ("Physics of Music", 1998). The process of testing range and selecting the three notes are explained in the Methodology section.

Table 2: Quantitative Descriptions of Vocalists

Voice Part	Full Range	Low Note	Middle Note	High Note
Mezzo	E^b3 – $A^\#5$	B^b3 (233Hz)	G4 (392Hz)	$D^\#5$ (622Hz)
Tenor	G2 – $F^\#5$	D3 (146Hz)	F4 (349Hz)	B^b4 (464Hz)
Baritone	$F^\#2$ – $F^\#5$	$C^\#3$ (139Hz)	C4 (262Hz)	B4 (494Hz)
Bass	B1 – G4	$F^\#2$ (93Hz)	E3 (165Hz)	C4 (262Hz)

IV. METHODOLOGY

i. S*patial Acoustics – Selecting a space*

To accurately capture and analyze the voices of our singers, numerous variables had to be independently isolated during recording. Other factors had to be considered and addressed during our data analysis. Before recording the first vocalist, a recording space had to be selected. It needed to fulfill a number of criteria:

1. Be as close to acoustically-neutral as possible – a space that produces resonances or noise as a product of the room, ventilation, or external foot or vehicle traffic would impede collection of accurate vocal frequency data.
2. Provide a consistent experimental set-up that would not change drastically from day-to-day – for example, a classroom where chairs or desks were often moved around would influence how the microphone, electronics, and singer would be positioned.
3. Be relatively resistant to changes in temperature and humidity – rooms with exposed windows, direct ventilation, or inconsistent heating would inadvertently create new variables that could affect the baseline control data.
4. Be available for testing vocalists over a span of two weeks – once started, the experiment cannot change the location in which it is executed.

---

[1] It is valuable to note that a cappella groups do not tune to equal temperament, but rather to just intonation. This latter tuning system relies on mathematical relationships between notes to define the intonation of each pitch, instead of the twelve-tone system found on a keyboard. Since this study does not discuss harmony in terms of intervallic content, the issue of tuning systems does not have an effect on our data.

The Recording Studio of Room 225 in Northeastern University's Shillman Hall was selected. Also known as Shillman Studio, it is a space on the university's campus that allows students to record music, but it is also one of the few spaces that meet the criteria above. Vocalists stood and recorded in the Tracking Room while the research team sat beyond the sound-insulated wall in the Control Room, operating the software and issuing directions to a pair of headphones on the vocalist via a "talk-back" microphone.

Certain variables persisted even after selection of the space. One concern was the noise floor of our set-up. Imperfect electrical connections along our signal chain are a constant risk. To understand the noise floor, a recording of the Tracking Room was taken while empty at the end of every session. Microphone and preamplifier settings were kept at the gain they were at during sample collection.

Another concern was that of natural resonance in the space. A vocal shield was placed around the microphone to minimize vocal reverberation and maximize acoustic absorption. This reduced unintentional resonance in the studio. The presence of a full drum kit and grand piano in the Tracking Room presented undesirable resonances. Permanent members of the studio, they could not be removed, and no formal isolation hardware was accessible to separate the vocalists' physical space from the instruments. The kick drum, snares, and toms were covered by soft clothing, and the cymbals were disassembled and placed on the floor away from the microphone to dampen their resonant power. The piano was kept shut. Finally, a "clap test" was conducted during each recording session to identify any unique resonances – two claps were recorded in nine sections of the room. This allowed for the creation of an "acoustic map," where certain frequencies could be identified as more or less prominent in different areas of the room. The clap recordings were aligned using Sonic Aligner and their spectrograms were compared for anomalies within the room and across sessions.

ii. *Range Tests*

Once spatial acoustics were identified and measured to create a set of control data, the range of each vocalist was tested. Range tests were conducted in front of a piano in the form of ascending and descending major scales. The purpose of singing scales when identifying range is to allow the voice to ease into its extremes by approaching one note at a time; it contrasts with the method of having the vocalist simply sing as high or low as they can, since this latter method may not accurately reflect the comfortable extremes that can be reached after warming up.

First, their lowest extremity was identified: in solfege, the scale was "sol-fa-mi-re-do" and descended by semitones until their voice became gravelly. They then sang the lowest note that they could sing each of the five primary vowels on – for example, if they could sing a note on each vowel except "oo," they moved up one semitone at a time until all five vowels were producible. This note is only producible in their modal voice.

Second, their highest extremity was identified. The solfege scale used was "do-re-mi-fa-sol-fa-mi-re-do" and ascended by semitones until their voice became squeaky. Then, they sang each of the five vowels on their highest note – if they could not achieve all five vowels, then they moved down one semitone at a time until they could. This note is only achievable in their falsetto.

Once the lowest and highest extremities of their range were identified, “low,” “middle,” and “high” notes were selected to act as examples for each register and synthesis mode. The low note was chosen to be a perfect fifth above their lowest note – this meant the note was comfortably in their chest voice. The high note was chosen to be a fifth below their highest note, so they could sing it both in their falsetto and in their belt. The number of semitones between low and high notes was counted, halved, and then added to the low note (or subtracted from the high note) to find the middle note. In the case of an odd number of semitones from low to high note, when the middle note is calculated to fall between two notes on a keyboard, then the final choice of middle note between the two options was decided on which pitch could be controlled more precisely.

iii. *Sample Collection*

Each of the vocalists was recorded singing the five primary vowels on their three unique notes in the three placements: bright, neutral, and dark. The seventy-five individual vowel permutations per person of these cases are shown in Figure 3.

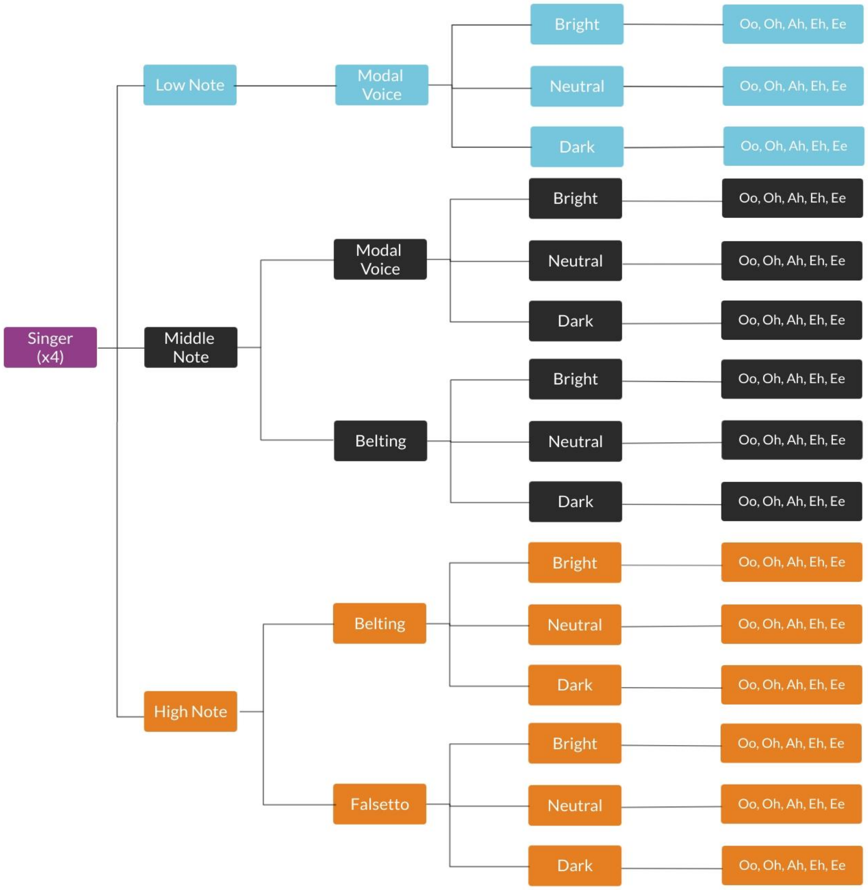


Figure 3. Chart of Individual Vowel Tests per Brightness Placement, Vocal Mode of Synthesis, and Pitch

Vocalists were then recorded singing the five vowels on their middle note in modal voice as a continuous morph from bright to neutral to dark over twelve beats at 120 beats per minute, as shown in Figure 4. This was analyzed in a spectrogram, which charted the relative amplitude of frequencies over a time duration. They were finally recorded doing the same spectrogram morph test, but on a note given by the experimenters, with the intention of digitally building a chord over parts for auditory study. These morphs yielded ten additional datasets per vocalist. In total, with seventy-five individual vowel samples and ten continuous placement samples per person, there were 340 samples to process and analyze.

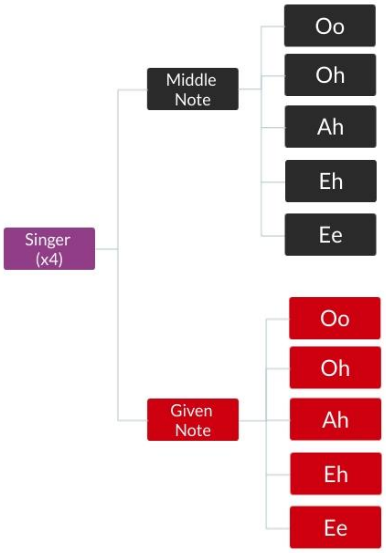


Figure 4. Chart of Continuous Vowel Spectrogram Tests per Pitch

iv. *Processing, Analysis, and Practical Summary*

1. **Selection.** Each sample represents 5 vowel sounds separated by pauses at 1 placement using 1 technique at 1 note. Recordings were repeated per sample until a satisfactory take (or recording) was captured. Satisfactory takes were judged on being: replicable by the vocalist to indicate precision, steady in the desired pitch for accuracy, opinion of the vocalist for performance, and having the necessary intensity and duration for reasonable analysis. Written documentation during the recording process indicated satisfactory takes per sample (see Appendix D). A later session was used to re-examine takes independent of the vocalist and export all audio samples from the Pro Tools file. Best takes were selected per sample, organized in a separate directory, and transferred to another computer at a different location for the remainder of the process.

2. **Trimming.** 1 ideal take of each of the 75 samples are imported into 1 Logic Pro X project for each of the 4 vocalists. Samples are then trimmed into 500 millisecond clips of each vowel sound with minimal fade-ins and fade-outs to prevent digital clipping. This clip represents 1 combinatorial case of all vocal variables. Once all clips are prepared, they are exported to the same lossless format of their import (24bit WAV and 44.1kHz). A batch of clips refers to the collection of 15 clips representing all combinations of note, technique, and placement per each vowel sound (i.e. groups made up of the “oo,” “oh,” “ah,” “eh,” and “ee” permutations).

3. **Fast Fourier Transform (FFT).** A program in MATLAB_R2020a (Appendix F) is written to recursively do the following for a batch of clips:

a. Select the clip from the directory and further trim the clip to a 100-millisecond median slice to reduce the effects of vibrato and pitch-drifting of the performers. In terms of sonic envelope, this is the sustain portion of ADSR. The transient variables of the attack, decay, and release are removed from the clip.
b. Plot the slice in amplitude vs. time graph for confirmation.
c. Calculate the FFT of the full slice at a sampling frequency of 44.1kHz and store the frequency and power data in a two-column array. This sampling rate creates a Nyquist frequency at 22.05kHz, which is above the upper limit of human hearing, making it a practical choice. Generate a label array from the clip name and vertically concatenate it with the numerical values for future identification.
d. Select the next clip and repeat the calculation, now horizontally-appending a labeled power data array to the previous array.
e. Once all clips are processed, output the fully-appended array as a tab delimited .txt file in the same directory, containing the FFT data for the entire batch.

This program is run 5 times for each vowel sound per vocalist. With all 75 cases now represented by numerical FFT arrays, quantitative analysis can begin.

4. **FFT Analysis.** The text files containing the FFT data for 1 vocalist are imported in 1 excel spreadsheet. It is an endless task and spatially irresponsible to produce and present discernible 75 FFTs. An overlay, even if divided into 5 separate vowel sounds, proved to be indecipherable and approached a computational threshold. For this reason, it became necessary to identify and perform a more rigorous analysis on only outstanding cases within the full set of data.

Outstanding cases are defined as: (A) Uncertain subsets of vowel sounds that represent the greatest range in harmonic content, (B) The extremes of sonic brightness for each note, (C) Cases that help define the multivariable picture of sonic brightness. The judgement and representation of these three novelties are thus:

a. *Uncertain Sounds.* The following is for 1 vowel sound at one note for one technique (ex. “Ah” belted on the middle note). FFT data for the three placements of each vowel sound are averaged into a mean FFT (mFFT). The standard error of the mean (SEM), σ_x, is calculated at each frequency, where σ is the standard deviation of the set, and N is the number of elements.

$$\sigma_x = \frac{\sigma}{\sqrt{N}} \quad . \qquad \text{(Eq. 1)}$$

The mFFT and SEM are divided by the maximum power of the mFFT to normalize the data.

The sum of all SEMs over the full frequency spectrum is used as an indicator of the harmonic uncertainty of that vowel at that note using that specific technique. The harmonic uncertainty is compared for all vowel sounds at the different techniques for the specific note. The condition that produces the greatest harmonic uncertainty presents a novel snapshot for further examination. The 3 placements of the most harmonically uncertain condition are individually normalized and plotted on the same graph to compare their harmonic content, understand the cause of the harmonic variation, and recognize how the most uncertain condition reflects the homogeneity of each vowel sound.

A major assumption of filtering by the greatest uncertainty is the comparative uniformity of all other samples. Samples of lesser uncertainty will assume to present conditions less affected by placement. Further studies may look at the uncertainty of all samples.

b. *Sonic Brightness*. How can the sonic brightness of cases be quantitatively compared? Within our definition, a perfect sine wave with a fundamental frequency at the examined note would be the least sonically bright sound of the group. With all data sets normalized, a sound with more power in the overtones would be expected to be ranked brighter against a sound with fewer overtones. Also, a sound with just a first harmonic would be expected to be less bright than a sound with just a second harmonic at the same power. From these assumptions, it can be assumed that brightness is directly related to the area under the FFT, or the spectral density.

For ranking, three quantities seemed appropriate: the spectral density, the centroid of the power spectrum, or the spectral density with squared-power. Each has its merits, but when comparing the final rankings, the power-squared ranking was found to be impractical. It minimized the weaker, upper harmonics so much so that they were effectively not present, which biased our rankings to favor sounds with an exceptionally powerful lower harmonic. Spectral density and centroid are closely related, with centroid being calculated as the frequency at which the integral of the FFT equals half of the full spectral density. The two calculations produce similar rankings, but differ in their biases – ranking by spectral density favors sounds with more numerous harmonics at higher frequencies, whereas centroid favors sounds with fewer, but more powerful harmonics. The two seem equally valid and produce reasonable rankings. Sonic brightness in our context seemed to be a two-variable quantifier.

However, in terms of ranking, it seemed spectral density better reflected one assumption of a bright sound – high harmonic content. Thus, conditions per note were all normalized and then ranked accordingly. The power spectra of the maximum, minimum, and median densities were plotted on the same graph to visualize the range of sonic brightness for the note.

c. *Spectrogram.* A spectrogram is a three-dimensional graph in the time domain representing power as color within the range of frequency. Though its image does not present specific quantities, it offers a vibrant visual representation for sound qualitatively, and is a great tool for comparing a sample's changing spectral density with respect to time. It is assumed spectrogram visualization of our continuous morph samples may offer a more objective perspective than just listening to the samples. Spectrograms are used to identify spectral similarities and contrasts, and hopefully provide an intuitive check on our findings from the FFT analysis. Spectrogram analysis is focused on the middle modal voice, assuming it is the most frequently used by vocalists, and examine only the given note. The "Melodic Spectrogram" of Sonic Lineup is chosen as the graph type because it emphasizes the harmonic content more clearly than the basic spectrogram.

This analysis is done for each vocalist to understand their individual intricacies.

5. **Harmonic Summary.** The purpose of this analysis is to identify novel cases and general trends that offer practical aid to vocalists, arrangers, and directors. By comparing all parts, a clearer relationship between vowels, techniques, and placements can be identified with a quantifiable sonic brightness.

V. RESULTS

i. *Noise floor*

By taking a Fourier transform of the noise floor sample from each session, the uncontrollable noise of our electronics and space can be understood. Averaging them together and calculating the standard deviation, the power spectrum of the average noise is plotted with SEM bars in Figure 5.

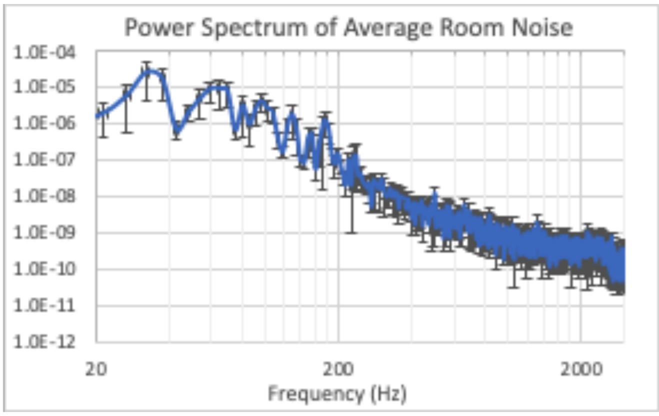


Figure 5. Average Room Noise with Error Bars

From our figure, three conclusions can be drawn. Firstly, that the magnitude of the noise is negligible in comparison to our experimental samples, especially at higher frequencies. Secondly, if the effects of noise on our analysis are considered, they would be most disruptive at frequencies below our domain of analysis and on the lower edge of human hearing. Thirdly, room noise is uniform across all four sessions.

ii. *Clap tests*

The spectrograms of the clap test from each of the four recordings are used to identify any regions of peculiar resonance or bias. The four clips are analyzed using Sonic Analyzer and their spectrograms are qualitatively compared in Figure 6. Each pair of lines represents two claps simultaneously in that section of the room. One-off lines were accidental claps that were not created with the standardized clapping technique.

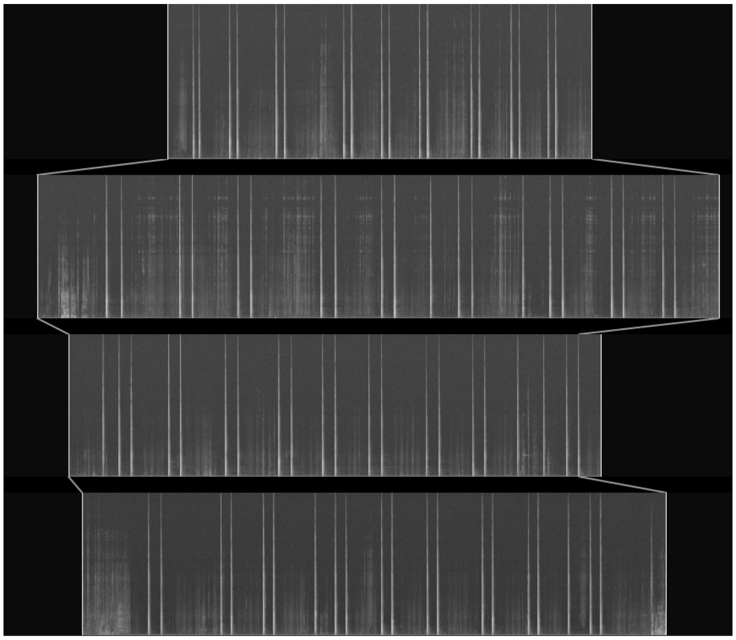


Figure 6. Spectrograms of Clap Resonance Test
(Top to Bottom, in order of latest session date: Bass, Mezzo, Tenor, Baritone)

Our comparison directs us to assume there are no exceptional quadrants within the room. With the y-axis being logarithmically-scaled frequency, x-axis representing time, and lightness an indication of power, the claps in front of the microphone show the greatest uniformity of power across the frequency range. There is also a tighter envelope, with more clear divisions around the attack and decay of the clap, indicating less spatial influence and a more direct translation of the physical production of the sound.

With these findings, there is high confidence that the data of each vocalist was collected in a reliably uniform environment, with minimal noise, and in the most direct position relative to the microphone.

BASS

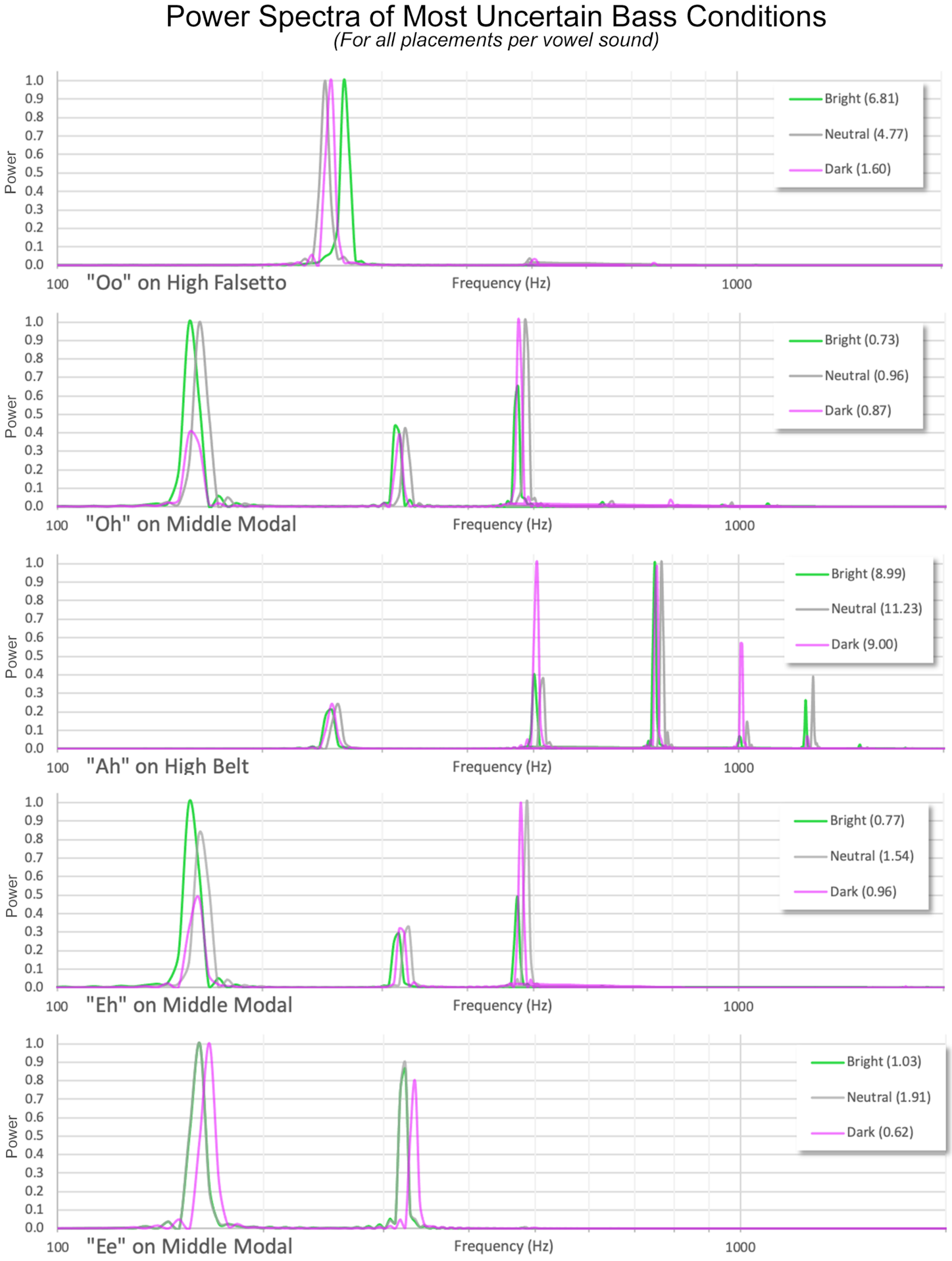


Figure 7. Comparative composite graph plotting placements of the most uncertain Bass vocal conditions. (Legend is defined as: '[Placement] [n=Normalization Constant]'. Domain is restricted to areas of activity.)

By examining the most variable sets of placements for each vowel sound, novel phenomena, or lack thereof, present in the bass' range and technique can be identified.

"Oo" on High Falsetto appears to be the most variable condition for that vowel, yet it is stable over all placements, with only a weak 2nd harmonic missing in the bright placement. Thus, that vowel is likely harmonically uniform across our bass' range and techniques, with the greatest difference arising from power. The stable, sine-like harmonic structure of "Oo" can be compared to "Ee" on Middle Modal, as both exhibit little variation. In applied terms, this means that these two vowels, "Oo" in falsetto and "Ee" in modal/chest voice, are good choices for basses who are tasked to sing in their high and middle registers, respectively.

These two are contrasted with the dynamically complex differences of "Ah" on a High Belt. Placement greatly affects the structure and, therefore, sound of this vowel. While all the FFTs share a significantly weak fundamental frequency and a peak at the 3rd harmonic, the distribution of power across harmonics varies greatest. Dark placement has the greatest distribution across the spectrum with a peak at the 2nd and a unique 4th harmonic, which is barely present in neutral. Bright and neutral placements share a weaker 2nd harmonic and similar presence of a 5th harmonic. This level of inconsistency indicates that the openness of the "Ah" vowel on a High Belt means neither the vowel, nor likely the high register, are good choices for a bass to sing.

"Oh" and "Eh" on Middle Modal exhibit similar patterns of variation: Dark placement dampens the fundamental, bright placement dampens the 3rd harmonic, neutral placement has the greatest normalized spectral density, and all exhibit a weak 2nd harmonic.

Normalization constants derived from the maximum power offer a perspective on power. For these cases, the normalization constant is generally greatest at the neutral placement. In the two aforementioned cases, a neutral placement offers the most present peaks.

From this comparison, Middle Modal appears often as the most harmonically variable and uniformly powerful position for the bass; however, it is worth noting that the Low Modal voice did not appear among the most variable cases. Therefore, the best advice for arrangers is to keep the basses singing low. Looking to Figure 8, directors could potentially benefit from the bass's resonance when using a bright placement in the Low Modal voice, although while paying attention to the bass's audible power when bright versus neutral or dark.

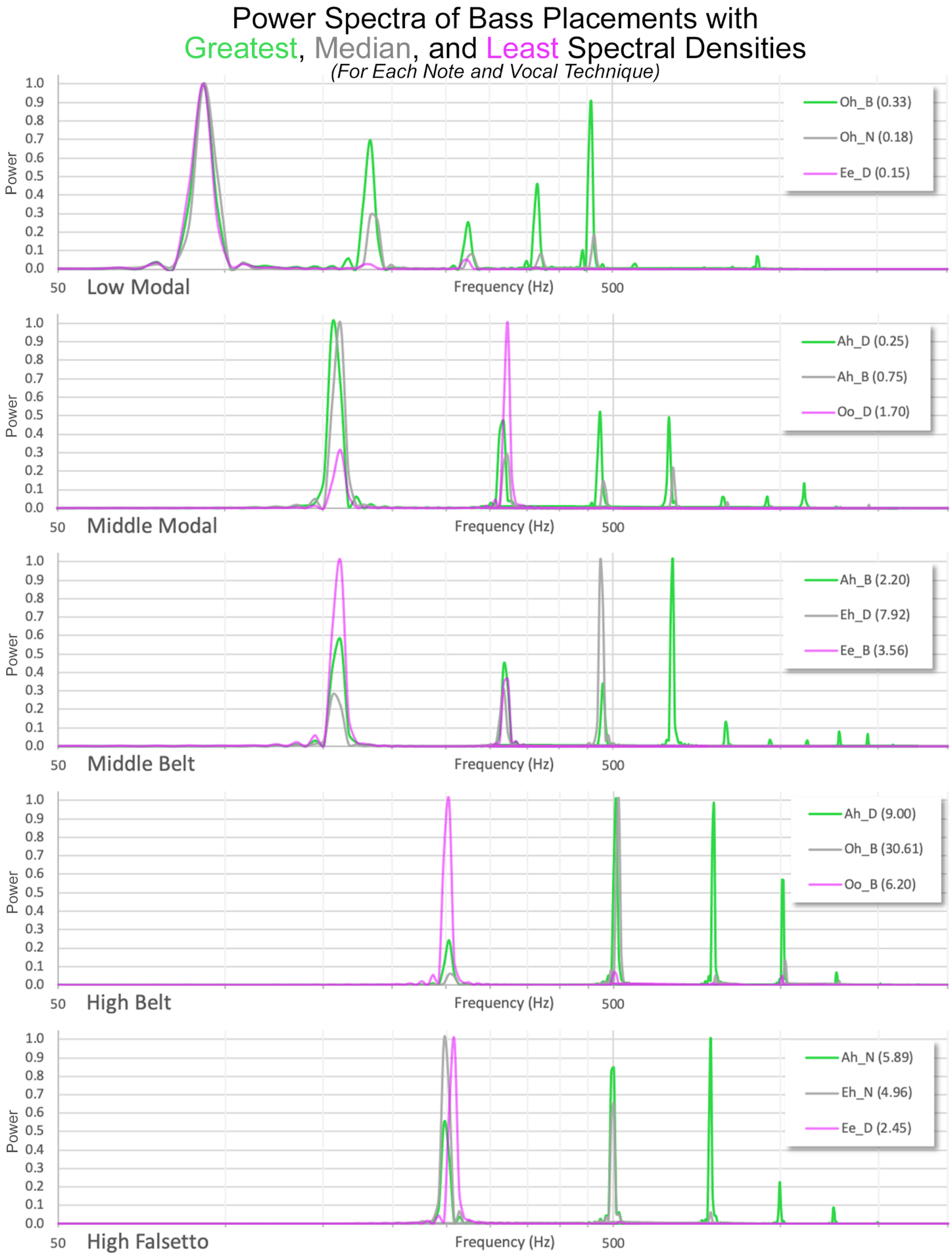


Figure 8. Comparative composite graph plotting Bass vowel placements for each range and technique. (Legend is defined as: ‘[Vowel_Placement] ([n=Normalization Constant])’. Domain is restricted to areas of activity.)

All voices present a broad range of spectral density over vowel and placement combinations. For Low Modal voice, there is a direct correlation between sonic brightness and upper harmonic presence. As spectral density increases for the Middle Modal voice, harmonic distribution shifts from a concentration at the 2nd harmonic to the fundamental frequency and more high frequency energy is introduced. Greater spectral density in this case reflects a wider distribution. Upper harmonic presence is also present in the Middle Belt; however, the fundamental frequency decreases with density. This pattern is very similar in the High Belt and High Falsetto voices. Spectral density in these cases is a result of diminished lower frequencies and excited upper frequencies.

The colloquial term “brightness” (as referenced in “bright/neutral/dark placement") is analogous to greater spectral density or overtone energy, yet our analysis of the bass does not correlate the two. While the orderly ranking of greatest, median and least cases of Low Modal presents an ideal example for supporting the term, other voices contradict it. Dark and neutral placements are the most spectrally-dense placements by significant measure in the Middle Modal, High Belt, and High Falsetto voices. The Middle Belt and High Belt both present bright placements ranked lowest in spectral density or “sonic brightness.” Bright placement does not present a strong direct relationship with spectral density over these vowel sounds. Thus, it can be said that for a bass to achieve the most solid, resonant tone while belting, the neutral and dark placements in middle and high registers are best to use.

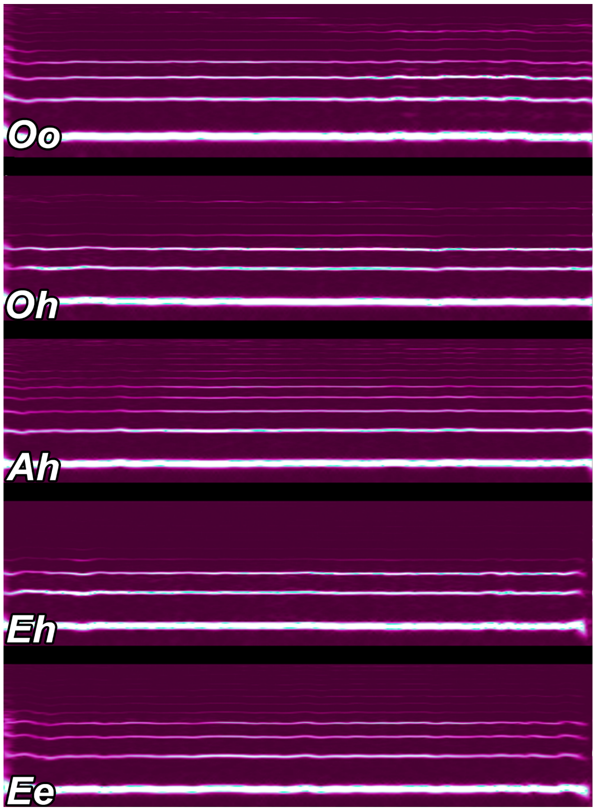


Figure 9. Spectrograms for Bass vowels over 6s continuous vocal morph from bright to dark placement.

These rankings do allow us to identify “Ee” as the darkest (or least spectrally dense) vowel generally across most vowels, with “Oo” a close second. On the other end, “Ah” tends to be generally ranked highest in spectral density. Unlike the previous comparison, power does not present any strong correlations with our variable, spectral density.

Spectrogram analysis presents a vivid picture of the bass’ resonance and foundational stability (Figure 9). Power is significantly concentrated at the fundamental with weakest harmonics at the bright placement consistently over all placements and vowels. “Oo” is a unique case where dark placement, contrary to the intent of the term, introduces more overtones. All other vowels present their strongest overtones at the neutral placement. Middle Modal voice presents rich harmonic content, a well-supported fundamental, and a powerful consistency over all variation.

BARITONE

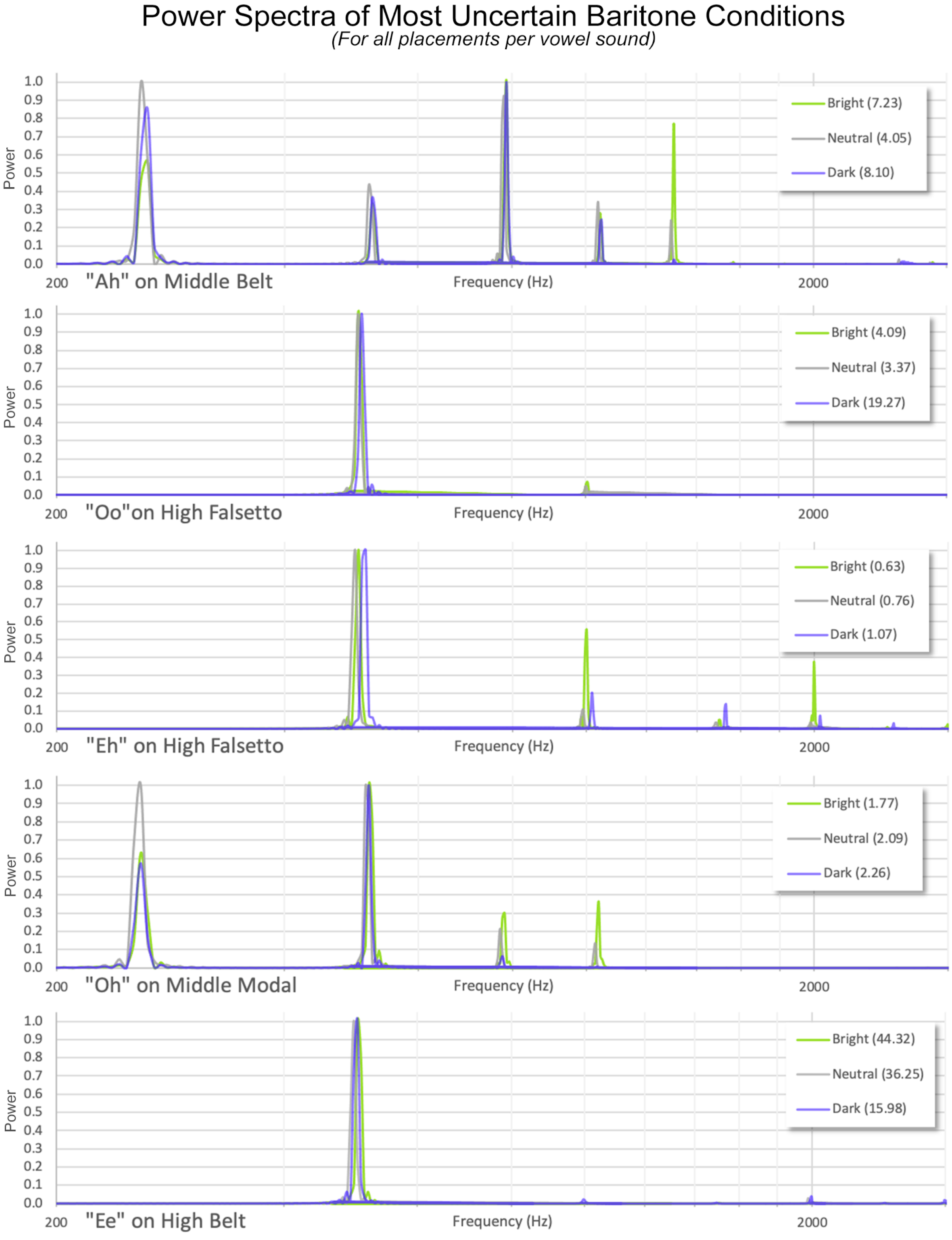


Figure 10. Comparative composite graph plotting placements of the most uncertain Baritone vocal conditions.
(Legend is defined as: '[Placement] [n=Normalization Constant]'. Domain is restricted to areas of activity.)

From a bird's-eye view of the Baritone's most uncertain conditions, the vowels "Oo" and "Ee" are the most uniform in regards to placement. Both have a near absolute concentration of harmonic energy at the fundamental across all placements, suggesting uniformity across all conditions for those vowels. This indicates that "Oo" and "Ee" are good choices of vowels for baritones to sing in any placement and register.

Other conditions distribute harmonic energy to create more uncertainty across placements. Examining "Oh" on the Middle Modal voice, we see that the bright placement has a strong 2nd harmonic, a comparatively weaker fundamental, and active 3rd and 4th harmonics. Dark placement lacks the energy of those upper harmonics. With the exception of a uniquely strong fundamental peak, neutral placement closely resembles the harmonic structure of the bright placement, albeit slightly weaker in overtones. In these conditions, it seems bright placement relies on an effect akin to a high-pass filter on the neutral placement. The effect of placement on the vowel sound "Ah" at Middle Belt voice is very similar to the Middle Modal "Oh". Bright and dark placements are harmonically similar and both lack fundamental energy, which neutral offers. Aside from a uniquely high peak above the shared structure of the middle harmonics, bright placement correlates steadily with greater overtone power. Because the neutral placement on "Ah" and "Oh" have the strongest fundamental, a baritone would sing with the fullest tone in this register when in this placement; a dark placement would likely "muddy up" an overall group sound because of its lack in upper harmonics; a bright placement would minimize this low-end boom and instead add a "tinny" presence to the group sound because of its emphasized upper harmonic.

"Eh" on high falsetto is an interesting case. While bright placement concentrates significant energy on the 2nd and 4th harmonics, dark placement presents the greatest number of overtones, along with additional 3rd and 5th harmonics. Neutral placement is the most pure with greatest concentration at the fundamental. The differences in structure represent significant variability in these conditions and represent the dynamic possibility in "Eh" as well as "Ah" and "Oh" vowel sounds.

A comparison of normalization constants for each placement of the most uncertain conditions offers comparisons of overall power. In regards to the baritone, there does not seem to be a direct relationship between maximum power and placement over all vowels; however each vowel set offers its own insights. The "Oo" High Falsetto possesses the greatest overall power, as well as the greatest range of power over all placements. Oppositely, "Eh" on the High Falsetto voice has the lowest maximum power and the smallest range. Considering these factors, we find that the baritone's falsetto voice at the upper range represents the greatest range in potential power.

Next, the maximum, median, and minimum of the 5 note/voice conditions are compared (Figure 11). There are strong relationships between harmonic features and spectral density for the baritone. All voices present the greatest number of harmonics and widest energy distribution on the most spectrally-dense cases. As spectral density decreases, the presence of these harmonics decreases to absence.

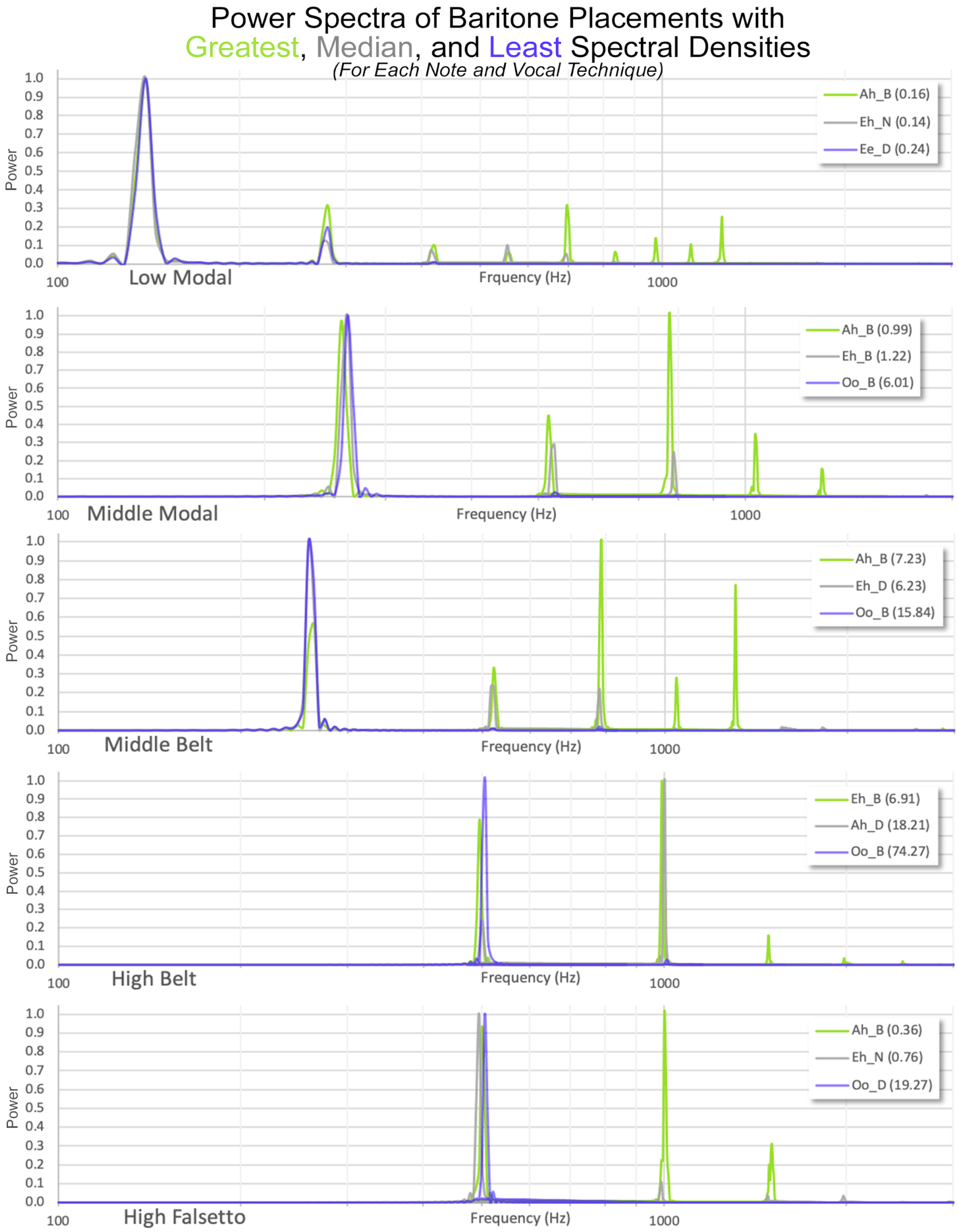


Figure 11. Comparative composite graph plotting Baritone vowel placements for each range and technique. (Legend is defined as: '[Vowel_Placement] ([n=Normalization Constant])'. Domain is restricted to areas of activity.)

Most voices also present a greater concentration at the fundamental as spectral density decreases. High Belt and Middle Modal offer unique cases where neutral placement diminishes the fundamental greater than the bright placement; however, the majority of the structures demonstrate neutral placement as a middle ground between bright and dark placements.

Comparing the normalization constants of each composite graph, a general indirect correlation with spectral density and maximum power is noticeable. This relationship is most dramatic on the High Belt, where there's a range of 65 between maximum power between the greatest and least spectrally dense samples of the vocal technique. Low Modal has the smallest range in power, signifying it as the most stable of the voices, albeit the weakest.

There are clear conclusions based on extreme vowel sounds. "Ah_B," or "Ah" sung on a bright placement, is frequently ranked as the brightest sound or close to the brightest sounding vowel sound for all voices. "Oo" sits at the opposite end of the spectrum confidently, but has variation based on placement. Most frequently the median, "Eh" offers the most neutral density of vowels.

Again, "bright" placement does not directly correlate with bright, or spectrally dense, sounds. Bright placements are the lowest-ranked in density for several voices and often are not the brightest-sounding placements for individual vowels.

The spectrogram of the baritone's Middle Modal placement morphs supports some of the previous observations (Figure 12). "Ah" clearly has the most number of strong harmonics at its brightest of all vowel sounds. Only its two highest overtones are affected by placement and are nearly absent by the neutral point. "Oo" is the one case where darker placement introduces greater overtone power. For all other vowels, there is an overtone bias to the brightest placement that diminishes as placement darkens. A powerful fundamental is constant through all vowels and placements. "Ee" presents the purest and most consistent sound with most concentration at the fundamental, with a very weak single harmonic that fades further through dark placement; this uniformity is suggested by the stability of its most uncertain case on high falsetto. Alternatively, "Oh" demonstrates its uncertainty over placement with more nuance than expected – generally consistent harmonics seem to fluctuate with power over placement, but never fully diminish or shift in any direction.

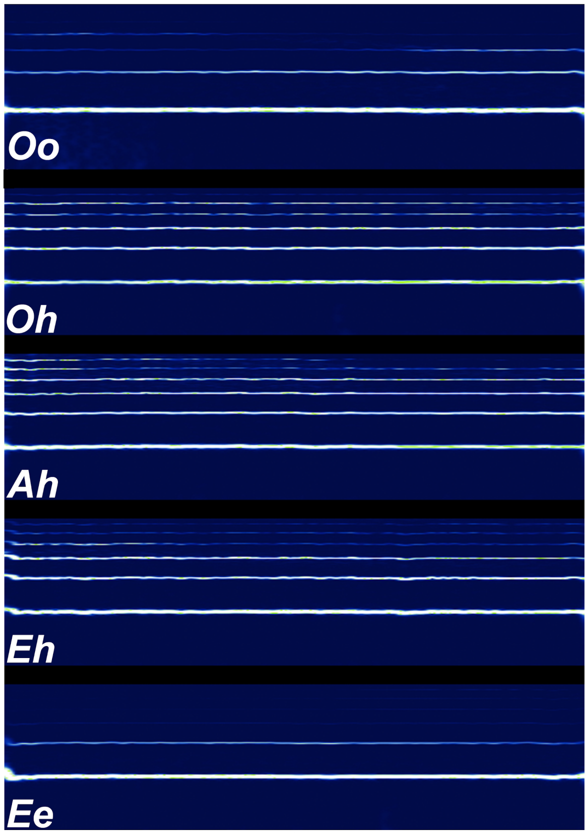


Figure 12. Spectrograms for Baritone vowels over 6s continuous vocal morph from bright to dark placement.

TENOR

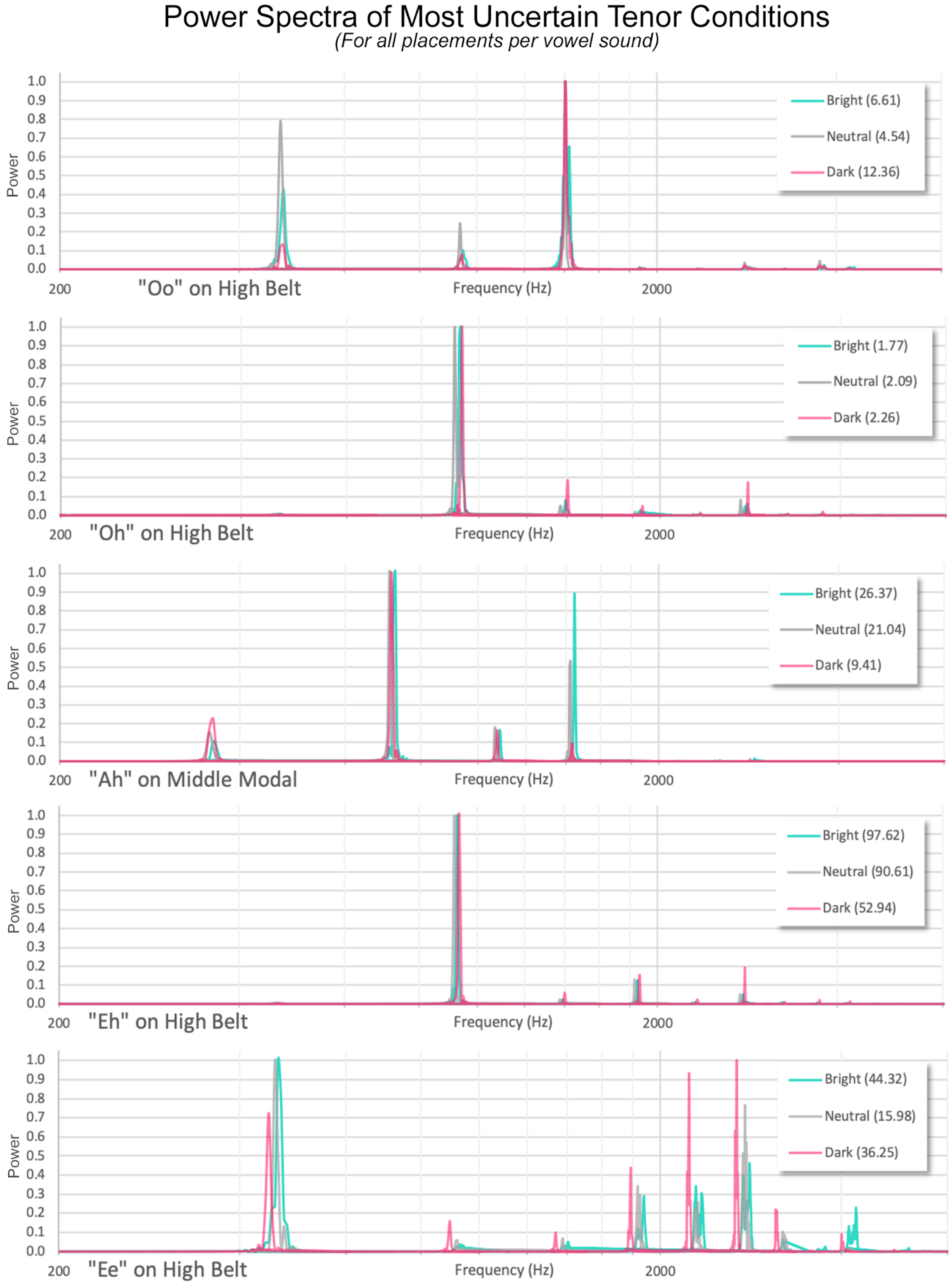


Figure 13. Comparative composite graph plotting placements of the most uncertain Tenor vocal conditions. (Legend is defined as: '[Placement] [n=Normalization Constant]'. Domain is restricted to areas of activity.)

The most uncertain cases of the tenor exhibit curious phenomena with little opportunity for general conclusions (Figure 13). For the most part, dark placements possess greater overtone energy and weakest fundamental presence; this is only contradicted by “Ah” on Middle Modal, where darker placement relates to the opposite. “Oo” on High Belt has the greatest concentration at the 3rd harmonic. “Oh” shares a strong fundamental energy across placements, with similar harmonic structures indirectly varying with brighter placement. “Eh” on the High Belt has comparable properties, but a uniquely direct relationship with overtone and power. “Ee” on a High Belt with a dark placement presents the relationship to an extreme: the 6th harmonic is the strongest frequency for dark placement, with power increasing as harmonics ascend to it and then decrease quickly as harmonics continue. Bright and neutral placements for this voice share this property to a lesser extent and peak at their fundamental. Bright does present a uniquely strong 8th harmonic. For “Ee” on High Belt, the peaks are less defined and display an increasing amount of partials as frequency ascends – a feature unique to this case for all ensemble members. In most voices, neutral placement acts as a middle ground between bright and dark; however in “Oo” on the High Belt it possesses the strongest fundamental and least overtone energy.

There is no clear relationship between placement and power. The tenor has the most nuanced variability between placement and vowel of all ensemble members. By examining the most variable cases, it was assumed that conclusions could be derived for the more uniform cases; however, the tenor’s comparisons do not offer coherent or meaningful suggestions. It seems the only takeaway is that the High Belt is an extremely inefficient note/voice combination, and should ideally not be used in an arrangement or performance.

Examining the tenor by range and voice presents a clearer picture with some suggestion towards trends (Figure 15). In all positions, the most spectrally-dense case corresponds with the most harmonically-active structure. This is most clear in the Low Modal voice, where brighter placement has more overtones. Generally, the least spectrally-dense material has the greatest fundamental frequency concentration. “High Belt” presents a lone position where the opposite is true – the most spectrally-dense case is the only one to have a fundamental frequency, while also having the greatest overtone energy.

“Ah” is often the most spectrally-dense vowel, while “Oo” is generally the least. Spectral density doesn’t correlate with power or placement. A neutral placement is brightest-sounding for the Middle Belt and darkest-sounding for both high voices.

The range of powers for each voice demonstrates the tenor’s dynamic quality. Low Modal is the most consistent and decidedly the weakest. Conversely, High Belt has the greatest range by far and the greatest power of all voices in its least dense case, even if it is technically inefficient.

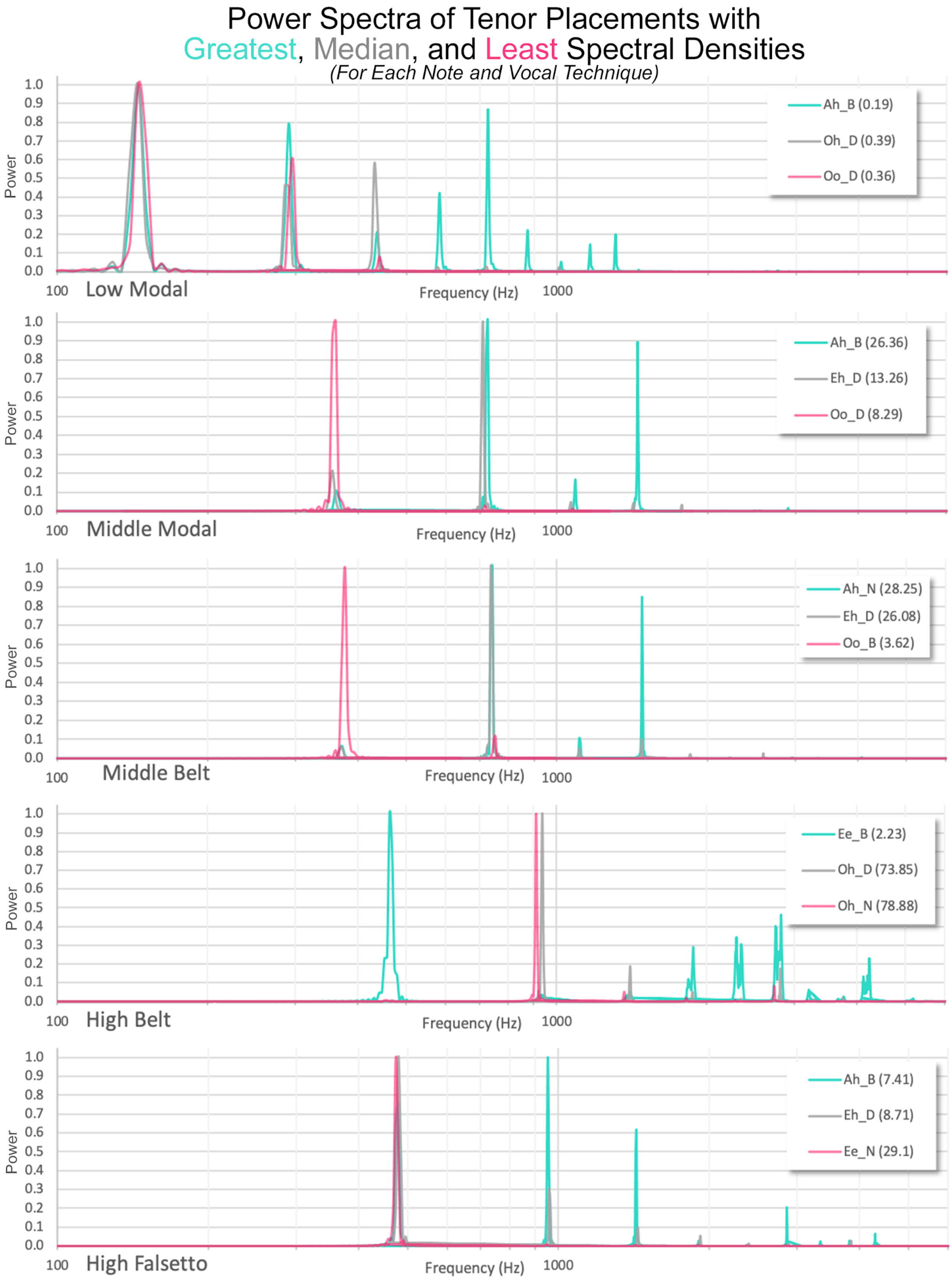


Figure 15. Comparative composite graph plotting Tenor vowel placements for each range and technique. (Legend is defined as: '[Vowel_Placement] ([n=Normalization Constant])'. Domain is restricted to areas of activity.)

A nuanced range of the tenor's contrast is clarified through spectrogram analysis of the Middle Modal voice (Figure 14). For the full duration of all morphs, there is a clear harmonically-active ranking of vowels (from greatest to least): Ah, Eh, Oh, Oo, and Ee. The blurriness of the upper lines of "Ah" describe formants or harmonic noise that were not apparent in our previous analysis. The morphs of the tenor are flawed, which is verified by listening to the clip. The best of performance intentions caused the tenor to brighten the placement of his voice in an effort to continue supporting the notes and gracefully end the morph. Ideally, this grace and professionalism would be absent from these morphs, so we may understand what features are exactly undesirable in a dark placement. The instinct to a brighter placement suggests a conventional preference for the overtones present in a tenor's middle modal voice; however, this is a subjective conclusion.

That said, there does not appear to be much variation in harmonic content for any vowels in the first half of all morphs, suggesting a similarity between bright and neutral placements. The morph of "Ah" presents a similar feature to the previous FFT analysis (Figure 13) – a diminishing highest overtone in the dark placement. Overall, unlike previous ensemble members, the fundamental is not consistently powerful over all vowel sounds, and spectral density is more clearly spread out across the numerous overtones.

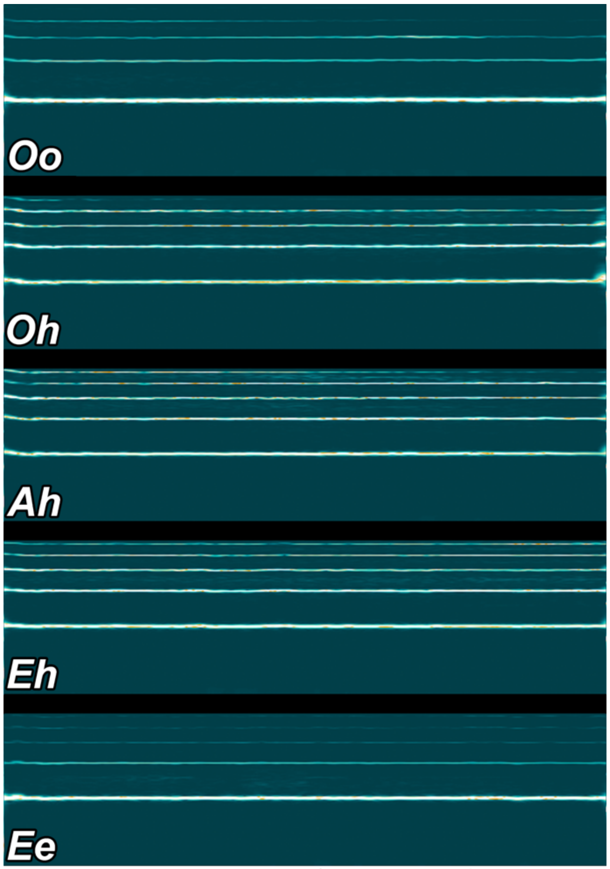
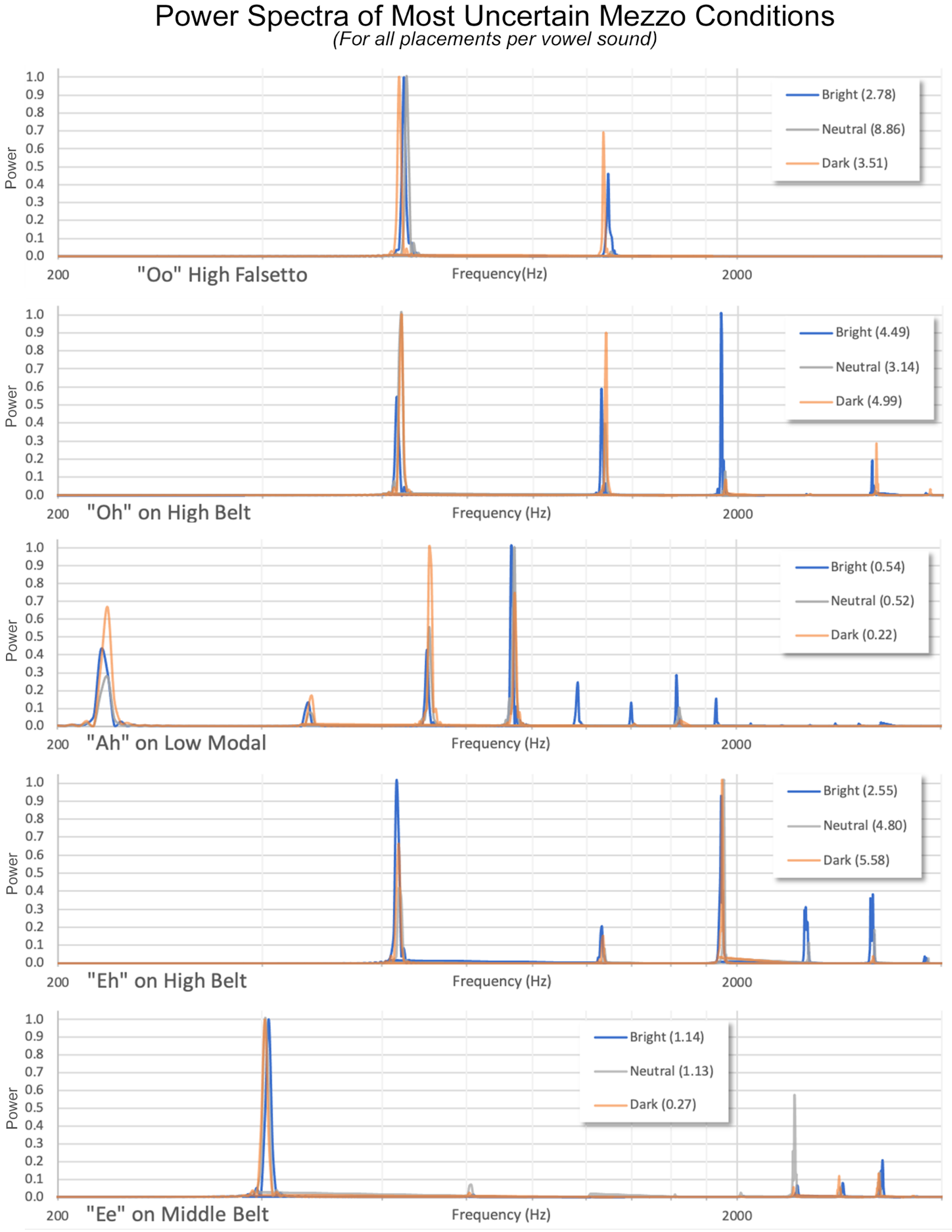


Figure 14. Spectrograms for Tenor vowels over 6s continuous vocal morph from bright to dark placement.

MEZZO



Power Spectra of Most Uncertain Mezzo Conditions
(For all placements per vowel sound)

Figure 16. Comparative composite graph plotting placements of the most uncertain Mezzo vocal conditions.
(Legend is defined as: '[Placement] [n=Normalization Constant]'. Domain is restricted to areas of activity.)

The most uncertain mezzo voices display a suggestion for the general behavior and uniformity of each vowel sound. “Oo” and “Ee” appear to be most uniform; the former exhibits variation in the 2nd harmonic, and the latter features varying energies in the highest-order overtones. Both demonstrate a neutral placement that deviates from the bright and dark – peaks shared by the two do not appear or the energy is shifted a single harmonic. This indicates that both vowels are strong choices in all registers and placements for a mezzo.

“Oh” and “Ah” offer a demonstration of darker placements shifting energy to the lower harmonics. Where the bright placement lacks energy (ie. the fundamental), the dark placement is more present; and vice versa, it lacks the upper harmonics. The neutral placement has the weakest fundamental on “Ah” but blends the powers of the other two placements. For “Oh,” a neutral placement offers a classic square wave type structure with a strong fundamental and diminishing odd harmonics. This suggests that the mezzo has a very direct role in “brightening the sound” of a group. If dark placements can predictably lead to a downward shift in harmonic energy, then arrangers can control how “dark and choral” or “bright and contemporary” the group sound is through the mezzo part; directors can make a passage lighter or heavier based on the mezzo’s placement; and mezzo singers themselves can serve as reference to other voice parts that are seeking to fit the overall group sound.

Bright placement for “Eh” on the High Belt has the strongest fundamental and the most overtones, a unique combination. Dark and neutral placements are weaker on their fundamental and lack those overtones; however, their power is about twice that of the bright placement. This means the mezzo can likely be loud on a high belted dark or neutral “Eh,” but the bright “Eh” will give the fullest tone to the sound.

Almost all voices are represented, with only “High Belt” appearing twice as the most variable. This suggests that no voice is significantly more inconsistent than the others. For all vowels and placements, normalization constants are low and have smaller ranges between placements, suggesting a strong uniformity of power for the mezzo. In short, the mezzo has a consistent, predictable impact on the group sound across her entire range.

The power spectra in Figure 17 for mezzo placements over notes and techniques illustrate the harmonic range of the mezzo. It is especially present in the Low Modal voice, where the most spectrally-dense case has a present fundamental juxtaposed by stronger 2nd and 3rd harmonics with energy in high order overtones barely present in the High voices. Neutral and dark placements for the Low voice lose the overtones – a feature of all mezzo voices – and exhibit stronger fundamentals, which is also exhibited in the Middle Belt. With the densest case features powerful overtones unique to itself and as placement shifts darker, the density concentrates closer to the fundamental. Similarly, Middle Modal, High Belt and High Falsetto shift density towards the fundamental as placement darkens.

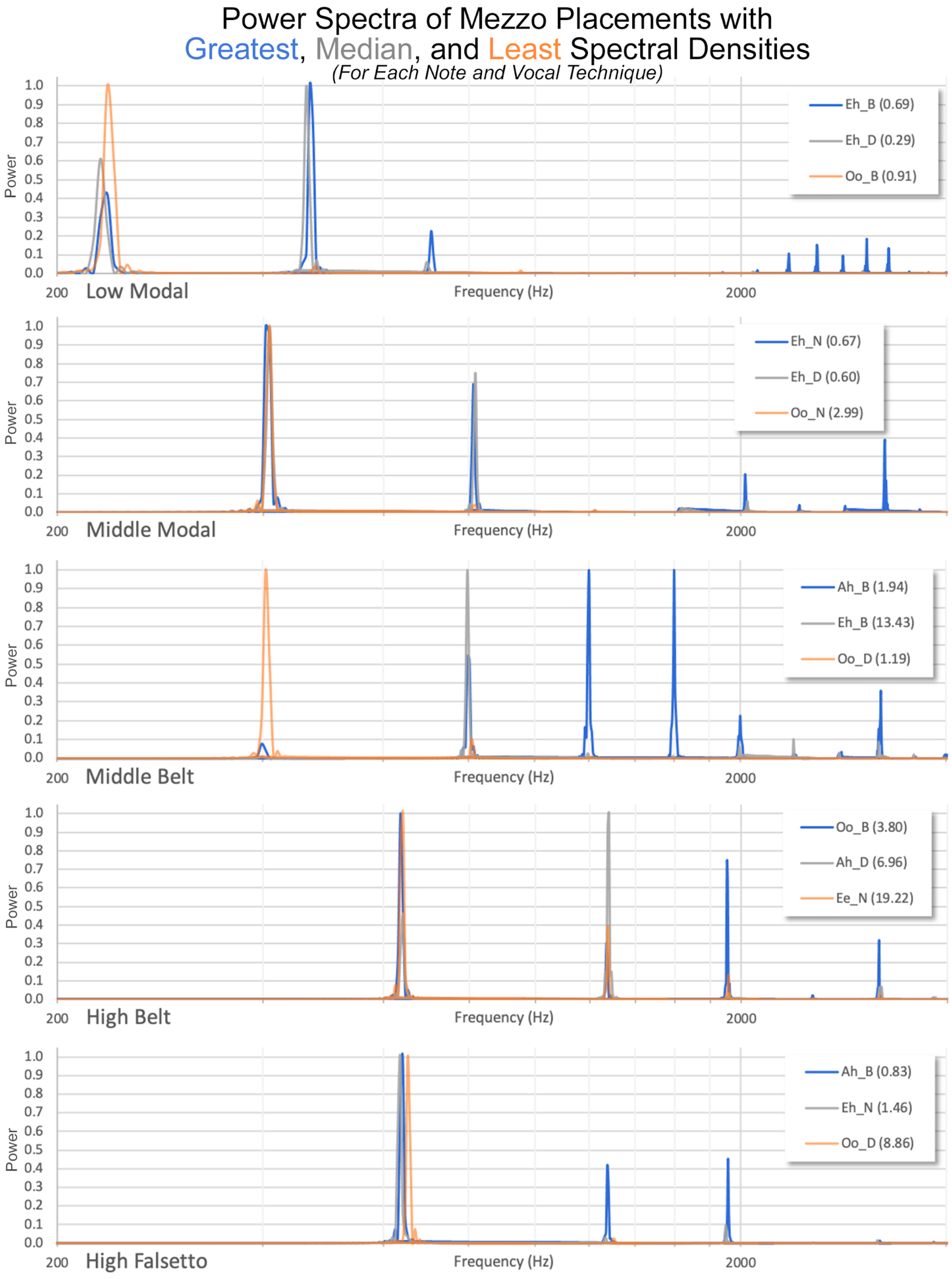


Figure 17. Comparative composite graph plotting Mezzo vowel placements for each range and technique. (Legend is defined as: '[Vowel_Placement] ([n=Normalization Constant])'. Domain is restricted to areas of activity.)

More than other ensemble members, the mezzo has a more concentrated range of power. The normalization factors indicate power is generally lower, and the greatest range is approximately 13. What also sets the mezzo apart is the number of harmonics and her range. While the mezzo is the most uniform in power, her voice offers a complex and dynamic range of harmonic structure, primarily through its most spectrally-dense cases.

Rankings amongst vowel sounds and placement vary greatly in terms of spectral density or sonic brightness. “Oo” represents both brightest sound for the High Belt and darkest sound for all other voices regardless of placement. It is difficult to draw conclusions on brightness and vowel sound. Brighter placement tends to favor brighter sounds; however, it also represents some of the darker sounds for each voice.

The consistency of the mezzo is pictured in her placement morphs (Figure 19). Bright spots at the beginning and ends of lines may be ignored; they reflect the envelope of the sound and are not within the scope of placement. “Oh” and “Ah” appear to have the most harmonic content, with the latter presenting a blip of an extra overtone. “Eh” loses some of the upper harmonics and features the comparatively weakest fundamental of all vowels. Contrastingly, “Oo” presents a similar structure, but concentrates more energy at its fundamental. Similar to all ensemble members, “Ee” is the purest vowel, with very faint harmonics and remains stable over all placements. Generally, brighter placement does seem to relate to more powerful overtones; however, there are fluctuations near the neutral-dark placement subtly apparent in most vowel sounds. The effects of placement on the mezzo over all vowel sounds appears to fit most neatly in the intuition of bright placement relating to spectrally-dense harmonic structure, colloquially referred to as “bright” sounds. The spectrogram also presents the mezzo as having the greatest consistency over all placements.

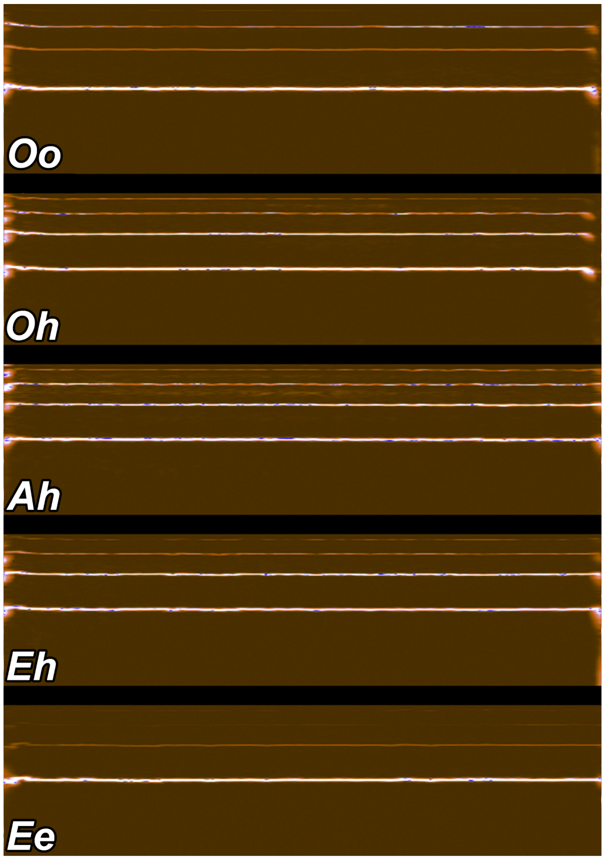


Figure 19. Spectrograms for Mezzo vowels over 6s continuous vocal morph from bright to dark placement.

THE ENSEMBLE

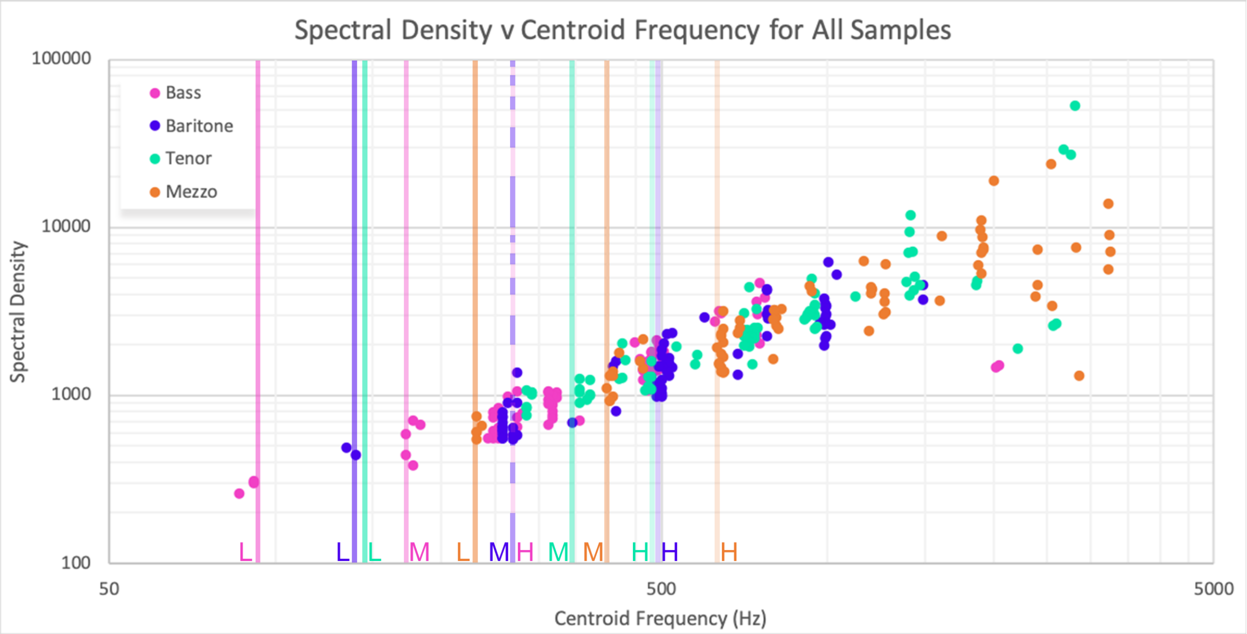


Figure 18. Plots of all vocal clips divided by ensemble members, with vertical lines indicating the approximate fundamental frequencies of the Low, Middle, and High notes of each member. Table in Appendix G.

For a vocalist, there is no harmony in isolation. Thus, we take a step back from the individual members and examine the ensemble as a whole. By plotting samples as the spectral densities of their centroid frequencies, we can see how harmonic focus and range vary between vocal parts (Figure 18).

For bass, baritone, and mezzo vocalists, there are points clustered around the fundamental frequencies of their tested notes; for example, the bass has 2 points with centroids just below the fundamental of the low note. Firstly, this means the intonation, or accuracy, of the bass vocalist was off in recording and he sang below the expected frequency. While not ideal, this is passable, as we care most about harmonic structure, not the fundamental frequency. However, their centroids being near the given note offers valuable insight – they lack significant overtones. If we carry this insight forward to examine all clusters around given notes, our data suggests potential for pure-sounding conditions – vocal combinations with harmonic structures that exhibit the most concentration at the fundamental frequency. For these clusters, and any points with similar centroids, we may understand that if their spectral density is greater, their harmonic content is greater as well. Yet, it is not great enough to significantly affect the centroid. Re-examining these clusters now, we see that the harmonic structures of the bass and baritone have strong biases to their fundamental frequencies, with overtones only slightly altering the distribution of power. They both are also solely responsible for frequencies below 200Hz.

Conversely, the tenor offers an introduction to understanding the data points with centroids that exist beyond the given note values. The lowest points of the tenor extend far above the fundamental frequency of the lowest given note. This indicates that while the tenor may sing that note, the harmonic content of their vocalization is distributed far beyond the fundamental frequency, and is likely more powerful in 3rd, 5th, or even higher order harmonics with numerous overtones. In fact, a majority of tenor data points do not cluster around given notes, but instead have centroids at their 3rd, 5th, and 7th harmonics. While the domain of the centroid frequencies may not be as great as the mezzo, the range of tenor densities is unparalleled, representing the most potential and bias for harmonic content with structures that exhibit significant higher order overtones. Instead of a tightly focused fundamental and low order harmonics, the tenor will bias its spectral power towards stronger overtone content. Ideally, the tenor's voice part should be arranged in such a way that they add harmonic color more than functioning as a fundamental frequency in chords.

If we re-examine our bass and tenor now with the combination of centroid clusters and spectral density range, we can understand their potential for harmonic variation. Similar to the tenor as discussed, the bass and baritone feature points with centroids at the harmonic intervals of their given notes, indicating a similar quality of overtone presence. There are two bass notes that present a very valuable perspective on effective high-pass filtering of sounds, meaning harmonic structures with diminished fundamentals but several very powerful overtones. This presents opportunities for the bass to shift from a role of providing foundational frequencies to adopting a role similar to a tenor, with spectrally-dense harmonic content. There are many more exaggerated points for the bass that demonstrate the harmonic potential of the voice in combination with a consistent power at fundamental frequencies. Alternatively, the baritone is more reliable in establishing a consistent low-middle frequency layer. There are also clusters at intervals of the given notes, again indicating rich harmonic content; however, they are fewer and do not extend as far. The domain and range of the baritone supplements comfortably between the bass and tenor.

The mezzo presents the most exciting potential, with centroid clusters and harmonic multiples dotted across a broad range of spectral densities. It presents similar features to the bass and baritone, with conditions that support a powerful fundamental frequency, with room for moderate overtone energy. The mezzo also displays a few points that exhibit significantly powerful overtones that outperform the fundamental, similar to the bass again. With the previous understanding of the tenor, we see the mezzo demonstrate comparable conditions of very spectrally-dense material; unlike the tenor, however, the mezzo maintains predictability as points still cluster together at these higher values. Special to the mezzo voice is also its domain of centroids, with the widest range and highest frequency, establishing it as versatile in providing lower frequency information like the bass, sitting within the ensemble along with the baritone, and introducing high frequency content through numerous strong overtones with the tenor. The mezzo vocalist possesses unparalleled versatility.

Discussing range, biases, and extremities brings harmonic role or responsibility into the conversation. The plot provides a quantitative parallel to understanding the rigor behind distributing certain harmonies to specific members. A tenor cannot provide the focus to low frequencies like the bass, nor could the bass provide the rich harmonic content of the tenor. A third dimension of power might allow an increased level of nuance to understand any relationships between volume or support. For now, this plot illustrates the potential for each member of the ensemble in the context of the others.

VI. DISCUSSION

i. *Evaluating Success*

To understand whether the experiment proceeded successfully, we must review the objectives: understanding physical characteristics of vowels in technically-trained singers; comparing vowel qualities based on vowel height and mode of vocal synthesis; comparing vowel qualities between singers; and applying findings to our musical understanding of collegiate a cappella singers.

Holistically, the project completed these tasks. Recording sessions successfully tracked each vocalist without causing vocal fatigue, and provided the opportunity to thoroughly collect sample points. One shortfall was the inability to complete testing with an alto and a soprano, as explained in the preface. This meant we were unsuccessful in specifically understanding vowel characteristics as it pertained across all voice parts.

Due to the individual nature of singing technique, there is no way to quantify how accurately each vocalist sang each vowel. Accuracy, in this sense, is how close the singer was to the brightness placement and synthesis mode that was defined. Judging vocal qualities by ear and by observing the singers' mouths were the primary methods of ascertaining whether or not the singer was properly achieving the desired placement. Additionally, since there is no "expected" harmonic value to compare our experimental data to, measuring accuracy relative to a mathematical fit is not possible.

The analysis of uncertain vowel techniques relied on one crucial assumption – intonation. We expected vocalists to deliver 75 different combinations all on the same pitch for 5 different pitches. Maintaining this constant and successfully identifying deviations from this constant during recording proved very difficult. Despite using a live digital tuner and retaking samples to ensure pitch accuracy, our data and visualization demonstrate that it was not kept consistent. The main focus has been harmonic structure – a property generally unaffected by a difference of several Hertz. However, this greatly affects how we identified novel cases. Especially in the tenor, the most uncertain placements did not only differ in their structure, but also greatly in their pitch. This obstructs successfully identifying which voices demonstrated the greatest variability in harmonic content.

Regarding precision, each voice part had just one singer to demonstrate the vowel characteristics. Precision, in this experiment, is how specifically we can draw conclusions from the data. It is not possible to calculate a standard error of mean values among numerous voices, since there was no more than one person per voice part; however, it does mean that when overlaying and comparing FFTs, we can see which cases of placement and synthesis mode affect the vowel on different pitches.

ii. *Sources of Error*

There were sources of systematic and probabilistic error through the experiment. Systematic error includes variations in the temperature and humidity of the Tracking Room where the singers recorded. Sound waves move faster in hotter environments, and are also slightly impacted by humidity. Additional sources include resonance caused by the shape of the room and instruments around the studio. Since vocalists were being recorded over a span of two weeks, temperature and humidity could vary, and instruments could be moved around the room. While there was no method to minimize the impact of the former variable, the latter variable regarding instrument placement was usually mitigated by moving the instruments back to the location they were in when the first session was completed. The "clap test" was useful in identifying whether there were resonances that were unique to that day or to a specific area of the room.

Probabilistic error includes noise fluctuations due to the hardware or software. This is not directly measurable since the signal flow is a combination of numerous electronic devices, each with its own unique amount of noise added to a signal. Statistically, the effects and intensity of added noise should be negligible compared to the intensity of the frequencies that are sent through the signal flow. Measuring the precision of the studio's equipment and wiring is not feasible, therefore creating a probability distribution for the FFTs based on probabilistic error is not possible.

Other sources of error include the limited number of samples and subjects that were tested. Since singing is unique to individuals, generalizing our results is not possible without testing multiple singers per each voice part and creating a normalized average for the FFT. Additionally, the room and equipment used for this experiment are used by many other students at Northeastern University. Therefore, it is not possible to know whether the electronic or spatial integrity changed while being used by others.

iii. *Areas of Further Study*

Through the process of the developing and executing this experiment, numerous questions came up in discussion that pointed out variables that were not being tested in this study. These factors, when studied in conjunction with the factors researched in this experiment, can deepen our understanding of singers and the specificity of technique.

Power & Intensity. The effects of projection or volume were not tested in this experiment. Volumes were adjusted to appropriate levels to ensure a constant gain on the preamplifier throughout the duration of a vocalist's session. Thus, there is no clear indication of the exact effect of varying volumes. It is expected that singing louder increases the partials as shown in the FFT, but there may be extraneous harmonics generated that are separate from the case of normal volume.

This factor depends as well on the singers' ability to sing louder or softer with proper technique, a task easier said than done. For example, singing in one's belt generally has less volume control than singing in one's modal voice, because belting requires a large amount of breath support in order to produce the note. But falsetto, which also uses a lot of breath, is softer yet can range in strength based on the individual. Testing this would require a rigorous definition for volume, including whether it is discussed while considering decibel-loudness or traditional pianissimo-to-fortissimo notation, and would require highly-trained vocalists who can properly differentiate and produce notes at varying loudness across their range.

Gender. In the approach for the study, there was much consideration for how vocalists conventionally fit into voice parts and how some of them break these conventions. Specifically, the division of traditionally-female voice parts and traditionally-male voice parts glosses over many individuals who sing outside their traditionally-assigned register. Further explorations would benefit from voice part comparisons between people of different genders, i.e. comparing a male tenor and a female tenor, or a countertenor and an alto/mezzo/soprano. Additionally, studying transgender vocalists in their range of preference would deepen our understanding of the overlap between gender and vocal range.

Age. As noted, the singers in this study were college students. Another area of study would be contemporary popular music singers who are older or younger, as this would allow for analysis of how voices evolve over time. One hypothesis is that the power spectrum for each vowel evolves over time; as a vocalist ages, changing aspects of the vocal folds and tract may boost or cut certain peaks from the frequency domain visualization. As a result, mixed-age ensembles may have more options when choosing placements, and could achieve a greater number of sonic styles.

Racial, Cultural, & Linguistic Background. The singers in this study were native English speakers, and have either grown up in New England or lived in the area for at least five years. Further research should look more deeply into how singers of a variety of races, who potentially speak other native languages or have other regional dialects, produce vowels differently. Physical differences in language production can carry over into singing vowels, so vowels outside of American English may offer a wider gamut of harmonic color.

iv. *Postulating a Novel Notational Method: Cube Charts*

In a traditional choral setting, vowels are not typically notated separately from the lyrics. Standard choir technique usually employs tall vowel placement and a combination of chest and head voice depending on the singer's range. Contemporary singers employ a variety of placements and registers, but in the case of a cappella, many singers use neutral-to-spread placement and combinations of chest, belt, and head voices. In all situations, the most-used language to identify vowels is the International Phonetic Alphabet. This is separated into close, close-mid, open-mid, and open vowels in front, central, and back placements. While the IPA is consistent among voices and dialects because of its specificity, its complex set of symbols makes memorization and application difficult for singers in the spur of the moment.

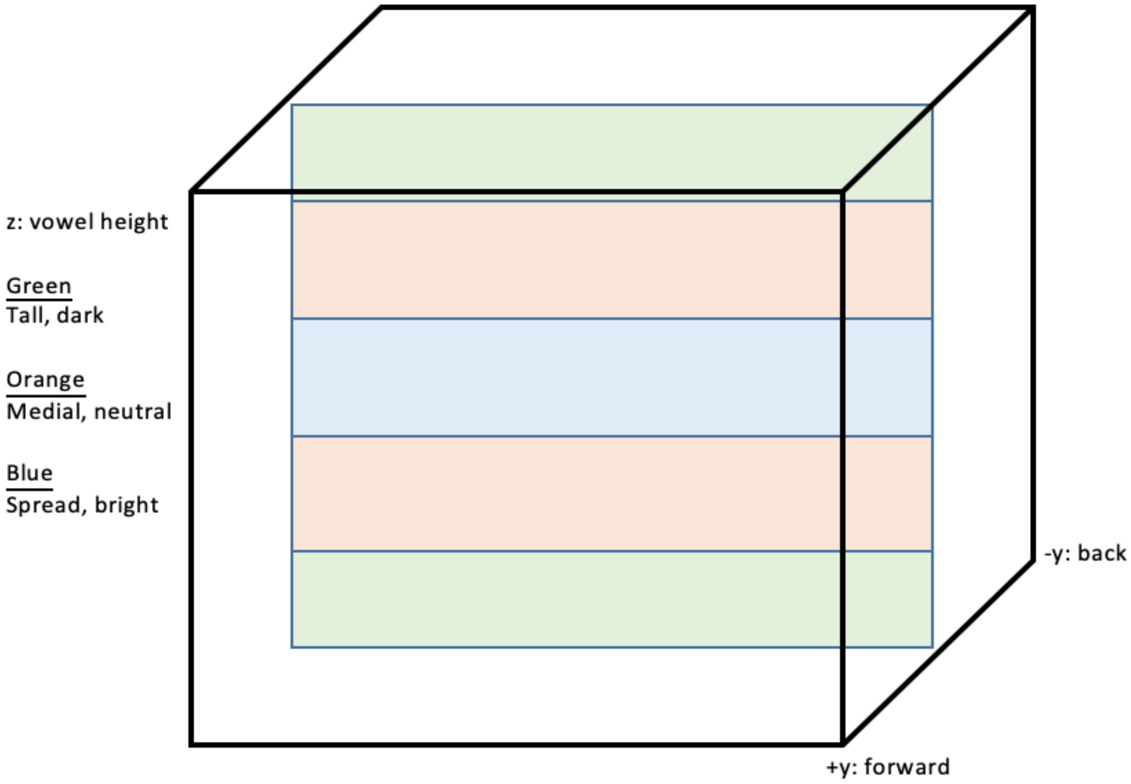
In collecting data for this experiment, we noticed patterns within vowel production that existed across phonemes, voice parts, and modes of synthesis. These patterns include how tall or spread the vowel is, how forward or back it is focused in the mouth, and how much breath the singer pushes through the pitch. These three characteristics currently are described independently from each other, and are only described qualitatively and verbally. There is no system that notates these qualities; thus, no way to teach young singers a system regarding placement, and no visualization to aid experienced vocalists.

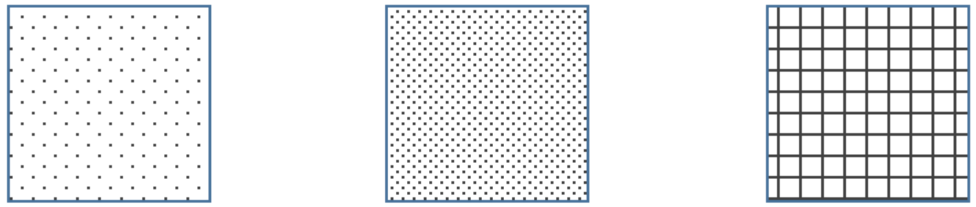
A proposed notational system, nicknamed a Cube Chart, follows in Figure A. This system can be applied to any vowel, regardless of IPA representation or lyrical context. A clear three-dimensional box represents the space in the mouth, with a patterned, shaded rectangle that lives within the box as in the Figures. The y-axis concerns how “forward” or “back” the vowel is. This form of “backness” placement was not directly measured in this experiment. The size of the rectangle in the z-direction indicates vowel height: each of the colored boxes illustrates the size of the rectangle as it corresponds to height and timbre. The type of patterned shading in the rectangle indicates how breathy the tone should be; more shaded indicates more tonal center and less breath should be used, and less shaded indicates less tonal center and more breath should be used. Using pattern-based shading is important because if one is printing on a faulty machine or drawing by hand, the shading may be difficult to create or differentiate between levels. Figure B shows a tall, centrally-placed, lightly breathy vowel. Figure C shows a spread, forward, very breathy tone. Figure D shows a neutral, back, half-breathy/half-tonal vowel. Figure E shows a spread, centrally-placed vowel sung with a full tone and no breathiness.

The reason for defining the shading of the rectangle as “breathiness” instead of mode of vocal synthesis is because the latter strongly depends on the individual’s range and cannot necessarily be regulated at all points in music. “Breathiness” is a byproduct of how much tonal center is used in conjunction with breath expulsion, and can be more easily controlled by singers across register or voice part. It also is a more practical instruction for singers. Breathiness has a greater effect on timbre than synthesis mode because an increased amount of expelled breath creates a greater amount of mid-to-high frequency content. In an arranging lens, this higher-frequency content created by the air translates to a sound that feels light and soft; a tone that is full, or with minimal breathiness, will sound heavy in comparison.

One advantage to Cube Charts is its scalability in usage. For novice singers who are grasping the concept of tall, neutral, and spread vowels, the inner rectangle can be separated from the cubic representation in a so-called “Square Chart." Viewed alone, the vertical size of the rectangle would indicate the height of the vowel. When the singer progresses to the intermediate step of learning forward, central, and back placement, the introduction of the cube illustrates the new concept while incorporating their foundational understanding of vowel height based on the rectangle. Finally, as a singer advances their technical skill, introducing the shading system adds the element of breathiness that can be modified in conjunction with the first two parameters. Another advantage is its applicability to any vocal music scenario. Vowels in choral, a cappella, and other settings can be visualized regardless of any "norms" in that style of music.

Figure A: Outline of Proposed Notational System: Cube Charts




Level 0: No tone, all breath
Level 1: Minimal tone, mostly breath
Level 2: Some tone, majority breath

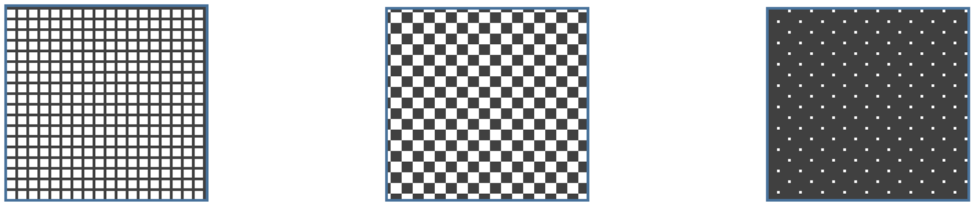

Level 3: Half tone, half breath
Level 4: Majority tone, some breath
Level 5: All tone, no breath

Figure B: Tall, centrally-placed, majority tone vowel

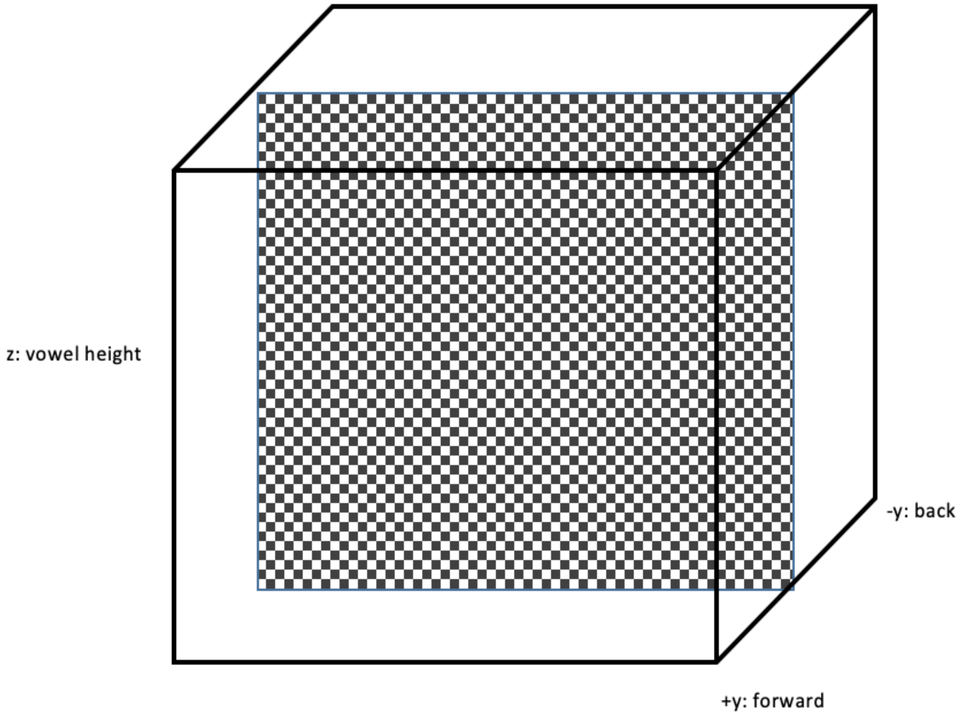


Figure C: Spread, forward, minimal tone vowel

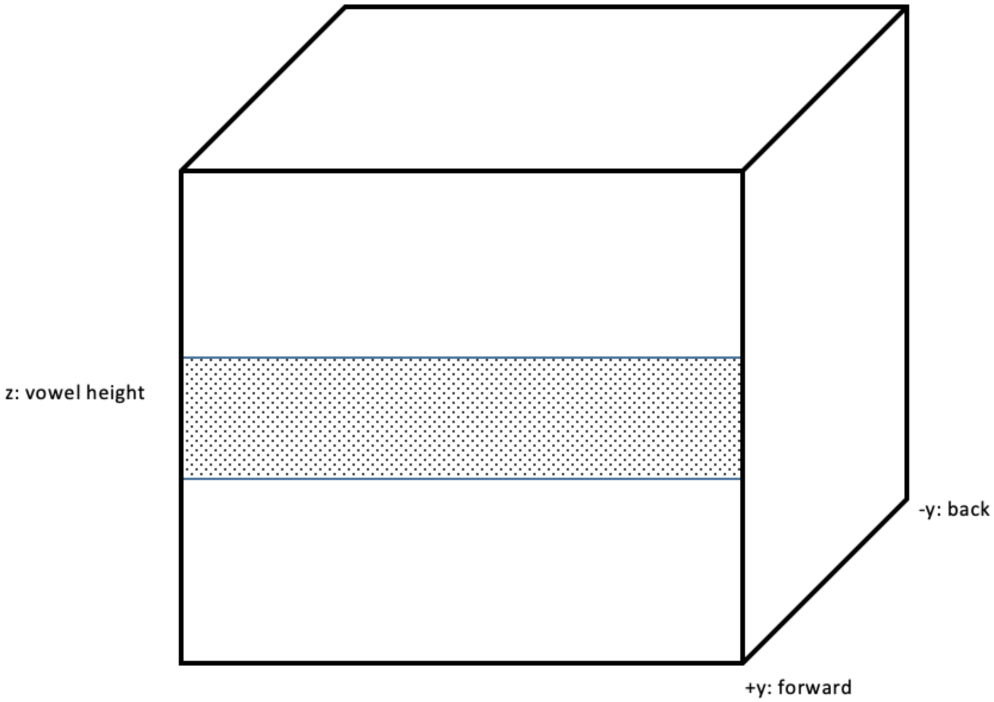


Figure D: Neutral, back, half-breathy tone vowel

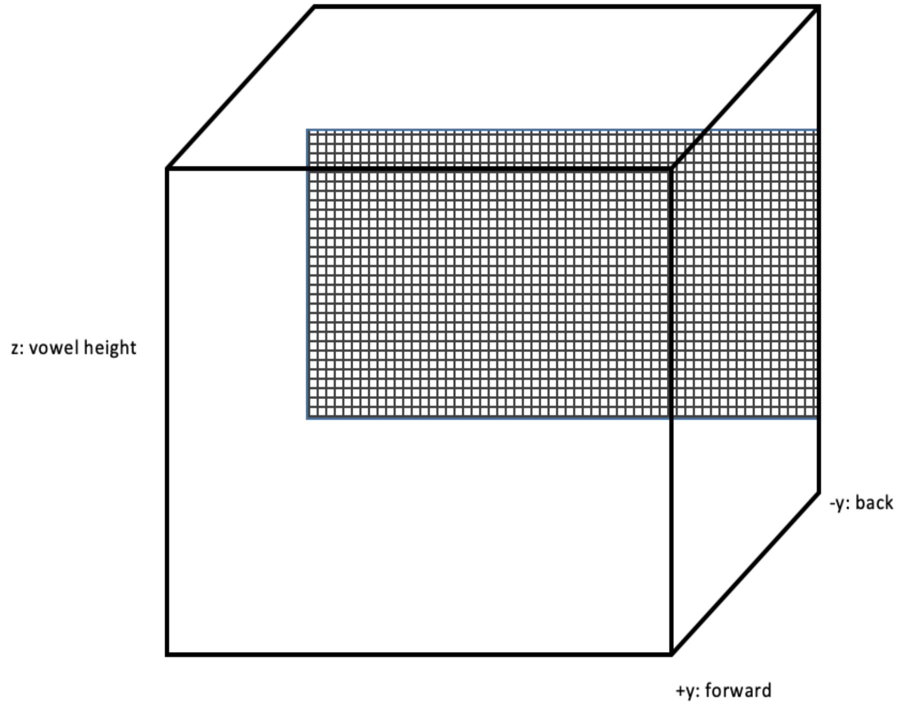
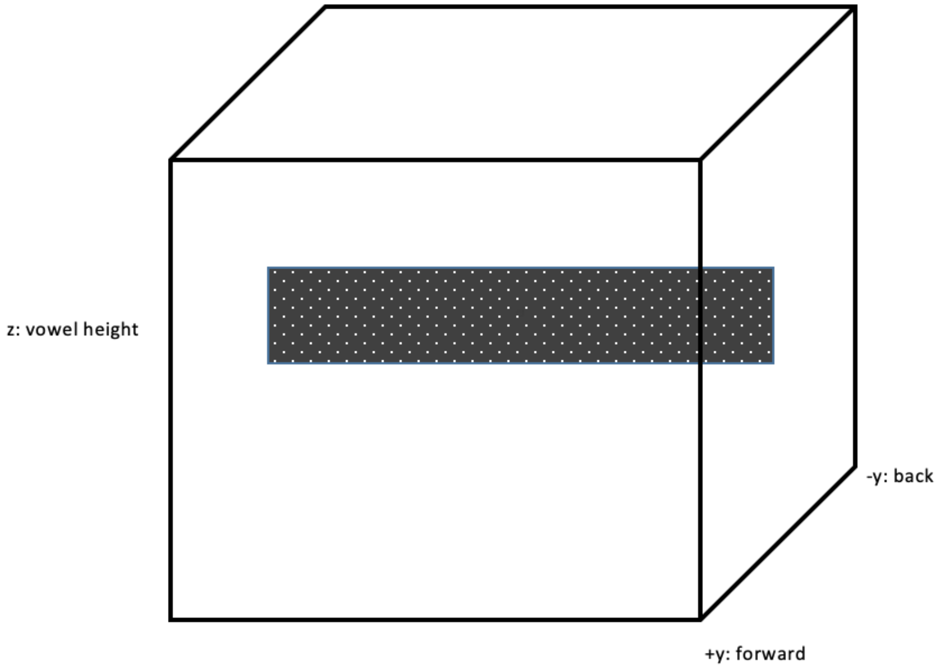


Figure E: Spread, central, full tone vowel



VII. CONCLUSIONS

This experiment examined the harmonic qualities of 4 vocalists in isolation to better understand their spectral potential and greater role within the ensemble. Much care was taken in controlling for room noise and resonance, microphone frequency response, and vocalist intonation. There are significant performance and anatomical variables related to vocalists as individuals that we did not explore with a large sample size. Thus, this experiment more accurately represents a case study with potential to expand through repetition. Our analysis of the specific 4 vocalists (Bass, Baritone, Tenor, and Mezzo) within their functional ranges suggests the following:

Musical Takeaways. For the bass, “oo” in falsetto and “ee” in modal voice are good choices to sing in their mid-to-high registers. The “ah” vowel on a high belt is inconsistent, and neither the vowel nor register are good choices to sing. When belting, neutral and dark placements are best in their middle and high registers. Arrangers should keep the bass singing low; directors can potentially utilize the resonance of bright placements in the low register as long as its power is controlled.

For the baritone, neutral placement on Middle Modal vowels like “oh” and “ah” will provide the fullest sound. In a High Falsetto, “eh” presents a variety of color depending on the placement. “Oo” and “ee” are strong choices of vowel to sing in any placement and register. The High Belt has a dramatic range in power; the Low Modal voice is the most stable, although the weakest. “Ah” sung in a bright placement is nearly always the brightest-sounding vowel no matter the register. Basses and baritones are responsible for frequency content below 200Hz, and arrangers and directors should focus on allowing the baritone to bridge the gap between the bass fundamentals and tenor harmonics.

For the tenor, the High Belt is an inefficient combination of note and vocal mode of synthesis and should be used sparingly. Low Modal voice is most consistent, albeit weakest. “Ah” is usually the fullest, spectrally-dense vowel, while “oo” is the opposite. Spectrograms suggest that there may be much similarity between bright and neutral placements. Arrangers and directors should capitalize on the tenor’s ability to add harmonic color due to the timbre of their voice.

For the mezzo, “oo” and “ee” are strong choices across their range. A high belted “Eh” can be sung loudly in dark or neutral placements, but will have the most full tone in a bright placement. Brightness placement on “oh” and “ah” correlates to the sonic brightness of the tone, and thus, the mezzo has an influential role in the overall brightness of the group’s sound. Arrangers and directors can possibly make the group sound bright or dark depending on how the mezzo sings.

Placement. Colloquially referred to as brightness placement, this technique is more accurately described by its spatial focus within the front of the mouth. When referring to sound, “bright” generally reflects spectrally-dense material with present overtones. We found that descriptions of placement are not commutable to descriptions of sound in terms of brightness or darkness because brighter placements do not always produce brighter sounds. For example, while the assumption may hold true for the tenor, it was not accurate for the bass, who produces the most spectrally-dense material in dark and neutral placements. Placement is more complex than spectral brightness in its effect on harmonic structure, and requires a more nuanced naming convention to accurately communicate the desired effect for the ensemble member, whether it be vocalist or director. The more physically- and anatomically-accurate terminology may be tall, neutral, and spread.

Vowel Sounds. A brightly-placed "Ah" tended to be the spectrally-densest vowel sound. "Oo" most often sat oppositely as the darkest or purest vowel sound, with placements varied per vocalist. The rankings of spectral density differed significantly between the extremities for each note and vocal technique for each vocalist, with little consistency for other vowels and placements. Similarly, there was little consistency for the effect placement had on the harmonic structure for each vowel – it is more accurately understood with the ensemble role individually.

Power. Our understanding of power is a secondary product of this study. The high belt is the most powerful vocal technique for all ensemble members with the greatest range in power between the extremes of the inversely-related spectral density. In all other situations, there was no definitive correlation between vowels, placement, range, technique and power.

The Ensemble. By plotting vocal samples on a spectral density v. centroid graph, the roles of ensemble members became clear as it pertains to "harmonic responsibility." By understanding the spectral content of each vocalist in context of the others, we begin to understand their functional differences. The tenor offers a spectacularly spectrally-dense voice and for that reason, should not be expected to offer the same balance and blend of the mezzo. Nor should a mezzo be prepared to offer the deeply low frequency, but richly resonant qualities of the bass. And none should be expected to compete with the humble and neutral support of the baritone filling in the mid frequencies. These are clearly more qualitative and practical understandings of the data; however, the physical implications and historical musical conventions of partitioning harmonies based on ensemble role correlate strongly with our spectral understandings of each member. Hopefully, this data quantitatively explains each member's harmonic role within the ensemble.

Notation. Our understanding of the data has also led to the postulation of Cube Charts, a new notational system that captures the various technical abilities of the vocalist. A visual system with three changeable parts representing three levels of vowel height, three levels of vowel backness, and a shading scale that defines tonal breathiness allows us to specify vowels like never before. Because we can isolate three variables, illustrate them, and correlate them to achievable musical techniques in a variety of singing styles, the notational system can be simplified for students or generalized as necessary for vocal ensembles in different genres.

The experiment is completed with modest success. In many senses there are opportunities for more ideal and controlled conditions. An anechoic chamber, privately reserved recording technologies, more rigorous checks for intonation, and a significantly larger sample size are just a few opportunities to increase the integrity of the study. Yet, even within our more functionally realistic recording conditions, we are able to understand the relationships of vowel sound, placement, vocalist range, and voice technique within the context of harmonic structure. Furthermore, we are able to quantitatively identify the differences of each member within the ensemble in agreement with musical convention. With a thorough understanding of the dynamic potential of our vocalists in isolation, we begin to understand how they may more vividly and vibrantly harmonize together.

VIII. REFERENCES

1. Vowels, Vowel Formants, and Vowel Modification. (n.d.) Retrieved from http://singwise.com/cgi-bin/main.pl?section=articles&doc=VowelsFormantsAndModifications
2. Singing Better Vowels. *(n.d.)* Retrieved from http://www.musicedted.info/SingingBetter2/Vowels.html
3. Harris, J. D., & Mehaffey, L. (2016). Pitch, dynamics, and vowel tuning in choral voice: Utilizing resonance strategies to train stylistic variance for choral singers. *The Journal of the Acoustical Society of America, 139*(4), 2067-2067. doi:10.1121/1.4950129
4. Chuang, Chiu- Kuang, and William S.- Y. Wang. "Influence of Vowel Height, Intensity, and Temporal Order on Pitch Perception." *The Journal of the Acoustical Society of America*, vol. 60, no. S1, 1976, doi:10.1121/1.2003608.
5. Hunt, A. (2019, November 01). How to Sing the Five Basic Singing Vowels. Retrieved from https://spinditty.com/learning/How-to-Sing-the-Five-Basic-Singing-Vowels
6. Interactive IPA Chart. (n.d.). Retrieved from http://www.ipachart.com/
7. "Recordinghacks." *Electro-Voice RE20 | RecordingHacks.com*, recordinghacks.com/microphones/Electro-Voice/RE20.
8. Suits, B. H. (1998). Physics of Music. Retrieved from https://pages.mtu.edu/~suits/notefreqs.html
9. "Sonic Lineup Documentation" (2019). Retrieved from https://www.sonicvisualiser.org/sonic-lineup/doc/reference/1.0.1/en/index.html

IX. ACKNOWLEDGEMENTS

The authors Arun and Christian would like to thank the following people for their direct or indirect contribution and influence on this paper:

Professor Tony De Ritis, Music Technology Capstone Director, for his guidance in selecting and researching the topic, support during the transitional phases of the project, and feedback on drafts and questions.

Professor Bryan Spring, Physics Capstone Advisor, for his guidance in the experimental methodology and analysis, suggestions for coding and visualizing data, and support during the evolution of the project's goals.

Dr. Victor Zappi, Music Technology Professor, for providing background information, articles, and guidance on the topics of vocal modeling, synthesis, and analysis.

Dr. Doug Bielmeier, Music Recording Professor, for granting access to Shillman Studio, its equipment, and providing insight on ways to test for acoustic controls.

Dr. Hubert Ho, Music Technology Professor, for inspiring our process of acoustic analysis.

Dr. Robin James, Music and Philosophy Professor, for inspiring our considerations of racial, cultural, and linguistic effects on music, perception, and research.

X. APPENDIX A: Link to Audio Samples and Data Sheets

https://drive.google.com/drive/folders/1GW_Xm0u31Lneot2aglcgGRtYkrn9tnY-?usp=sharing

APPENDIX B: Vocalist Biographies

ELLA JOHNSON

Ella is a second year from Massachusetts, studying Cell and Molecular Biology, with minors in Math and Chinese. She has been singing mezzo-soprano for *The Nor'easters* since 2018 and is the Treasurer. Her vocal inspirations are Adele, Maggie Rogers, Stevie Nicks, and Donna Missal.

Range: $E^b3 – A^\#5$

JUAN ONG

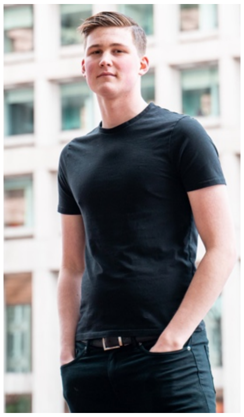
Juan is a graduate student and former undergraduate of Northeastern University originally from Hong Kong, currently pursuing a Masters in Computer Science. He was a tenor for *The Nor'easters* from 2016-2019, and was formerly the Music Director. His vocal inspirations are mainly impressive tenors old and new, such as Freddie Mercury, Shawn Mendes, Bruno Mars, and (most notably) Billy Joel.

Range: $G2 – F^\#5$

ARUN VENUGOPAL

Arun is one of the authors of this paper, and sings baritone when he is not the vocal percussionist for *The Nor'easters*. He is from Massachusetts, was in the group from 2015-2020, and has formerly been President, Assistant Music Director, arranger, and webmaster. His vocal inspirations are mainly jazz and Bollywood baritones, such as Frank Sinatra, Sammy Davis Jr., and Sukhvinder Singh.

Range: $F^\#2 – F^\#5$



BENJAMIN OCKERT

Ben is a second year student from Connecticut studying Cybersecurity and Business Administration. He has sung bass for *The Nor'easters* since 2019, is the current Business Manager, and plays on the club ultimate frisbee team. Since 2014, he has been running his own production company, 'Jamin Productions LLC, where he and his friends produce music and DJ weddings, proms, homecomings, and Sweet 16s.

Range: $B1 – G4$

APPENDIX C: Microphone Response Curve

Electro-Voice RE20

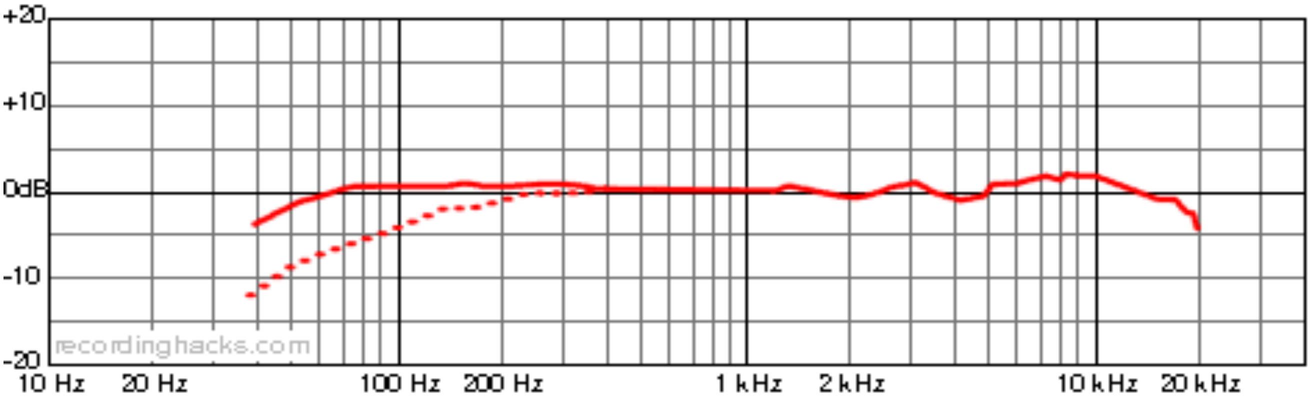


Figure courtesy of RecordingHacks.com.

APPENDIX D: Session Notes

Microphone: Electro-Voice RE20
Preamplifier: API 512C
Audio Interface: Universal Audio Apollo 16
Powerbank Rack: Furman Power Conditioner PL-Pro
Computer: MacOS Sierra 10.12.6
    Processor: 3.5 GHz 6-core 1866 MHz DDR3
Board: Raven MTX with Big Knob Studio Command System

JUAN - tenor
3pm February 24th 2020
Shillman Recording Studio
Lowest note: G2
    Note chosen: D3
Highest note: F#5
    Note chosen: B4
Middle note: F#4
Notes
    High note belts were hard
        Dark was hardest
    C#4 for chord spectrogram was in modal voice

Similarities between belt/falsetto in high tests and belt/modal in mid tests

ELLA - mezzo
7pm February 25th 2020
Shillman Recording Studio
Low note: Eb3
Bb3
High note: A#5
D#5
Middle note:
G4
Notes:
Middle modal note was in mix register because of ease of singing pitch vs in chest voice
Tried covered belt
D#5 feels like too easy to be a belt bc lower notes are harder to belt - feels more like a loud mix
Falsetto placement - made sure full raising of soft palette
Moving from bright to dark spectrograms is hard on oh
Stay on loud mix from Ah bright to dark
Less controlled belt, more controlled mix

BEN - bass
3pm March 11th 2020
Shillman Recording Studio
Low note: B1
F#2
High note: G4
C4
Middle note:
E3
Notes:
Low range is absurdly resonant
Struggled to produce belt at first
Upper placements difficult but manageable

APPENDIX E: Sonic Lineup Documentation

“The Melodic Spectrogram view is a time-frequency plot derived from a succession of short-time Fourier transforms, truncated to a limited frequency range (80 to 1500 Hz, roughly D#2 to F#6 in ASA octave terms). It may be useful for observing melodic and harmonic progressions in the music. The spectrogram uses an 8192-sample Hann window and FFT size, with 2048-sample advance between frames. If the audio file has multiple channels, they are mixed to provide the source of the spectrogram. The colour scale uses a mapping that is linear in magnitude within each column, but which scales each column down so that its maximum magnitude is set to the base-10 logarithm of its original value. This flattens out the difference in level between subsequent columns, making it easier to see pitch content but obscuring some of the volume and timing information.” (Sonic Lineup, 2019)

APPENDIX F: MATLAB Script

```
fs = 42100;
dir "[vocalist]".wav
FolderInfo = dir;
dircount = max(size(FolderInfo));
filecount = dircount - 2;

for i = 4:dircount
    clip = FolderInfo(i).name;
    header = string(extractBetween(clip,"","_"));

[x,fs] = audioread(clip);

plot(x)
xlabel('Sample Number')
ylabel('Amplitude')

snip = x(fs*0.2:fs*0.3);
g = max(size(snip));
t = (0:1/fs:(g-1)/fs);

plot(t,snip)
xlabel('Time (seconds)')
ylabel('Amplitude')
xlim([0 t(end)])

n = pow2(nextpow2(g));
y = fft(snip,n);
f = (0:n-1)*(fs/n);
power = abs(y).^2/n;

f2 = f(1:floor(n/2));
p2 = power(1:floor(n/2));

plot(f2,p2)
xlabel('Frequency')
ylabel('Power')

freq = rot90(f2,3);



if i == 4
    labels = ["freq" header(3,1);"" header(4,1);"" header(5,1);];
    data2 = [freq p2];
    FullExport = [labels; data2];
else
    labels = [header(3,1); header(4,1); header(5,1)];
    data2 = [labels; p2];
    FullExport = [FullExport data2];
end

end
writematrix(FullExport,'[vocalist]_[vowel].txt','Delimiter','tab
```

APPENDIX G: Spectral Density and Centroids for All Normalized Cases

Vocalist	Note	Vowel	Placement	Centroid (Hz)	Spectral Density
1_Bass	M_Mod	Ah	Dark	624.462891	2743.024506
1_Bass	M_Mod	Oh	Neutral	479.11377	1829.940863
1_Bass	M_Mod	Eh	Neutral	479.11377	1669.30275
1_Bass	M_Mod	Ah	Bright	479.11377	1230.384002
1_Bass	M_Mod	Eh	Dark	473.730469	1403.261292
1_Bass	M_Mod	Ah	Neutral	473.730469	1066.281547
1_Bass	M_Mod	Oh	Dark	468.347168	1529.531599
1_Bass	M_Mod	Oh	Bright	462.963867	1433.297506
1_Bass	M_Mod	Eh	Bright	462.963867	1231.126392
1_Bass	M_Mod	Oo	Neutral	322.998047	1035.768982
1_Bass	M_Mod	Ee	Neutral	322.998047	972.8809634
1_Bass	M_Mod	Oo	Bright	317.614746	802.143836
1_Bass	M_Mod	Oo	Dark	317.614746	761.7566712
1_Bass	M_Mod	Ee	Bright	312.231445	1049.195715
1_Bass	M_Mod	Ee	Dark	312.231445	1035.108781
1_Bass	M_Belt	Ah	Neutral	640.612793	3100.063013
1_Bass	M_Belt	Ah	Bright	635.229492	3162.631404
1_Bass	M_Belt	Eh	Neutral	479.11377	1782.438946
1_Bass	M_Belt	Oh	Neutral	479.11377	1538.824817
1_Bass	M_Belt	Eh	Bright	473.730469	1594.876761
1_Bass	M_Belt	Oh	Bright	473.730469	1307.207778
1_Bass	M_Belt	Eh	Dark	468.347168	1406.606401
1_Bass	M_Belt	Ah	Dark	462.963867	1410.92418
1_Bass	M_Belt	Oh	Dark	457.580566	1638.337416
1_Bass	M_Belt	Oo	Neutral	317.614746	1018.346732
1_Bass	M_Belt	Ee	Neutral	317.614746	872.3506787
1_Bass	M_Belt	Oo	Bright	317.614746	730.776197
1_Bass	M_Belt	Ee	Dark	312.231445	946.1588155
1_Bass	M_Belt	Oo	Dark	312.231445	895.300022
1_Bass	M_Belt	Ee	Bright	312.231445	671.4614108
1_Bass	Low	Oh	Bright	446.813965	2059.468311
1_Bass	Low	Eh	Neutral	355.297852	708.3565468
1_Bass	Low	Ah	Bright	279.931641	777.4623856
1_Bass	Low	Eh	Bright	274.54834	741.0187671
1_Bass	Low	Oh	Neutral	274.54834	648.8963944
1_Bass	Low	Ah	Neutral	269.165039	583.1441229
1_Bass	Low	Oo	Bright	263.781738	979.1704546
1_Bass	Low	Ee	Neutral	183.032227	671.2567186
1_Bass	Low	Oo	Neutral	177.648926	705.3788794
1_Bass	Low	Ah	Dark	177.648926	381.1662151
1_Bass	Low	Ee	Bright	172.265625	586.729577
1_Bass	Low	Oh	Dark	172.265625	440.1495165
1_Bass	Low	Oo	Dark	91.5161133	307.037902
1_Bass	Low	Eh	Dark	91.5161133	301.3115939
1_Bass	Low	Ee	Dark	86.1328125	259.8453858
1_Bass	H_Fals	Ah	Neutral	742.895508	3615.396107
1_Bass	H_Fals	Ah	Dark	742.895508	2287.509962
1_Bass	H_Fals	Oh	Bright	506.030273	1619.260429
1_Bass	H_Fals	Oh	Neutral	500.646973	1667.709883
1_Bass	H_Fals	Eh	Dark	495.263672	2016.372071
1_Bass	H_Fals	Eh	Bright	495.263672	1807.071109
1_Bass	H_Fals	Oh	Dark	489.880371	2141.186039
1_Bass	H_Fals	Eh	Neutral	489.880371	1410.719348
1_Bass	H_Fals	Ah	Bright	274.54834	1051.882267
1_Bass	H_Fals	Oo	Bright	258.398438	562.8107611
1_Bass	H_Fals	Ee	Neutral	253.015137	636.197038
1_Bass	H_Fals	Ee	Dark	253.015137	555.7844964
1_Bass	H_Fals	Ee	Bright	247.631836	610.2358026
1_Bass	H_Fals	Oo	Dark	247.631836	557.855985
1_Bass	H_Fals	Oo	Neutral	242.248535	558.6284925
1_Bass	H_Belt	Ee	Neutral	2040.271	1502.971681
1_Bass	H_Belt	Ee	Bright	2018.73779	1470.551574
1_Bass	H_Belt	Ah	Neutral	769.812012	3813.486879
1_Bass	H_Belt	Ah	Dark	753.662109	4649.124107
1_Bass	H_Belt	Oh	Dark	753.662109	2031.824222
1_Bass	H_Belt	Ah	Bright	748.278809	3024.997646
1_Bass	H_Belt	Eh	Neutral	506.030273	1811.241331
1_Bass	H_Belt	Oh	Bright	506.030273	1566.176372
1_Bass	H_Belt	Oh	Neutral	500.646973	1854.009387
1_Bass	H_Belt	Eh	Dark	500.646973	1702.027571
1_Bass	H_Belt	Eh	Bright	495.263672	1423.414627
1_Bass	H_Belt	Oo	Neutral	253.015137	840.5678374
1_Bass	H_Belt	Ee	Dark	253.015137	758.3724513
1_Bass	H_Belt	Oo	Dark	247.631836	789.9209567
1_Bass	H_Belt	Oo	Bright	247.631836	742.030533
2_Baritone	M_Mod	Ah	Bright	775.195312	4234.681613
2_Baritone	M_Mod	Ah	Neutral	775.195312	3020.816489
2_Baritone	M_Mod	Oh	Bright	597.546387	2924.245768
2_Baritone	M_Mod	Eh	Bright	522.180176	1466.318396
2_Baritone	M_Mod	Eh	Neutral	516.796875	1517.708355
2_Baritone	M_Mod	Oh	Neutral	511.413574	2326.743254
2_Baritone	M_Mod	Ah	Dark	511.413574	1572.107554
2_Baritone	M_Mod	Eh	Dark	511.413574	1547.761453
2_Baritone	M_Mod	Oh	Dark	511.413574	1389.008217
2_Baritone	M_Mod	Ee	Neutral	258.398438	652.3729555
2_Baritone	M_Mod	Oo	Dark	258.398438	626.6764471
2_Baritone	M_Mod	Ee	Bright	258.398438	609.9622332
2_Baritone	M_Mod	Oo	Neutral	258.398438	585.62987
2_Baritone	M_Mod	Ee	Dark	258.398438	558.0405179
2_Baritone	M_Mod	Oo	Bright	258.398438	557.8540064
2_Baritone	M_Belt	Ah	Bright	1038.97705	5228.992353
2_Baritone	M_Belt	Ah	Dark	780.578613	3238.890884
2_Baritone	M_Belt	Eh	Bright	780.578613	2855.413586
2_Baritone	M_Belt	Ah	Neutral	775.195312	4315.547027
2_Baritone	M_Belt	Eh	Neutral	775.195312	2256.105631
2_Baritone	M_Belt	Oh	Bright	522.180176	2349.17147
2_Baritone	M_Belt	Oh	Neutral	516.796875	1669.301535
2_Baritone	M_Belt	Eh	Dark	516.796875	1562.00405
2_Baritone	M_Belt	Oh	Dark	516.796875	1315.749994
2_Baritone	M_Belt	Oo	Dark	263.781738	906.5565978
2_Baritone	M_Belt	Ee	Bright	258.398438	793.4725305
2_Baritone	M_Belt	Oo	Neutral	258.398438	745.3225218
2_Baritone	M_Belt	Ee	Neutral	258.398438	739.442013
2_Baritone	M_Belt	Ee	Dark	258.398438	689.0352766
2_Baritone	M_Belt	Oo	Bright	258.398438	647.2952591
2_Baritone	Low	Ah	Bright	968.994141	2707.978128
2_Baritone	Low	Eh	Bright	689.0625	1763.793118
2_Baritone	Low	Ah	Neutral	689.0625	1325.033318
2_Baritone	Low	Oh	Bright	414.51416	1602.932635
2_Baritone	Low	Eh	Neutral	414.51416	802.0134976
2_Baritone	Low	Oh	Neutral	409.130859	1493.175024
2_Baritone	Low	Oh	Dark	344.53125	686.4109587
2_Baritone	Low	Ee	Bright	274.54834	1373.762713
2_Baritone	Low	Oo	Bright	274.54834	902.9722723
2_Baritone	Low	Eh	Dark	274.54834	582.8337278
2_Baritone	Low	Oo	Neutral	269.165039	640.2381287
2_Baritone	Low	Ah	Dark	269.165039	576.4154222
2_Baritone	Low	Ee	Neutral	269.165039	550.8760666
2_Baritone	Low	Ee	Dark	139.96582	441.180671
2_Baritone	Low	Oo	Dark	134.58252	488.7991832
2_Baritone	H_Fals	Eh	Bright	1491.17432	4540.661491
2_Baritone	H_Fals	Eh	Dark	1012.06055	2620.719871
2_Baritone	H_Fals	Ah	Bright	1001.29395	6180.923678
2_Baritone	H_Fals	Oh	Bright	995.910645	3046.434672
2_Baritone	H_Fals	Ah	Dark	990.527344	3330.872704
2_Baritone	H_Fals	Ah	Neutral	985.144043	1975.662809
2_Baritone	H_Fals	Oh	Neutral	506.030273	2025.511945
2_Baritone	H_Fals	Eh	Neutral	500.646973	1868.513145
2_Baritone	H_Fals	Oh	Dark	500.646973	1712.699055
2_Baritone	H_Fals	Ee	Neutral	500.646973	1020.494234
2_Baritone	H_Fals	Ee	Dark	500.646973	991.8865902
2_Baritone	H_Fals	Oo	Dark	500.646973	978.0163287
2_Baritone	H_Fals	Oo	Bright	495.263672	1188.182296
2_Baritone	H_Fals	Oo	Neutral	495.263672	1049.283542
2_Baritone	H_Fals	Ee	Bright	489.880371	987.0453063
2_Baritone	H_Belt	Eh	Dark	1491.17432	3732.845128
2_Baritone	H_Belt	Eh	Neutral	995.910645	3431.744833
2_Baritone	H_Belt	Ah	Dark	995.910645	2265.878306
2_Baritone	H_Belt	Ah	Neutral	990.527344	3493.119676
2_Baritone	H_Belt	Oh	Bright	990.527344	2918.713917
2_Baritone	H_Belt	Oh	Neutral	990.527344	2686.213602
2_Baritone	H_Belt	Ah	Bright	990.527344	2180.210989
2_Baritone	H_Belt	Eh	Bright	985.144043	3750.374212
2_Baritone	H_Belt	Oh	Dark	985.144043	2639.410156
2_Baritone	H_Belt	Oo	Neutral	500.646973	1493.279844
2_Baritone	H_Belt	Oo	Dark	500.646973	1234.693426
2_Baritone	H_Belt	Oo	Bright	500.646973	1097.685422

2_Baritone	H_Belt	Ee	Dark	495.263672	1497.92617
2_Baritone	H_Belt	Ee	Bright	495.263672	1186.28901
2_Baritone	H_Belt	Ee	Neutral	489.880371	1160.699002
3_Tenor	M_Mod	Oo	Neutral	355.297852	1244.921645
3_Tenor	M_Mod	Ee	Neutral	355.297852	1079.38091
3_Tenor	M_Mod	Ee	Dark	355.297852	1044.353834
3_Tenor	M_Mod	Oo	Dark	355.297852	905.8616842
3_Tenor	M_Mod	Oo	Bright	366.064453	939.9374351
3_Tenor	M_Mod	Oh	Dark	705.212402	3076.60939
3_Tenor	M_Mod	Oh	Neutral	705.212402	2444.309828
3_Tenor	M_Mod	Eh	Neutral	705.212402	2237.731996
3_Tenor	M_Mod	Eh	Dark	705.212402	1966.95773
3_Tenor	M_Mod	Ah	Dark	715.979004	2465.53066
3_Tenor	M_Mod	Oh	Bright	721.362305	2105.928374
3_Tenor	M_Mod	Eh	Bright	721.362305	1949.39393
3_Tenor	M_Mod	Ee	Bright	732.128906	1529.119555
3_Tenor	M_Mod	Ah	Neutral	1405.0415	3932.807064
3_Tenor	M_Mod	Ah	Bright	1437.34131	5068.125731
3_Tenor	M_Belt	Oo	Neutral	371.447754	1229.665557
3_Tenor	M_Belt	Oo	Dark	371.447754	1015.359988
3_Tenor	M_Belt	Oo	Bright	371.447754	1000.800912
3_Tenor	M_Belt	Oh	Neutral	732.128906	2225.305644
3_Tenor	M_Belt	Eh	Neutral	737.512207	2515.523044
3_Tenor	M_Belt	Eh	Dark	737.512207	2437.347474
3_Tenor	M_Belt	Oh	Bright	737.512207	2313.609572
3_Tenor	M_Belt	Oh	Dark	737.512207	2193.258288
3_Tenor	M_Belt	Ah	Dark	742.895508	3277.61162
3_Tenor	M_Belt	Eh	Bright	748.278809	2516.038428
3_Tenor	M_Belt	Ah	Bright	1125.10986	3887.877611
3_Tenor	M_Belt	Ah	Neutral	1469.64111	4515.464406
3_Tenor	M_Belt	Ee	Neutral	2207.15332	1892.088322
3_Tenor	M_Belt	Ee	Bright	2562.45117	2599.535129
3_Tenor	M_Belt	Ee	Dark	2594.75098	2657.077809
3_Tenor	Low	Ee	Bright	285.314941	1074.207206
3_Tenor	Low	Ee	Dark	285.314941	856.3216885
3_Tenor	Low	Oo	Neutral	285.314941	838.4240441
3_Tenor	Low	Oo	Dark	285.314941	755.338199
3_Tenor	Low	Oo	Bright	290.698242	1046.672676
3_Tenor	Low	Ee	Neutral	290.698242	1036.851349
3_Tenor	Low	Eh	Dark	290.698242	1016.803325
3_Tenor	Low	Oh	Dark	419.897461	1258.461972
3_Tenor	Low	Oh	Bright	425.280762	2031.257976
3_Tenor	Low	Eh	Neutral	425.280762	1278.578722
3_Tenor	Low	Oh	Neutral	430.664062	1631.227333
3_Tenor	Low	Eh	Bright	532.946777	1953.173338
3_Tenor	Low	Ah	Dark	576.013184	1538.810217
3_Tenor	Low	Ah	Neutral	581.396484	1753.322813
3_Tenor	Low	Ah	Bright	721.362305	4385.706399
3_Tenor	H_Fals	Ee	Neutral	468.347168	1075.191964
3_Tenor	H_Fals	Oo	Neutral	473.730469	1283.432296
3_Tenor	H_Fals	Ee	Dark	473.730469	1134.699628
3_Tenor	H_Fals	Ee	Bright	479.11377	1589.544298
3_Tenor	H_Fals	Oo	Dark	479.11377	1287.458423
3_Tenor	H_Fals	Oo	Bright	479.11377	1087.055974
3_Tenor	H_Fals	Ah	Neutral	942.077637	3085.114567
3_Tenor	H_Fals	Oh	Bright	947.460938	4037.39233
3_Tenor	H_Fals	Oh	Neutral	947.460938	3471.166675
3_Tenor	H_Fals	Ah	Dark	947.460938	3416.52706
3_Tenor	H_Fals	Eh	Neutral	947.460938	3015.654246
3_Tenor	H_Fals	Oh	Dark	947.460938	2491.235953
3_Tenor	H_Fals	Eh	Dark	958.227539	2572.460171
3_Tenor	H_Fals	Ah	Bright	1426.57471	7222.860809
3_Tenor	H_Fals	Eh	Bright	1431.95801	4234.257709
3_Tenor	H_Belt	Oh	Neutral	909.777832	2815.446428
3_Tenor	H_Belt	Eh	Neutral	915.161133	2927.768254
3_Tenor	H_Belt	Ah	Neutral	920.544434	3059.164887
3_Tenor	H_Belt	Ah	Bright	925.927734	3190.133941
3_Tenor	H_Belt	Eh	Bright	925.927734	3118.899191
3_Tenor	H_Belt	Oh	Bright	936.694336	4935.625428
3_Tenor	H_Belt	Ah	Dark	1388.8916	4717.328221
3_Tenor	H_Belt	Oo	Neutral	1399.6582	7036.455056
3_Tenor	H_Belt	Oo	Dark	1405.0415	9355.98111
3_Tenor	H_Belt	Oo	Bright	1415.80811	11829.05566
3_Tenor	H_Belt	Eh	Dark	1857.23877	4552.377132
3_Tenor	H_Belt	Oh	Dark	1862.62207	4801.969144
3_Tenor	H_Belt	Ee	Neutral	2670.11719	29213.98306
3_Tenor	H_Belt	Ee	Dark	2756.25	26978.0003
3_Tenor	H_Belt	Ee	Bright	2810.08301	52838.53304
4_Mezzo	M_Mod	Eh	Neutral	3235.36377	8974.055019
4_Mezzo	M_Mod	Ah	Bright	2400.95215	7370.862579
4_Mezzo	M_Mod	Eh	Bright	2400.95215	4548.074189
4_Mezzo	M_Mod	Ah	Neutral	1200.47607	4389.494208
4_Mezzo	M_Mod	Eh	Dark	812.878418	2505.118884
4_Mezzo	M_Mod	Oh	Neutral	807.495117	2914.049527
4_Mezzo	M_Mod	Ah	Dark	807.495117	2586.33865
4_Mezzo	M_Mod	Oh	Bright	802.111816	3223.874916
4_Mezzo	M_Mod	Oh	Dark	419.897461	1786.961859
4_Mezzo	M_Mod	Ee	Bright	409.130859	1384.120272
4_Mezzo	M_Mod	Ee	Neutral	409.130859	1304.790382
4_Mezzo	M_Mod	Oo	Dark	409.130859	976.3393713
4_Mezzo	M_Mod	Ee	Dark	403.747559	932.1458711
4_Mezzo	M_Mod	Oo	Neutral	403.747559	924.9868266
4_Mezzo	M_Mod	Oo	Bright	398.364258	1096.973511
4_Mezzo	M_Belt	Ee	Bright	3251.51367	7139.621354
4_Mezzo	M_Belt	Ee	Dark	3229.98047	13862.37321
4_Mezzo	M_Belt	Ee	Neutral	3219.21387	5641.31767
4_Mezzo	M_Belt	Eh	Bright	2379.41895	3885.537184
4_Mezzo	M_Belt	Ah	Bright	1997.20459	19043.11969
4_Mezzo	M_Belt	Ah	Neutral	1609.60693	8850.760028
4_Mezzo	M_Belt	Oh	Bright	1593.45703	3686.399985
4_Mezzo	M_Belt	Oh	Dark	1205.85938	4301.450987
4_Mezzo	M_Belt	Ah	Dark	1200.47607	4036.01712
4_Mezzo	M_Belt	Oo	Bright	1189.70947	2412.095468
4_Mezzo	M_Belt	Eh	Neutral	823.64502	3257.625836
4_Mezzo	M_Belt	Eh	Dark	802.111816	3208.770963
4_Mezzo	M_Belt	Oh	Neutral	796.728516	2850.20734
4_Mezzo	M_Belt	Oo	Neutral	796.728516	1651.99353
4_Mezzo	M_Belt	Oo	Dark	403.747559	1304.440901
4_Mezzo	Low	Ee	Bright	2858.53271	1302.042279
4_Mezzo	Low	Eh	Bright	2820.84961	7610.894183
4_Mezzo	Low	Ah	Bright	1162.79297	6268.942541
4_Mezzo	Low	Ah	Neutral	936.694336	4191.013983
4_Mezzo	Low	Ah	Dark	925.927734	4444.090674
4_Mezzo	Low	Eh	Neutral	694.445801	2798.594113
4_Mezzo	Low	Oh	Bright	689.0625	2361.435492
4_Mezzo	Low	Oh	Neutral	462.963867	2172.853771
4_Mezzo	Low	Oh	Dark	462.963867	1444.48585
4_Mezzo	Low	Eh	Dark	457.580566	1606.34738
4_Mezzo	Low	Ee	Neutral	236.865234	662.0043529
4_Mezzo	Low	Oo	Dark	231.481934	744.9372023
4_Mezzo	Low	Ee	Dark	231.481934	607.4498234
4_Mezzo	Low	Oo	Neutral	231.481934	601.7171711
4_Mezzo	Low	Oo	Bright	231.481934	545.5710425
4_Mezzo	H_Fals	Ah	Bright	1894.92188	5289.711503
4_Mezzo	H_Fals	Oo	Bright	1275.84229	3122.736927
4_Mezzo	H_Fals	Oo	Neutral	1265.07568	3051.625839
4_Mezzo	H_Fals	Ah	Dark	694.445801	2543.113998
4_Mezzo	H_Fals	Ee	Dark	651.379395	1382.567194
4_Mezzo	H_Fals	Oh	Bright	645.996094	2064.190429
4_Mezzo	H_Fals	Oo	Dark	645.996094	1373.741561
4_Mezzo	H_Fals	Ah	Neutral	640.612793	2305.609078
4_Mezzo	H_Fals	Eh	Bright	640.612793	2249.475788
4_Mezzo	H_Fals	Ee	Bright	640.612793	1773.820571
4_Mezzo	H_Fals	Oh	Dark	640.612793	1469.544879
4_Mezzo	H_Fals	Eh	Dark	640.612793	1378.979455
4_Mezzo	H_Fals	Oh	Neutral	635.229492	1563.373807
4_Mezzo	H_Fals	Ee	Neutral	635.229492	1523.733284
4_Mezzo	H_Fals	Eh	Neutral	629.846191	1915.300841
4_Mezzo	H_Belt	Ee	Bright	2551.68457	3403.534527
4_Mezzo	H_Belt	Eh	Bright	2535.53467	23965.51149
4_Mezzo	H_Belt	Eh	Neutral	1911.07178	7636.64129
4_Mezzo	H_Belt	Oh	Dark	1911.07178	7401.19296
4_Mezzo	H_Belt	Oo	Bright	1905.68848	8743.053478
4_Mezzo	H_Belt	Ah	Neutral	1900.30518	10970.67652
4_Mezzo	H_Belt	Eh	Dark	1894.92188	7111.165888
4_Mezzo	H_Belt	Oh	Bright	1889.53857	9674.02625
4_Mezzo	H_Belt	Ah	Bright	1873.38867	5919.338201
4_Mezzo	H_Belt	Ah	Dark	1275.84229	6033.647673
4_Mezzo	H_Belt	Oo	Dark	1270.45898	4044.069108
4_Mezzo	H_Belt	Oh	Neutral	1270.45898	3635.609509
4_Mezzo	H_Belt	Oo	Neutral	645.996094	3189.232764
4_Mezzo	H_Belt	Ee	Neutral	645.996094	2473.474616
4_Mezzo	H_Belt	Ee	Dark	645.996094	1702.640874