Quantitative Phonetic and Frequency Domain Characterization of Vocal Blend for Sung Vowels

by

Jennifer Lu

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Mechanical Engineering

at the

Massachusetts Institute of Technology

May 2020

Signature of Author ______________________________
Department of Mechanical Engineering
May 8, 2020

Certified by: ______________________________
Barbara Hughey, PhD
Senior Lecturer Thesis
Supervisor

Accepted by: ______________________________
Maria Yang
Professor of Mechanical Engineering
Undergraduate Officer

Quantitative Phonetic and Frequency Domain Characterization of Vocal Blend for Sung Vowels

by

Jennifer Lu

Submitted to the Department of Mechanical Engineering
on May 8, 2020 in Partial Fulfillment of the
Requirements for the Degree of

Bachelor of Science in Mechanical Engineering

ABSTRACT

Choral performance focuses on achieving the illusion of many voices singing as one smooth, cohesive tone, known as vocal blending. Choral pedagogy includes instructions on how to blend, but there is little information in the literature on how these qualitative instructions result in quantitative changes in the frequency domain. Voice samples of two altos with choral backgrounds were collected. They sung vowels [i], [u], [e], [o], and [ɑ] together on a D4 (293 Hz), transitioning from unblended to blended tones, and data were analyzed using frequency domain techniques.

For a musical pitch, the presence of higher harmonics amplifies dissonant intervals, which are undesirable in a blended choral sound. Choral singers are taught to "darken" their tones (reduce the amplitude of higher harmonics) and modify their vowels by dropping their jaws and rounding their lips to achieve blend.

The summed Fourier amplitudes for higher harmonics as a fraction of the sum of the fundamental frequency and second harmonic decreases by (82 ± 8)% for vowels [i] and [u], and (62 ± 12)% for vowels [e], [o], and [ɑ] after transitioning to a blended sound. It can be concluded that significant overtone suppression occurs after blending.

Vowel formants, the resonance frequencies of cavities in the vocal tract, amplify higher harmonics close in frequency to a given formant frequency. Across all vowels, the 2nd, 3rd, and 4th formant drop 1-3 harmonics between the unblended and blended time states, indicating that vowel modification reduces the intensity of dissonant intervals in higher harmonics by amplifying lower, more consonant harmonics.

Thesis Supervisor: Barbara Hughey, PhD
Title: Senior Lecturer

Acknowledgements

I would first like to thank my thesis advisor Dr. Barbara Hughey for her invaluable knowledge and endless enthusiasm while guiding me throughout the entirety of this project, from its conception as a 2.671 project to its expansion beyond music and into the world of linguistics. It is always a joy to share and discuss new learnings and discoveries with her.

I would also like to thank Dr. Norvin Richards, who helped inspire my interest in linguistics and basis for this thesis. Taking 24.917 this fall was one of my most enjoyable experiences here at MIT, and I am thankful for the new perspectives that came out of it.

Many thanks, goldfish crackers, and pancakes to Nathan Weckwerth and Nova Zhang, who sang with me to collect data and without whose help this thesis would not be possible. Special thanks and mint Oreos to Jake Kinney, linguist/pianist/trumpetist/vocalist/ arranger extraordinaire, who helped me collect data whose passion for all he does inspires me to be a better, more creative musician and arranger.

I would like to thank all my friends at MIT for their continual support and friendship throughout my career here.

Finally, I would like to thank my sister and parents, many of whose calls I ignored while working on the last few drafts of this thesis. Their unwavering love and support have carried me through the past four years at MIT.

Contents

List of Figures

List of Tables

1. Introduction

For choral performance, there is an emphasis for an individual singer to blend with the rest of the choir to contribute to a cohesive sound. In a choir, the voices of individual vocalists are desired to be indiscernible from the rest of the group. Vocalists are typically instructed to blend through timbre and vowel formation modification by dropping their jaws and rounding their lips. The ‘‘darker” and purer the tone, the easier it becomes for the vocalist to blend.

Vocalists adjust and blend to the tone of others in a choir based on qualitative assessment. They are taught to apply physical adjustments, such as dropping their jaw, and listen to the choir until they settle on a more pleasant sound and do not ‘‘stick out” from the rest of the individual singers. While there exist applicable studies comparing the timbres, different harmonic series, and blends of various instruments in an orchestral setting [1], fewer studies exist comparing these traits of individual vocalists in a choral setting.

Similarly, vocalists are taught to change their vowels based on qualitative instruction. For instance, vocalists are taught to modify the sound ‘‘ah” to something closer to an ‘‘aw”. In phonetics, the different resonance frequencies of the vocal tract for different spoken vowels have been studied closely, but the effect on sung vowels in the context of choral blending is less well studied. There exist several studies that examine these resonance frequencies in the context of sung vowels, particularly for opera singers [1] and how these frequencies interact with the frequencies of the harmonic series for any particular note. However, these studies are often conducted outside of the context of vocal blend.

The following study utilizes music theory, phonetics, and frequency domain analysis techniques to examine the phenomenon of vocal blend through multiple lenses.

2. Background

Musical instruments come in all different shapes and sizes, and as a result have distinct timbres. Timbre, also known as tone, or color, is the quality in a sound that creates variation of a musical note of the same pitch and volume. Variation in timbre is closely connected to the harmonics, or higher frequencies, of a perceived pitch. Just as there are discernible differences from instrument to instrument, no vocalist is the same. Every singer is physically built differently, from the length of their vocal cords to the shape of the inside of their mouth and the way they pronounce their words.

While this variation in tone may be desirable in establishing a distinct soloist's voice and style, choral performance focuses on achieving the illusion of many voices singing as one smooth, cohesive tone. This is known as vocal blending. Choir singers are taught to make their tones as similar as possible to those around them through physical adjustments of their vocal tract and adjusting the formation of their vowels.

2.1 Sound, Pitch, and Timbre

Sound propagates as waves with defined sinusoidal frequencies. The lowest frequency in a given single musical note is known as the fundamental frequency, which the listener can assign to a pitch, which is used to define the particular musical note [2]. Pitch is a perceived quality that is used to distinguish higher and lower frequencies as higher and lower notes on a musical scale.

The beginning and end of a musical scale are defined by its lowest pitch and the pitch with a fundamental frequency twice that of the lowest pitch, respectively. This higher pitch is known as a perfect octave. The notes in a scale are a set of pitches between the lowest note and highest note, ordered by fundamental frequency. A discrete step division of the octave space represents the musical scale.

Western music operates on the twelve-step, 13-note chromatic scale, where the octave is divided into logarithmic intervals of a ratio equal to $\sqrt[12]{2}$ (the 12th root of 2). Each step is known as a semitone. The notes are named A, A#/B♭, B, B#/C♭, C, C#/D♭, D, D#/E♭, E, F, F#/G♭, G, G#/A♭, and the scale can be translated up and down octaves. Thus, for the standard 7-octave piano, 88-key piano tuned to A = 440 Hz the frequency of any given note n on the piano can be defined as

$$f(n) = \left(\sqrt[12]{2}\right)^{n-49} \times 440\,\mathrm{Hz} \tag{1}$$

where $n = 1$ represents the lowest note on the piano, an A_0 and $n = 88$ represents the highest note, a C_8. Pitches are commonly referred to relative to their position on the piano, where A_0 represents an A in the 0th octave of a piano, and C_8 represents a C in the 8th octave of a piano. 440 Hz corresponds to the set frequency of the note A_4, which is the note to which Western orchestras '"tune", i.e. ensure that every instrument plays the same frequency for a given note [3].

By this standard, every note in western music has a certain frequency associated with it. Divergence from these specified frequency steps, where fundamental frequencies of a pitch are too high or low compared to the expected fundamental frequency of the note, create the perception of sharpness or flatness, respectively. This is commonly known as a note being out of tune.

Two notes with a discrete number of semitones between them are known as an interval. Table 1 indicates the corresponding naming conventions for musical intervals within an octave space.

Table 2.1.1 – Naming convention for semitone intervals

Number of Semitones	Interval Name	Abbreviation
0	Tonic / Perfect Unison	P1
1	Minor 2nd	m2
2	Major 2nd	M2
3	Minor 3rd	m3
4	Major 3rd	M3
5	Perfect 4th	P4
6	Augmented 4th / Diminished 5th / Tri-tone	A4 / d5 / TT
7	Perfect 5th	P5
8	Minor 6th	m6
9	Major 6th	M6
10	Minor 7th	m7
11	Major 7th	M7
12	Octave / Perfect Octave	P8

In music theory, chords are groupings of three or more notes played at the same time. When two notes in an interval help form a chord, the combination of the sound waves' sinusoidal frequencies will yield different perceived qualities of an interval. Major and perfect intervals are regarded as concordant and pleasant to the ear, while minor intervals can create musical tension and are considered dissonant by Western music standards. Combinations of notes at different intervals contribute to the perceived quality of chords. For instance, the major triad, composed of the tonic, major third, and major fifth, is considered one of the most consonant chords and is a building block of Western music [4].

Above the perceived pitch and its respective fundamental frequency, higher frequencies known as overtones, or harmonics, will be present in the waveform of a note. An overtone refers to all harmonics above the fundamental frequency, also known as the first harmonic. The reason some chords and intervals are more pleasant to the ear can be explained by the presence of a larger number of overlapping overtones [3].

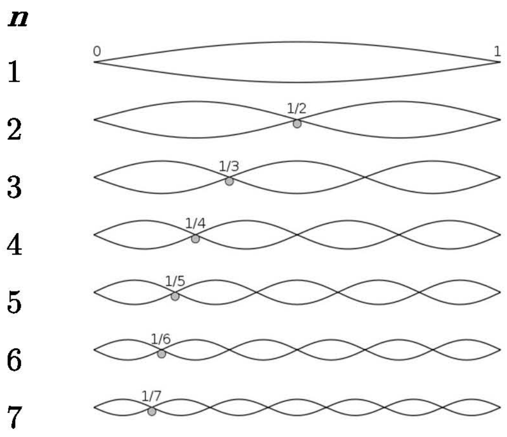


Figure 2.1.1 – Vibrational modes of an ideal string, depicting the vibration at its fundamental frequency and frequencies n times the fundamental, which correspond to the n^{th} harmonic.

Figure 2.1.1 depicts the simultaneous vibrational modes of an ideal string. The first row depicts vibration at its fundamental frequency f, reflecting its perceived pitch. The second row depicts the vibration of string in halves, which corresponds to a frequency twice that of the fundamental, $2f$. This pattern continues for all frequencies f_n [4].

The frequencies f_n can be present in the waveform of a musical note. For a note at fundamental frequency 350 Hz, there will be higher frequencies at 700 Hz, 1050 Hz, and so on, with amplitude depending on the specific musical instrument.

While these frequencies are all integer multiples of the fundamental frequency, they all correspond to different musical notes. These notes form the overtone series. The first note is known as the fundamental, or first harmonic.

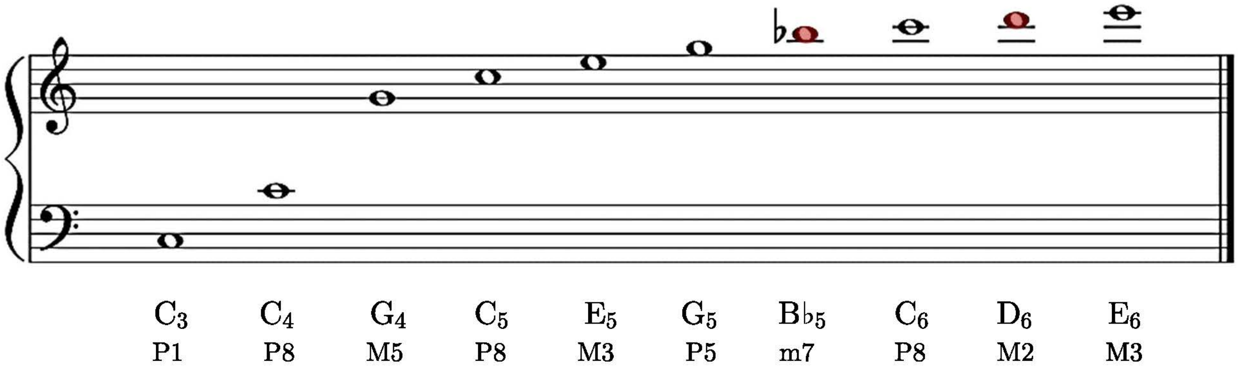


Figure 2.1.2 – the overtone series of C_3. The first note depicts a C_3 and the pitches corresponding to the frequencies of the higher harmonics are indicated by the placement of the notes on the staff and labeled below the staff. For each note in octave i, the intervals are labeled with respect to C_i. Notes shaded in red form dissonant intervals with the fundamental.

Figure 2.1.2 depicts the intervals of the different harmonics of the fundamental C3. The second harmonic is a C4, an octave above the first (also called “middle C”); the third, a G4, a perfect fifth above the second; the fourth, a C5, a perfect fourth interval above the third. For the human voice, these different notes are present when the fundamental pitch is played or sung; however, the fundamental frequency is the perceived pitch because it is also the spacing between overtones [2].

Lower overtones, like the 2nd and 3rd harmonics, correspond to harmonies pleasant to the ear like octaves and fifths. As the overtones get higher, their intervals get closer together and can create dissonance with respect to each other as well as the pitch of the note. In a chord, when singers sing different notes, one reason some chords and intervals are more pleasant to the ear can be explained by the presence of a larger number of overlapping overtones [5].

Timbre, tone, and color are often defined as the qualities in a sound that create variation in a note of the same fundamental frequency and same volume. In the same way that different intervals in a chord determine the quality of the sound of several pitches played together, the different overtones in a pitch determine the quality of the sound of a single pitch. The relative amplitude of the harmonics present in a pitch are what distinguish different mechanisms of sound production [3].

A common binary scale used to describe tones distinguishes between “dark” and “bright” tones. In bright tones, there exist large ratios of the amplitudes of higher harmonics to the fundamental frequency. In dark tones, these ratios are small. Darker tones emulate pure tones, where there is only one frequency present, such as in a tuning fork [3].

The nature of the harmonic series distinguishes a singing voice from a reed instrument from a computer-generated drone, which has only one fundamental frequency. Similarly, the different distribution and amplitudes of the overtones distinguish different tones of the same instrument, like the voice. For instance, in solo vocal performance, if a country artist and an R&B artist were to sing the same song together at the same time and volume (loudness), there would be a clear distinction of their individual voices due to their characteristic tones. An opera singer's tone is distinctive such that it can ring loudly and clearly above an orchestra or choir [1].

2.1.1 Fourier Analysis and Power Spectral Density

The frequencies of sound waves, which determine pitch, are most appropriately analyzed using Fourier analysis. A Fourier series breaks down complex waveforms into the infinite sum of sines and cosines. The following waveform function S(t) is defined as

$$S(t) = \frac{1}{2}a_0+\sum_{n=1}^{\infty} a_n cos(2\pi f_n t)+\sum_{n=1}^{\infty} b_n sin(2\pi f_n t) \tag{2}$$

where a_o is a constant, and a_n and b_n are the Fourier cosine and sine coefficients, respectively. Fourier analysis is a powerful tool that allows the breakdown of a waveform into sinusoids with different frequencies f_n, where f_1 is the lowest frequency. A Fourier transformation takes a complex waveform in the time domain and breaks it down into its multiple associated frequencies in the frequency domain [3]. Analysis of physical signals that do not have infinite duration in time uses the discrete Fourier transform (DFT), often employing the more efficient Fast Fourier Transform (FFT) algorithm. The data in Fig 2.1.3 of a D4 sung by a single singer were analyzed using the fft command in MATLAB

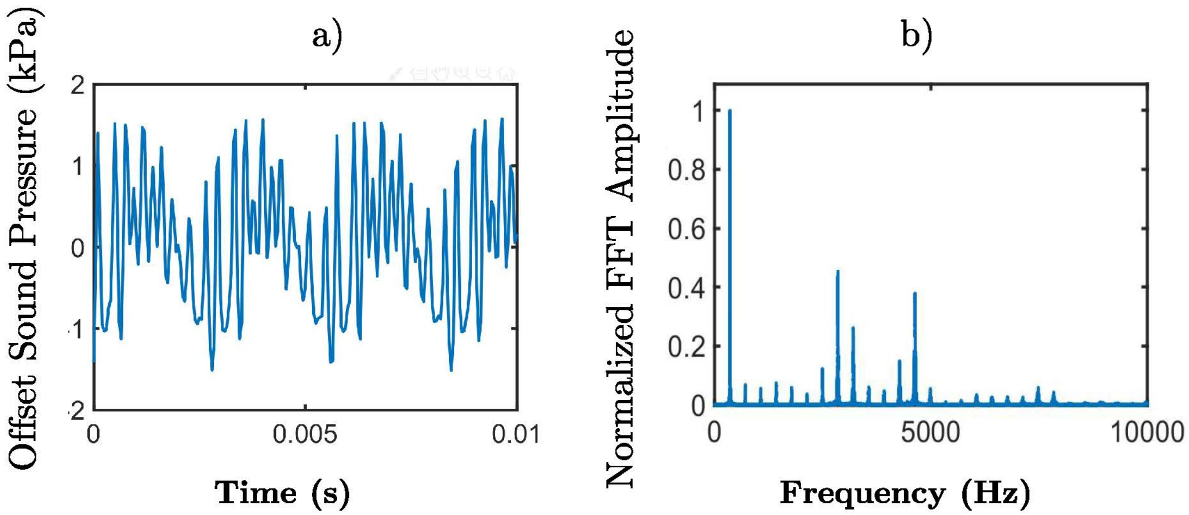


Figure 2.1.3: a) Recorded microphone signal of a sung F above the middle C (F4, 350 Hz) on an [i] vowel. The full waveform is 1 second long; a short segment of the waveform allows observation of the shape of the periodic signal in time b) The Fourier transform of this note, normalized by the FFT amplitude of the fundamental frequency. The n^{th} peak represents the normalized FFT amplitude at frequency f_n of the constituent sinusoids making up the complex waveform.

Figure 2.1.3 depicts the recorded waveform of a sung note. The lowest frequency peak is approximately 350 Hz, corresponding to an F4, the sung pitch. However, there also exist peaks at higher frequencies f_n, where n is an integer value - these peaks represent FFT amplitude at the frequencies associated with the harmonics (overtones) of the note.

The power spectral density (PSD) is a normalization of the FFT data by the frequency spacing [6], given by

$$PSD\,(f) = \frac{A^2(f)}{\Delta f} \tag{3}$$

where A is the magnitude of the signal component at frequency f, and Δf is the size of the frequency bin. As a result, the PSD amplitude gives the strength or intensity of the signal per unit frequency.

Integrating the PSD over the frequency range gives the total intensity of the signal. A useful quantity for determining the signal strength of a complex signal as a function of frequency is the ‘‘running” integral, defined as the Cumulative PSD, given by

$$Cumulative\ PSD = \int_0^f PSD\,(f) \tag{4}$$

The cumulative PSD can be normalized by the total integrated PSD and plotted over the frequency range to visualize the percentage of the intensity of the signal at a given frequency or in a selected frequency band.

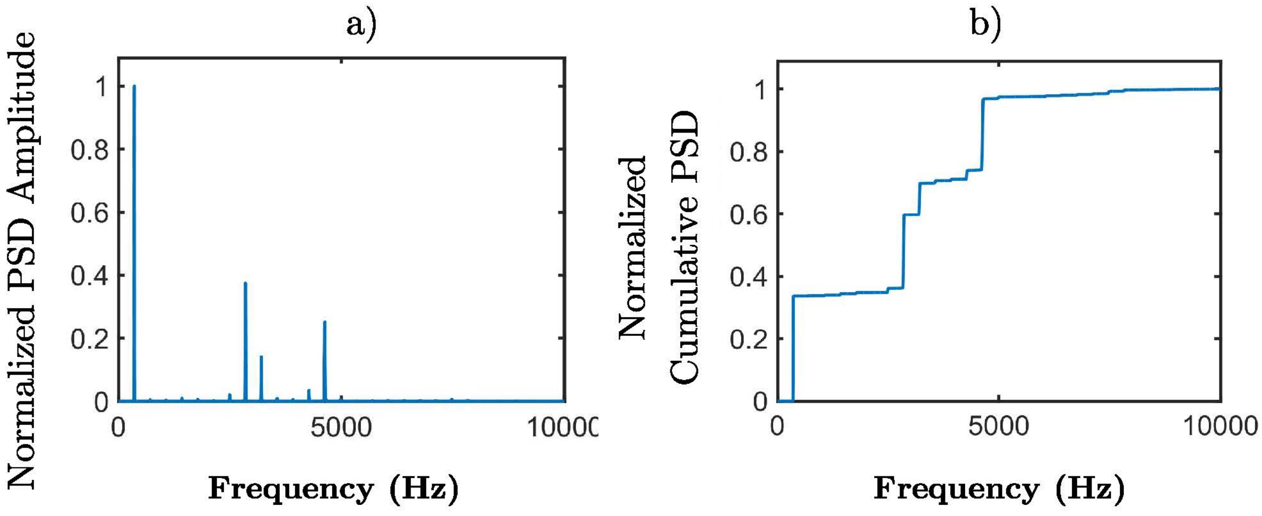


Figure 2.1.4 a) The PSD of the signal given in Fig. 2.1.3a, D4 sung on an [i] vowel. Notice that the relative peak heights are different than in Fig. 2.1.3b because the PSD is proportional to the square of the FFT amplitude of each peak. b) The normalized cumulative PSD for this signal. A sharp increase in amplitude indicates the presence of a large harmonic peak.

The normalized cumulative PSD can be used to determine that the frequency of the first large step, 350 Hz, contributes about 35% of the power of the signal. Frequencies around 3000 and 5000 Hz, corresponding to the other large increases in the cumulative PSD, are each responsible for around 20% of the power of the signal.

2.2 The Vocal Tract and Vowel Formants

Sound is produced when speaking or singing by the vibration of vocal cords within the larynx. Sound is amplified through the resonance chambers of the mouth, head, and nose, shaded in Figure 2.2.1 [7]. Manipulating the soft palate, the lips, the tongue, and the jaw opens and closes air passages to the different resonance chambers, changing the tone color, or timbre, of the pitch.

Singers naturally have different timbres based on the shape and length of their larynx, the housing of the vocal cords. A male bass will typically have a different timbre than a female soprano, because the male's vocal tract will likely be significantly longer. Thus, sound will resonate differently within their vocal cavity. Singers can further manipulate the quality of their sound by shaping their mouths differently and are often taught to drop their jaws and avoid drawing their cheeks back to achieve a darker sound.

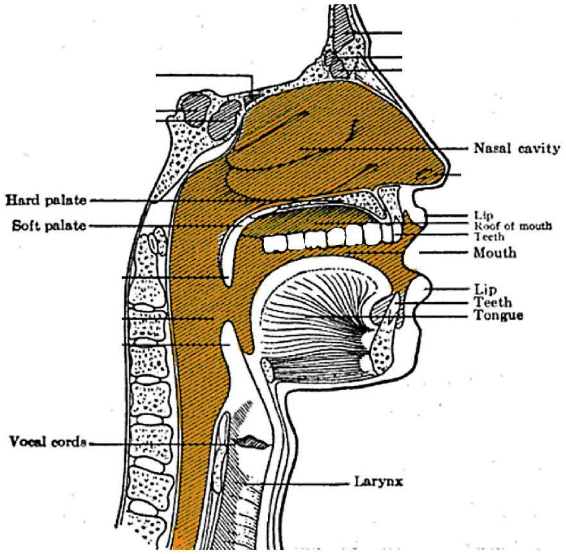


Figure 2.2.1 – cross-sectional diagram of the vocal organs, where the vocal cavity is shaded in orange. The vocal cords vibrate and sound and resonates in the mouth and nasal cavity. The soft palate can cut off sound into the nasal cavity, often making the tone harsher [7].

The soft palate controls the opening of the sound into the nasal cavity. Drawing the soft palate back often results in a nasal and harsher sound, created by higher amplitudes of the higher overtones [7].

Moving the jaw results in opening and closing of the mouth cavity. By dropping the jaw in vocal performance, the sound has more room to resonate within the vocal cavity and this creates a richer, fuller sound, with higher amplitudes in lower overtones [6]. For the same reason, proper singing technique requires that the tongue rest in the bottom of the mouth, except when articulating consonants. Raising the tongue can draw the soft palate back, cutting off the opening of the nasal cavity to sound and therefore brightening and harshening the sound [1].

When the lips are pursed for pronunciation of the syllable ‘‘oo”, the corners of the mouth are moved forward and the lips are drawn away from the teeth, creating another resonance cavity. Dropping the jaw and pursing the lips effectively create two resonance cavities and often result in a darker, purer tone [7], where the fundamental frequency is more prevalent.

2.2.1 Qualitative Identification of Vowels – Vowel Position

Changes in position of the tongue, lips, and jaw opening are used to characterize vowels. In singing, pitches are held on the vowels, which can be drawn out for longer than a consonant. Vowels are spoken with an open vocal track conducive to singing, while consonants involve constriction, closure, and contact of the tongue, lips, teeth, or other parts of vocal tract [8]. For instance, it is not physically possible to sing a pitch on a ‘‘p” consonant without a vowel. During vowel vocalization, articulators like the tongue and lips are held in one place, unlike consonants, which are characterized by articulator movement. This rest of the thesis will focus on the attributes of the five vowels studied: [i], [u], [e], [o], and [ɑ].

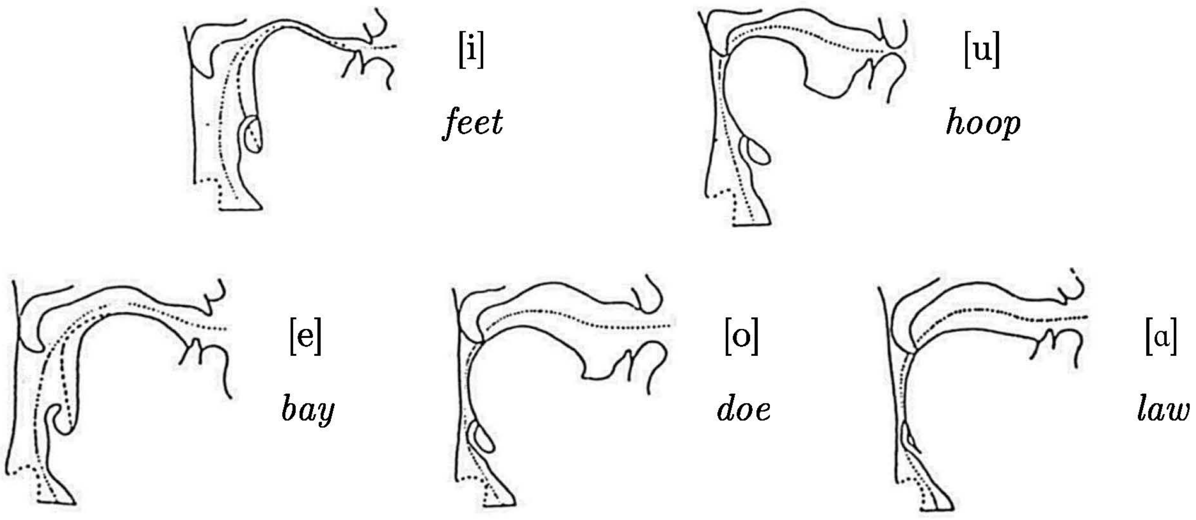


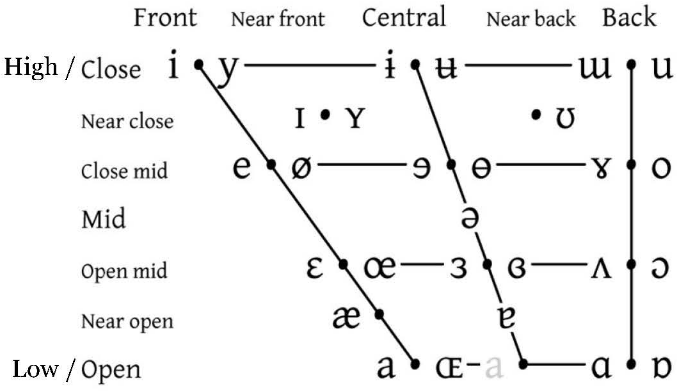
Figure 2.2.2 – shape of mouth cavity for different vowel shapes [i], [u], [e], [o], and [ɑ]. Modified figures from Sundberg 1989 [9].

In linguistics, vowels can be qualitatively characterized by the position of the tongue with respect to the roof of the mouth. The International Phonetic Association defines vowels with respect to their height, backness, and roundedness, as described below [10].

For ‘‘high” vowels, such as [i] in *feet* or [u] in *food*, the tongue is positioned close to the roof of the mouth. This decreases the open volume in the mouth cavity and closes the mouth, so high vowels are commonly referred to as ‘‘close” vowels. Conversely, ‘‘open” or ‘‘low” vowels, such as [ɑ] in *law* position the tongue low in the mouth. The International Phonetic Alphabet defines the following vowel heights: close, near-close, close-mid, mid, open-mid, near-open, and open [10].

Vowel ‘‘backness” is qualitatively characterized by position of the tongue relative to the back of the mouth. Front vowels, such as [i] (*feet*), position the tongue closer to the teeth, while back vowels, such as [u] (*food*), position the tongue closer to the back of the mouth. The International Phonetic Alphabet defines the following backnesses: front, near-front, central, near-back, and back [10].

Vowel ‘‘roundedness” is easily visible, referring to the pursing of the lips in a vowel. Unlike height and backness, there is only a binary distinction, in that vowels are considered either rounded or unrounded, with no graduations in rounding. Differences in vowel roundedness is most prominent in back vowels. Front vowels draw the lips and cheeks back, making it less natural to round out a vowel.



Vowel	*Word*
i	feet
u	hoop
e	bay
ə	dumb
o	doe
ɛ	bet
a	bat
ɑ	law

Figure 2.2.3 – IPA vowel chart and examples of English vowels. Vowels to the right of the bullet point in the IPA vowel chart indicate rounded vowels, and those to the left indicate unrounded vowels [10].

For the five vowels of interest, [o] and [u] are already considered rounded and there is no unrounded counterpart in English. For unrounded vowels [i], [e], and [ɑ], there are also no rounded counterparts in English. Instead, the English vowels [i], [e], and [ɑ] can be compared to Scandinavian rounded vowels [y], [ø], and [å] [11]. There are several resources online that help visualize and hear subtle differences between vowel sounds, including videos and audio files. A notable resource is ipachart.com [12], which hosts audio files for all the vowels depicted in Figure 2.2.3.

2.2.2 Quantitative Identification of Vowels – Vowel Formants

As the vocal cavity constricts and relaxes with the change in tongue position and mouth shape, the shape of the resonance chambers in the mouth change. As a result, different resonant frequencies correspond to different vowel shapes. These resonance frequencies are known as the vowel formants and are a quantitative measure of vowel characterization [1]. Vowel formants vary from person to person, depending on the length of their vocal tract and shape of their resonance cavities; for example, the frequency of the first (lowest frequency) formant varies from 300 – 1000 Hz depending on vowel and gender/age. Frequencies of the first 4 formants for common English vowels [i], [u], [e], [o], and [ɑ] are given in Table 2.2.1 [13].

Table 2.2.1 – Formant frequencies F1, F2, F3, and F4 for American vowels [i], [u], [e], [o], and [ɑ], spoken by men, women, and children. Adapted from Hillenbrand et al. (1995) [13].

		[i]	[u]	[e]	[o]	[ɑ]
F1	M	342	378	476	497	768
	W	437	459	536	555	936
	C	452	494	564	597	1002
F2	M	2322	997	2089	910	1333
	W	2761	1105	2530	1035	151
	C	3081	1345	2656	1137	1688
F3	M	3000	2343	2691	2459	2522
	W	3372	2735	3047	2828	2815
	C	3702	3145	3323	2987	2950
F4	M	3657	3557	3649	3384	3687
	W	4352	4092	4319	3927	4299
	C	4572	4320	4422	4167	4307

Lower vowel formants F1 and F2 are the formants which characterize the vowel. Studies have established a general trend and placement of vowel formants, which are are used to define the height and backness of a vowel. The first formant F1 has an inverse relationship with the height of the vowel. As the frequency of the first formant increases, the height of the vowel decreases. The first formant of the high vowel [i] *(feed, cheat)* lies between 350 – 450 Hz, while that of the low vowel [ɑ] (*father, law)* ranges from 750 – 1000 Hz [13].

The second formant F2 is also inversely related to the backness of the vowel. Vowels with higher second formant frequencies are closer to the front of the mouth, while vowels with lower second formant frequencies are closer to the back of the mouth. The vowel [i] has a second formant range of 2300 – 3000 Hz, while the vowel [u] has a range from 1100 – 1500 Hz [13].

F2 decreases with the rounding of a vowel. From the rounding of an [ʌ] (*cup, duck, plum)* to [ɔ] (*on, cot)*, F1 decreases from a range of 1200 – 1550 Hz to a range of 1000 – 1200 Hz [13].

Figure 2.2.4 is a graphical representation of the formant values given numerically in Table 2.2.1, showing the trend of the first formant increasing with low/openness and the second formant increasing with backness [10]. Formants vary from person to person based on physical build and manipulation of the vocal tract. The frequency ranges have been superimposed on the IPA vowel chart of Fig 2.2.3.

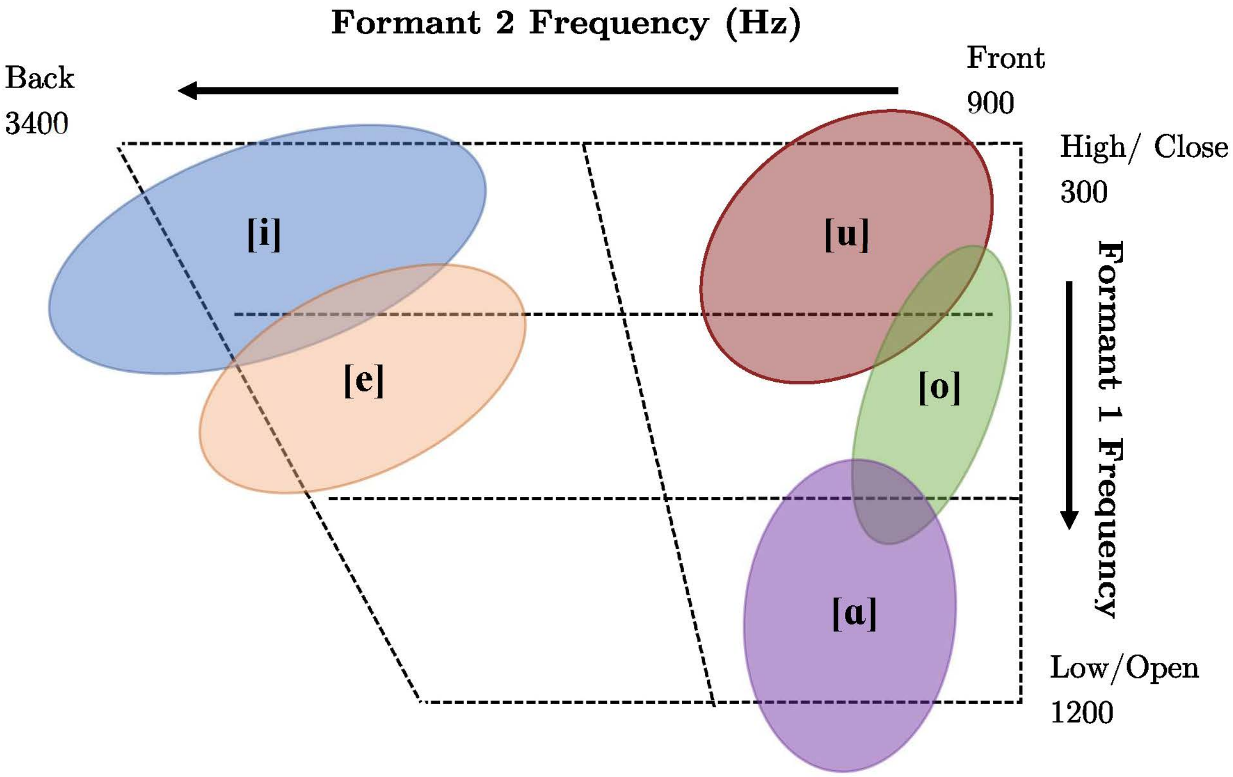


Figure 2.2.4 – Range of F1 and F2 values for men, women, and children for vowels [i], [u], [e], [o], and [ɑ]. Adapted from Hillenbrand et al. (1995) [13].

2.3 Applications to Group Singing

Differences in tone create stylistic distinctions from one vocalist to another, as mentioned in Sect. 2.1. However, in a choral or group singing context, the goal is not for hundreds of soloists to be singing at once. Rather, the priority is to create one cohesive tone [14], so that each section (soprano, alto, tenor, bass) sounds like a single voice. The focus draws away from the tone of the individual and towards the tone of the entire choir. This is the concept known as blend [15].

To achieve blend, choir singers are often instructed to blend in the following ways: pitch matching, darkening one's tone, matching vowels, and modifying vowels. These terms are defined below.

The most basic requirement for blend is pitch matching, where singers tune to the same note, or are in tune to each other relative to their note placement in a chord. For singing the same note, the fundamental frequencies must be the same. If one singer is sharp or flat, the fundamental frequencies and overtones will not align and create an unpleasant sound, due to the close proximity but mismatch of the pitch's fundamental frequencies and overtones. Tuning to the correct pitch becomes even more important when singers create chords, as different pitches have fewer overlapping overtones.

Choral pedagogy emphasizes "darkening" a singing tone in order to achieve blend with the rest of the choir [16]. As defined in Sect. 2.1, a "darker" tone corresponds to the decreased presence of higher harmonics, so that most of the signal is in the fundamental and lowest 6 harmonics. Singers are taught to drop their jaw and round their lips in order to create a darker tone [17]. By emphasizing only the fundamental and lower harmonics and avoiding the dissonant clashes of the higher overtones, choral singers can produce a uniform, less discordant, and blended sound.

Specific instructions to singers to produce a darker tone are to emulate open vowels like [ɑ] *(father, law)* back vowels like [u] *(mood, tune)*, and to purse their lips and round out their vowels. In doing so, they change the vowel they are singing. For instance, by rounding out an [i] and pursing the lips, the vowel instead becomes a [y], similar to the ü in the Chinese word *yü,* meaning fish.

The distinct resonance frequencies associated with each vowel play an important role in vocal blend. When singing, the specific frequency of a singer's natural vowel formants may not necessarily line up with the overtone series. Singers, both choral and soloistic, are taught to adjust their vocal tracts to achieve resonance between the harmonics of the notes and their vowels, either raising or lowering their formants as a result [9]. When this resonance is achieved, singers often describe a buzzing feeling within their throat [1]. When singers achieve this resonance, it amplifies the corresponding harmonic of the note. By matching vowels, singers can match similar formant frequencies and align the upper resonating harmonics.

Physical vowel modification like rounding vowels are associated with lower values of F1, F2, and F3 [18]. Another physical adjustment recommended is dropping the jaw slightly. While fully dropping the jaw close to an [ɑ] syllable is associated with an increase in F1, dropping the jaw slightly naturally draws the tongue back, primarily decreasing the value of F2 [10]. Lowering F1 and F2 will result in less perception of higher harmonics, avoiding intervals that clash and are more dissonant.

Spectrograms are useful for visualizing both overtones and formants on a sung pitch by displaying the intensity of each frequency at any given point in time [3]. The PSD amplitude of each frequency bin at each time segment is displayed on a color scale, as shown in Fig. 2.3.1. In a spectrogram plot, the horizontal axis is the time, the vertical axis is the frequency, and the color scale represents the intensity of each frequency component at a given time.

The phenomena of tone darkening through overtone suppression and formant and harmonic resonance can be observed in Figure 2.3.1, the spectrogram of a sung vowel [i] on a D4 (293 Hz). The evenly spaced bands correspond to the harmonics (overtones), and the bright yellow bands correspond to the formants. The time from 0-2 seconds represents two unblended singers, and the time from 2-4 seconds represents two blended singers.

(The rest of this page is left intentionally blank.)

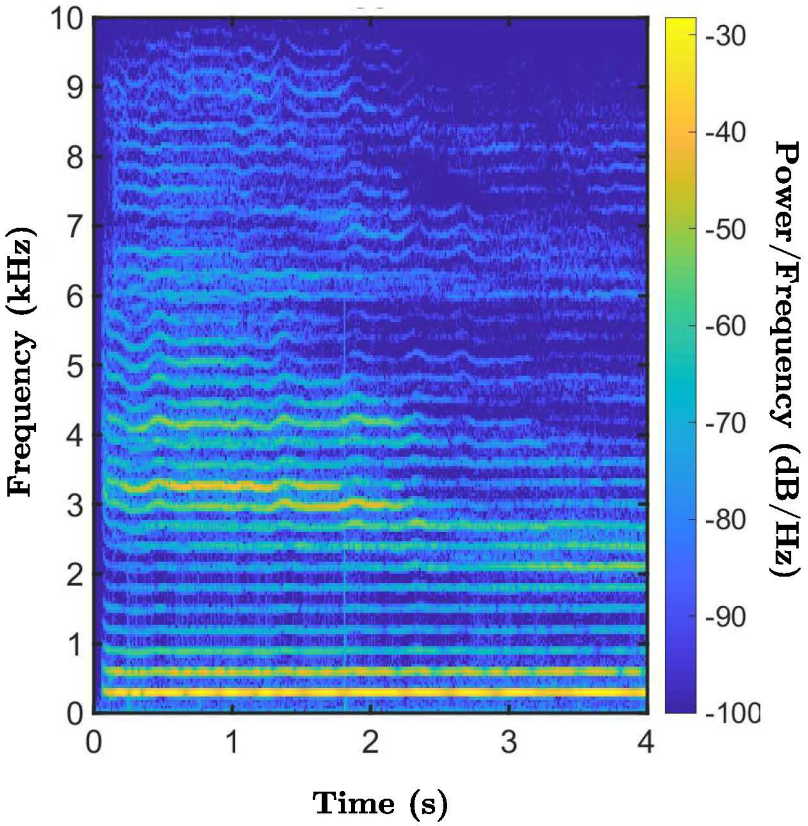


Figure 2.3.1 – Spectrogram of two altos on a sung vowel [i] with fundamental frequency 293 Hz. The time from 0-2 seconds represents two unblended singers, and the time from 2-4 seconds represents two blended singers. The change in the spectrogram after blending is discussed in the text. The evenly spaced bands correspond to the harmonics (overtones), and the bright yellow bands correspond to the formants. F1 resonates with the fundamental, as noted by the bottom bright yellow band. The yellow band between 3000 – 4000 Hz corresponds to the resonance of F2 with higher harmonics of the fundamental.

In the unblended region of the spectrogram shown in Fig. 2.3.1 (0 – 2 sec), the bottom bright yellow band corresponds to the alignment of the fundamental frequency of 293 Hz to the first formant. The yellow band between 3000-4000 Hz corresponds to the resonance of a higher harmonic and the second formant. In the blended region, the intensity of the F2 band decreases and shifts down a few harmonics, representing the lowering of the second formant. Furthermore, the intensity of all the higher harmonic bands decreases in the blended region, displaying overtone suppression. Spectrogram analysis was used in the present work to characterize vocal blend for five different sung vowels.

3. Methods

Data were collected using two altos with choral backgrounds, including the author of the thesis. In each trial, subjects were recorded as they sang various vowels together, transitioning from unblended to blended tones. In all trials, both subjects sang the same musical note, D4 (293 Hz).

The collected voice samples were analyzed using frequency domain techniques as described below.

3.1 Recording Voice Samples

Data were collected using a miniDSP UMIK-1 USB microphone. WAV samples were recorded directly into Audacity with a 44.1 kHz sample rate. The microphone was positioned 6 inches away from both test subjects' mouths to prevent signal clipping.

The singers were instructed to sing with unblended, bright tones on a D4 (293 Hz) for 2 seconds (8 beats at 240 bpm). After 2 seconds, the singers were instructed to blend their tones and hold the blended note for another 2 seconds. In line with traditional choral practices, as discussed in Sect. 2.3, the bright tone was sung by drawing the corners of the mouth back as far as possible, while the dark tone was sung by dropping the jaw and rounding the lips as much as possible.

The singers blended on the vowels [i] as in *eat*, [e] as in *ate*, [ɑ] as in *caught*, [o] as in *oat*, and [u] as in *food.* Each vowel blending trial was recorded 5 times, resulting in a total of 25 measurements.

3.2 Identifying Overtone Suppression

In Sect. 2.1 and 2.3, the role of higher harmonics on the perceived blend of a pitch is discussed. Due to the more dissonant tones in the higher harmonics of a pitch, vocalists are taught to suppress these overtones when blending in a group. In order to identify the change in these higher harmonics over time, Fourier analysis was performed at three different time points: once in the unblended time zone, the blending time zone, and blended time zone.

The FFT of each wave file was computed every 0.01 seconds and the resulting spectra were averaged over the unblended time state (0 – 1.5 seconds), mid-blend time state (1.5 – 2.5 seconds), and blended time state (2.5 – 4 seconds). Although the vocalists were instructed to begin blending at 2 seconds, the mid-blend time state contains a 0.5 second buffer on either end to account for unconscious changes in mouth shape in preparation for blending. Time-averaged data was subsequently averaged for all of the trials.

The FFT amplitude of each overtone was normalized by the FFT amplitude of the fundamental frequency to allow comparison of the relative strength of each higher harmonic, relative to the fundamental. Results for the [i] vowel are given in Figure 3.2.1. These higher harmonics were grouped into harmonics 3-6, 7-15, 16-25, and 26-40, and the total intensity of each group was defined as the sum of the normalized harmonic amplitudes in that group, as shown in Figure 3.2.2 for the [i] vowel. The motivation for these particular groupings is explained below in terms of the overtone series and expected formant frequencies.

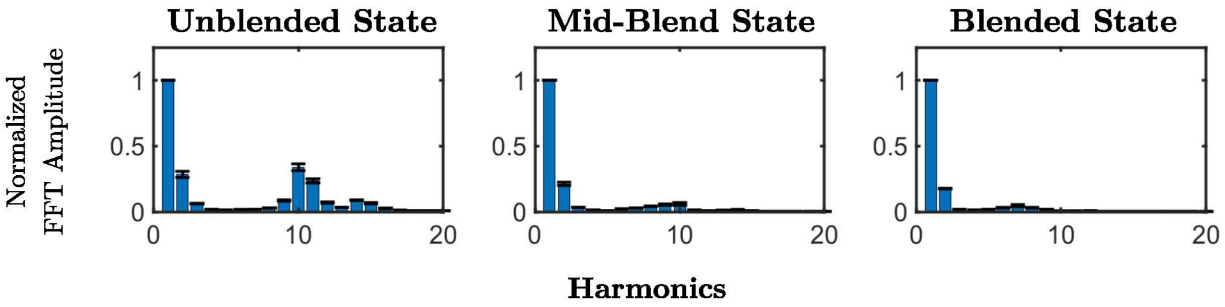


Figure 3.2.1 – Normalized FFT amplitudes for the 1st – 20th harmonic on an [i] vowel for the unblended, mid-blend, and blended states. Data were averaged over the defined time state boundaries 0 – 1.75 seconds, 1.75 – 2.25 seconds, and 2.25 – 4 second. Time-averaged data were normalized by the FFT amplitude at the fundamental frequency for that trial, and then averaged over five trials for the [i] vowel and. Error bars indicate the 95% confidence interval.

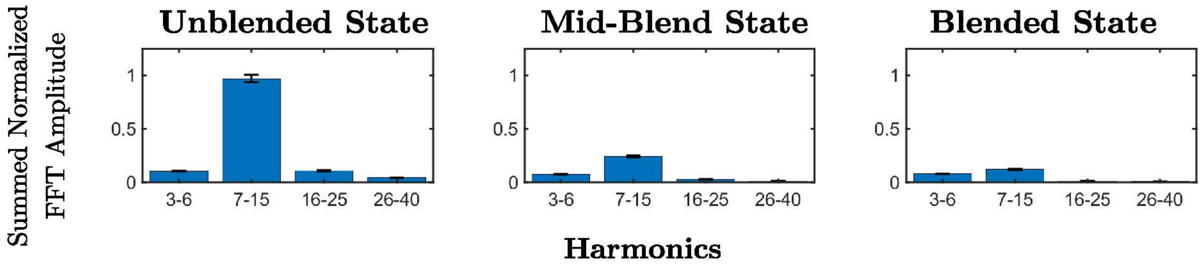


Figure 3.2.2 – The results of Figure 3.2.1, plotted as the sum of the average normalized harmonic amplitudes for harmonics 3-6, 7-15, 16-25, and 26-40. Error bars indicate the 95% confidence interval of the sum.

When examining the amplitude of higher harmonics, the 2nd harmonic is not included for two reasons. For lower vowels such as [ɑ] and [o], the first formant is closer to 600-800 Hz. As previously discussed in Sect. 2.3, vocalists will adjust their vocal tract to achieve resonance between their formants and a similar frequency harmonic. For lower vowels, the first formant amplifies the second harmonic, which is one octave above the fundamental. From Sect. 2.1, the octave represents the same musical note, so it does not create any unusual dissonances and should not have as great of an effect on the overall tone as higher harmonics.

The first higher harmonic grouping was determined based on the overtone series. The grouping of harmonics 3-6 includes all consonant overtones following the first octave jump. As shown in the musical notation in Figure 3.2.3 the 3rd overtone is a 5th interval, the 4th is two octaves above the fundamental, the 5th is a major 3rd interval, and the 6th is three octaves above the fundamental. These notes are all included in the major triad, one of the most common, consonant chord structures in western music.

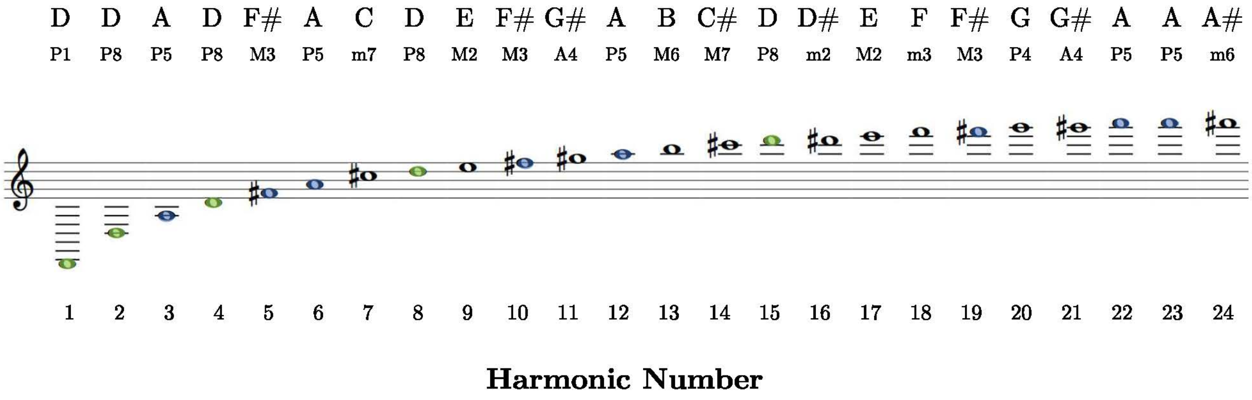

D D A D F# A C D E F# G# A B C# D D# E F F# G G# A A A#
P1 P8 P5 P8 M3 P5 m7 P8 M2 M3 A4 P5 M6 M7 P8 m2 M2 m3 M3 P4 A4 P5 P5 m6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Harmonic Number

Figure 3.2.3 – The harmonic series for the note D, from the 1st to 24th harmonic (n). The M3 and P5 intervals are highlighted in blue, and the octave (P8) intervals of the note D are highlighted in green. As the harmonics get higher, there are fewer M3 and P5 intervals. Additionally, the harmonics get closer together.

The second group of harmonics studied includes many harmonics not contained in the major triad as well as the harmonics with which the vocal formants resonate. This group begins with the 7th harmonic, which corresponds to a note that is a major 7th interval above the fundamental. It is the first overtone that is not a major 3rd, perfect 5th, or octave. The second group of harmonics is grouped through the 15th harmonic. Beyond the 15th harmonic, octave intervals of the fundamental frequency are not present. For a pitch sung on D4 (293 Hz), the 15th harmonic (4395 Hz) corresponds to the upper bound of 4th formant frequencies for the vowels [i], [u], [e], [o], and [ɑ].

The third group of harmonics 16 – 25 approximately represents the remaining harmonics in the overtone series that are at least a semitone apart. In Fig. 3.2.3, the 22nd and 23rd overtone are both written as the note A, because the difference in pitch corresponds to less than a semitone step. The final group of harmonics 26 – 40 represents remaining range of the frequencies which were analyzed from the sound sample.

The grouped amplitudes were analyzed at the same time states: unblended, blending, and blended. These results will be discussed further in the Sect. 4, Results and Discussion.

3.3 Tracking Formants

As discussed in Section 2.2 and 2.3 the vowel formants are the resonance frequencies of the different cavities in the vocal tract. When sung, these resonance frequencies tend to align with a certain harmonic. By analyzing a sound signal's power spectral density, one can determine the frequencies that have the most energy. Vowel formants correspond to these peaks in the frequency spectrum and can be determined using cumulative PSD analysis for each vowel.

Spectrogram data were averaged across trials for each time point and formant frequency band using MATLAB. The frequency bands were defined by the formant frequency ranges of each vowel, as given in Table 2.2.1. The power spectral density of each point in a given frequency band was computed for each time point, and then converted to a cumulative PSD for this formant frequency band as a function of time. This cumulative PSD was normalized by dividing by the total cumulative PSD over this frequency band, as given in Eq. 5.

$$\mathit{Normalized}\ \mathit{Cumulative}\ PSD\,(f) = \frac{\int_{f_{i,1}}^{f} PSD}{\int_{f_{i,1}}^{f_{i,2}} PSD} \tag{5}$$

where f is any given frequency, i is the formant number, and $f_{i,1}$ and $f_{i,2}$ are the lower and upper bounds of the formant frequency range for a given vowel and formant i, as determined by Table 2.2.1. The integral of the PSD over the entire formant frequency range represents the total strength of the signal in that frequency range.

The frequency at which the normalized cumulative PSD reaches 50% is known as the median strength frequency, where half of the signal strength is below this frequency, and half is above. This was therefore chosen as the formant frequency, since it represents where the strength of the signal in this frequency band is centered. This is depicted in Figure 3.3.1 for the vowel [u]; the power of the signal for all the formants exhibits a sharp increase around 50% of its normalized cumulative PSD, indicating the concentration of signal strength around the median value. This graph is labeled with the measured frequencies of formants 1 – 4.

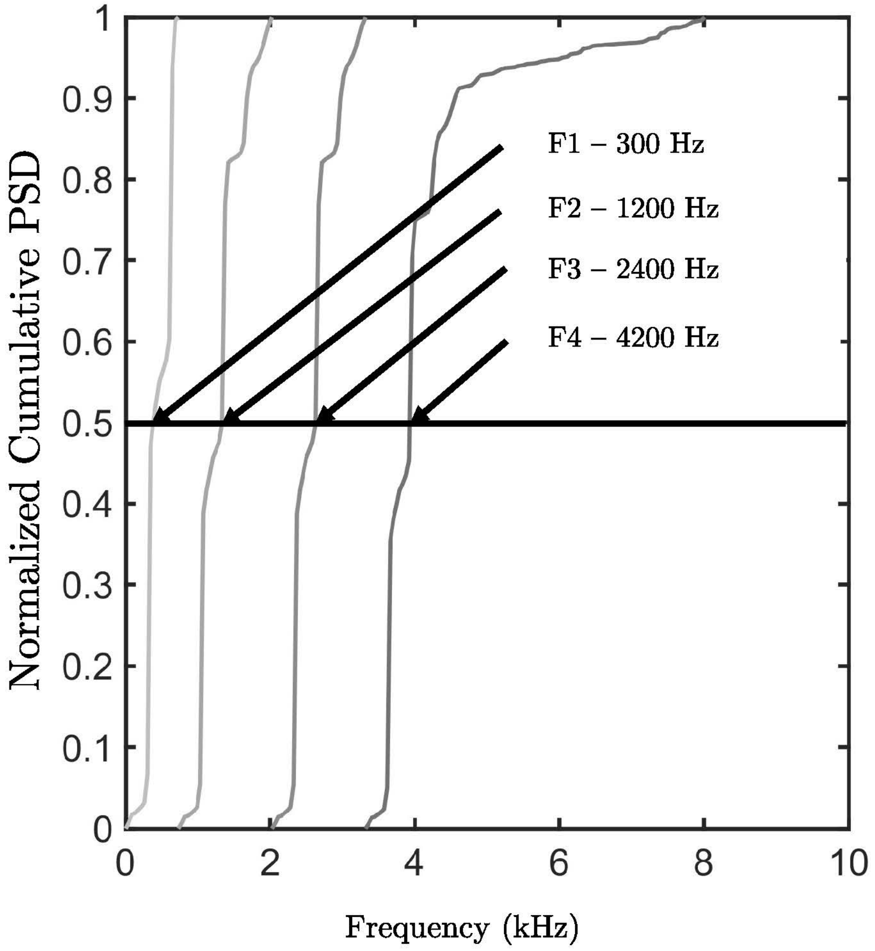


Figure 3.3.1 – Normalized cumulative PSDs for four formants in the vowel [i]. The frequency of each format is defined by where the curve intersects the horizontal line at 0.5, indicating the median strength frequency for each formant.

The formant frequencies defined by the median strength frequency as shown in Fig. 3.3.1 were found at every time point and plotted. These values were also averaged over predefined time zones for unblended, mid-blend, and blended pitches between the two vocalists, as shown in the central large plot of Fig. 3.3.2. The smaller '"wing" plots in Fig 3.3.2 give the normalized FFT amplitude data for the unblended and blended time states on a log scale with the same frequency scale as the vowel tracking plot to compare the alignment of formants to the intensity of the harmonics.

As before, the time states were defined as follows: unblended time state (0 – 1.5 seconds), mid-blend time state (1.5 – 2.5 seconds), and blended time state (2.5 – 4 seconds). Vocalists were instructed to blend at 2 seconds. The mid-blend time state accounts for 0.5 seconds of adjustment in preparation for blending for 1.5 – 2 seconds, and the time it takes to each full blend from 2 – 2.5 seconds.

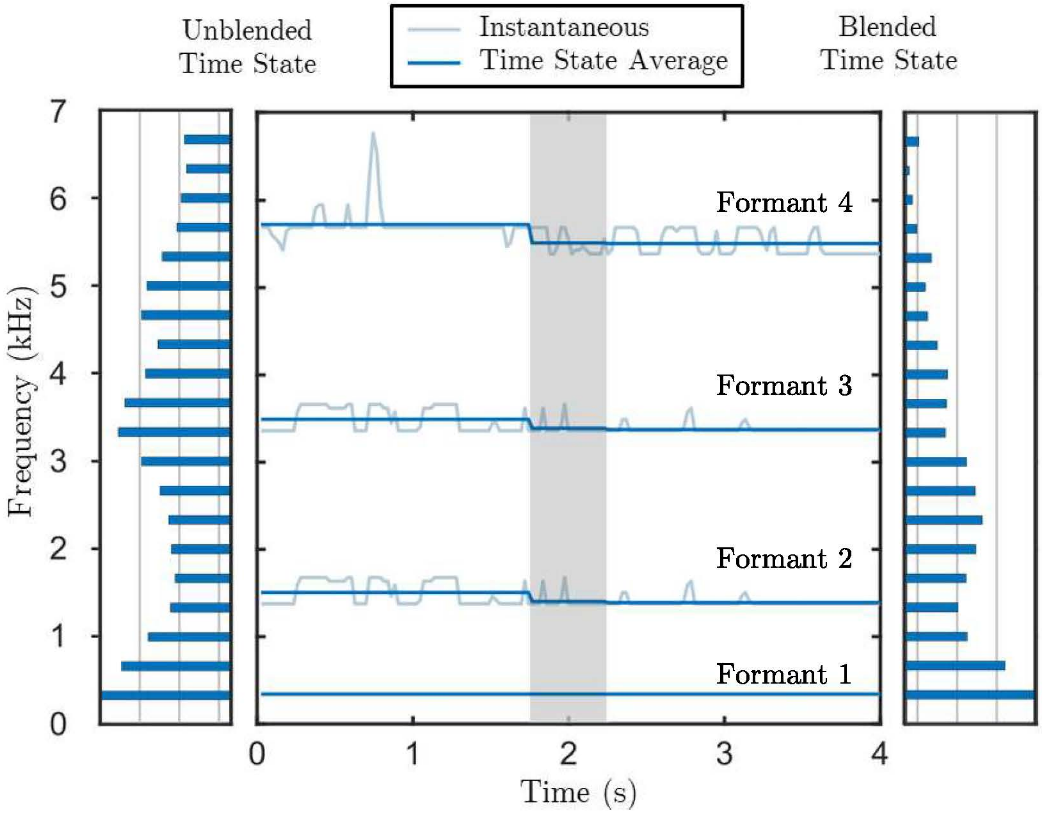


Figure 3.3.2 – Formant frequencies as a function of time for the vowel [i] for formants 1-4. The blending time zone is noted by a grey background. The unblended and blended time zones are on the left and right, as indicated. The instantaneous formant frequencies defined above are the light lines in the central plot, and the dark lines indicate the average frequency over each of the three time states. The normalized FFT amplitudes of each harmonic are plotted on a log scale in the '"wing" plots on the sides.

Visible in Fig. 3.3.2 is is an obvious decrease in the frequency of the higher formants after blending. This type of composite graph will be used in the Results & Discussion effect to discuss changes observed in formant frequency after blending in further detail.

4. Results and Discussion

The results for two altos blending on vowels [i], [u], [e], [o], and [ɑ] for five trials each on a D4 pitch (293 Hz) displayed overtone suppression and a decrease in formant frequencies. These data support existing claims that blending of a pitch is associated with fewer clashing overtones [19]. Furthermore, they also support the theory that lowering vowel formant frequencies through mouth shape adjustment amplifies lower, less dissonant harmonics [19].

Analysis of overtone suppression will be presented first, followed by the analysis of vowel adjustments and changing formants.

4.1 Overtone Suppression

As discussed in Sect. 3.2, amplitudes of each overtone for each vowel, as identified in the FFT spectrum, were averaged over the predefined unblended, mid-blend, and blended time periods and then normalized by the amplitude at the fundamental frequency. The normalized FFT amplitudes were grouped by harmonics and summed to consolidate data.

During the unblended and blended time states, the singers are assumed to have been singing with consistent tone brightness and vowel shape. During the mid-blend time state, the singers were dynamically changing their tone and vowel shape over a short time period, which resulted in a comparatively larger degree of uncertainty in overtone amplitude.

Figures 4.1.1 and 4.1.2 show the shape of the normalized amplitudes for vowels [i], [u], [e], [o], and [ɑ].

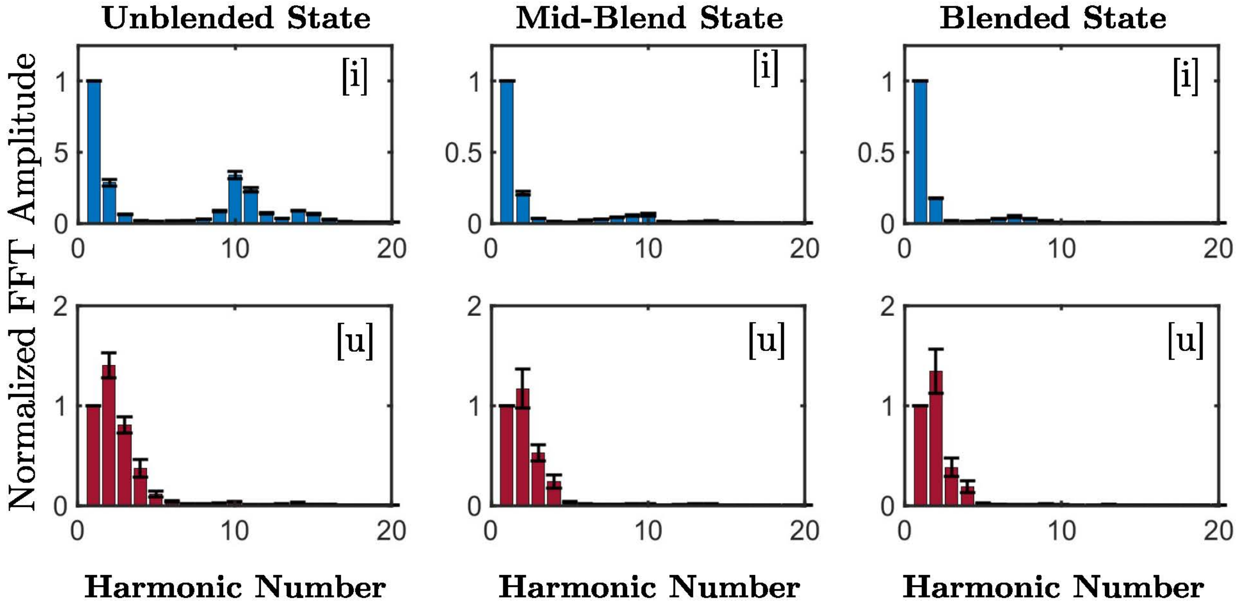


Figure 4.1.1 – Normalized FFT amplitudes of the harmonics of vowels [i] and [u] for unblended, mid-blend, and blended time states. Higher harmonics are more present in the unblended time state, and the mid-blend state looks similar to the blended state.

For each vowel, it is apparent that harmonics above the second decrease drastically between the unblended and blended time states. For unblended time states, amplitudes of harmonics above the 10th are present. For blended time states, harmonics above the 5th are comparatively smaller to the harmonics in the blended time state. The first and second harmonic dominate, with comparatively lower amplitudes of harmonics above the second. The mid-blend state closely resembles the blended state.

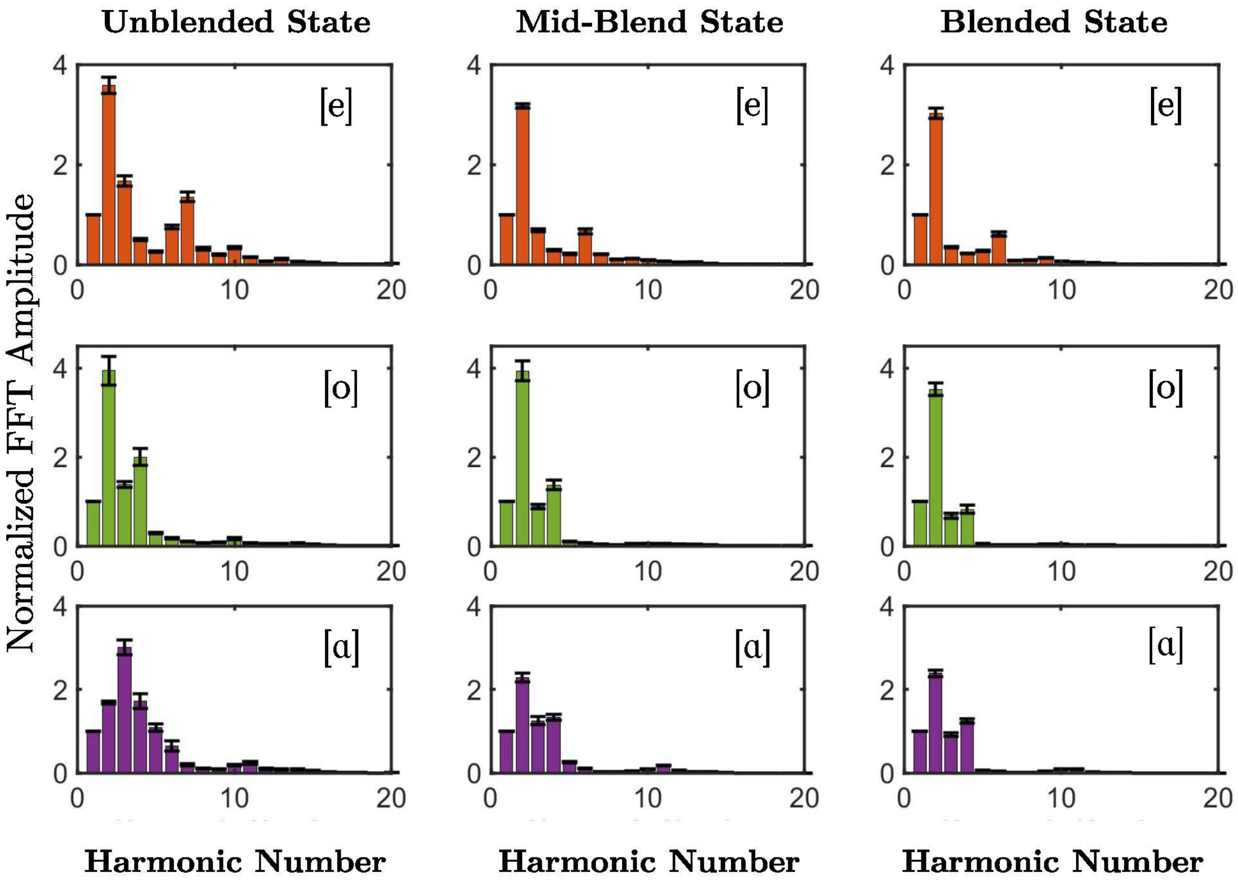


Figure 4.1.2 – Normalized FFT amplitudes of the harmonics of vowels [u], [o], and [ɑ] for unblended, mid-blend, and blended time states.

As shown in Fig. 4.1.1 and 4.1.2, all vowels [i], [u], [e], [o], and [ɑ] displayed similar trends in decreasing amplitudes of higher harmonics when comparing the unblended and blended states. The results in the mid-blend state closely resembles the blended state.

Figure 4.1.3 shows the summed amplitudes of the defined harmonic groupings, as described in Sec. 3.2. The grouped data make the trends of the ungrouped harmonic amplitude data more clear.

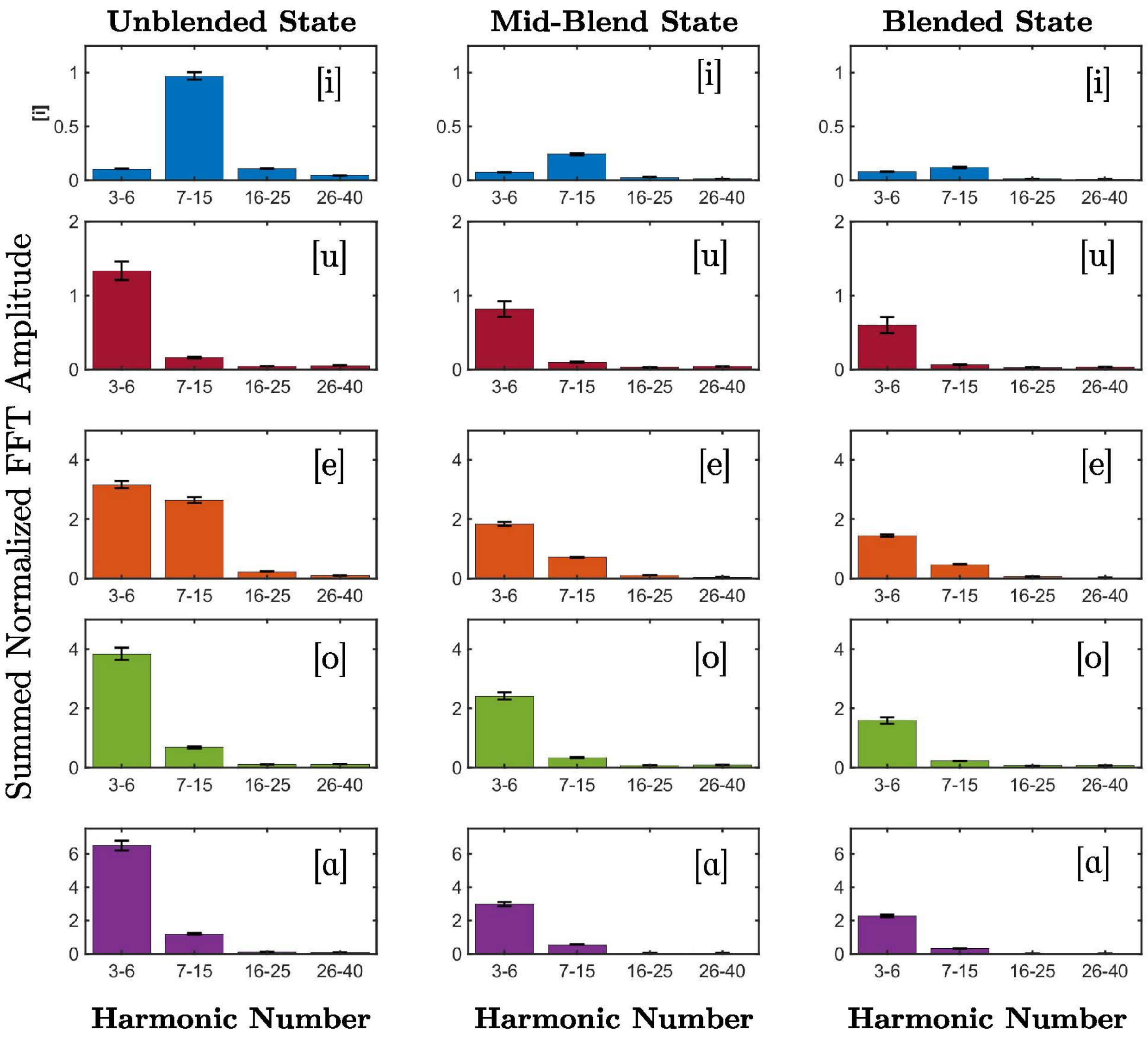


Figure 4.1.3 –Summed normalized FFT amplitudes of grouped harmonics of vowels [i], [u], [e], [o], and [ɑ] for unblended, mid-blend, and blended time states

The data in Figure 4.1.3 show similar trends to the data in Figs. 4.1.1 and 4.1.2, where the amplitude of higher harmonics is greater in the unblended region than the blended region. With the exception of [i], the mid-blend state closely resembles the blended state.

The variation in the mid-blend state can possibly be attributed to the relative differences in blending time between [i] and the rest of the vowels. The differences are likely due to the physical act of rounding, which vocalists do to blend their tones. The dynamic act of rounding the mouth varied amongst vowels. For vowel [i], it is a high vowel, where the unblended state typically involved drawing back the lips and cheeks, and blending involved bringing the lips forward to a more pursed position. For vowels [ɑ],

[o], and [u], the vowels are more naturally rounded than [i] and likely took less time to round. Although [e] is also a high vowel and physically more similar to [i], the mid-blend state closely resembles the blended state, indicating that [e] also took a short time to blend despite its characterization as a high vowel. For future studies, it could be worth investigating the average time it takes to round a vowel.

The summed amplitudes for each harmonic grouping were then divided by the summed amplitudes of the fundamental frequency and the second harmonic and referred to as the Harmonic Group Amplitude Ratio. This serves as a comparison for the amplitude of tone-affecting overtones to the perceived pitch. As shown in Figure 4.1.4, the mid-blend time state was not included in this analysis because of the dynamic nature of the mid-blend time state and the observed differences in blending times between vowels.

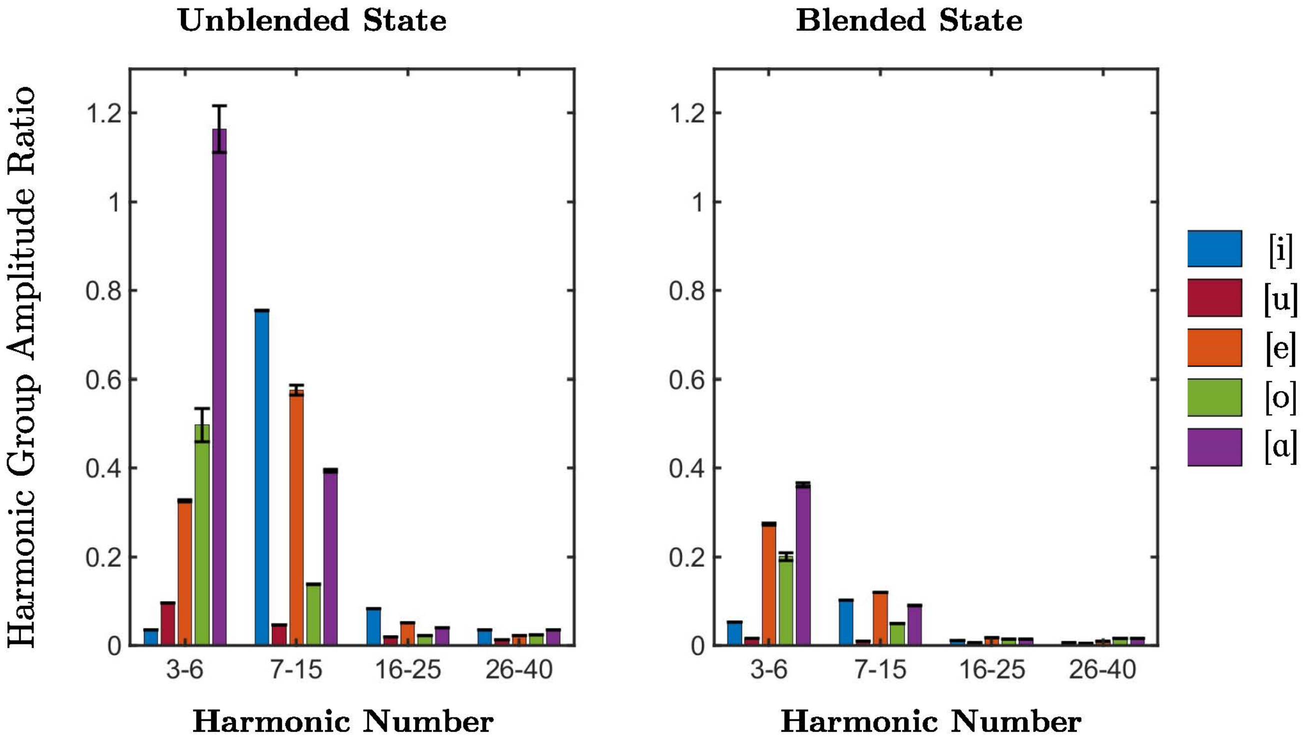


Figure 4.1.4 – Harmonic Group Amplitude Ratios for vowels [i], [u], [e], [o], and [ɑ]. For all vowels, there is a reduction in in Harmonic Group Amplitude Ratios between the unblended and blended states, which point to increased dominance of the perceived pitch.

As shown in Fig. 4.1.4, there is a consistent decrease between the Harmonic Group Amplitude Ratios, indicating that overtone suppression is present for all vowels when singers transition from an unblended to a blended time state. Reductions in Harmonic Group Amplitude Ratio indicate the dominance of the perceived pitch and darkness and pureness of the sound after blending.

The largest changes in ratio are in the first harmonic grouping 3-6 in the low vowels [ɑ] and [o], which are likely due to the placement of the vowel formants, which will be discussed in Sec. 4.2. For vowel [ɑ], the value of the Harmonic Group Amplitude Ratio is greater than 1 for the unblended state. Referring to Fig. 4.1.2, it is clear that the 3rd harmonic dominates this signal, which is equivalent to a P5 interval. In the unblended state, the 4th harmonic is approximately equal in amplitude to the 2nd harmonic, which is almost a factor of 2 larger than the fundamental. The decrease in amplitude ratio for the 3-6 harmonic group for [ɑ] after blending results from the decrease in the 3rd harmonic by a factor of 1.5 while the amplitude of the 4th harmonic remains constant. Therefore, the twofold reduction of the Harmonic Group Amplitude in Fig. 4.1.4 for harmonics 3-6 can be largely attributed to the shift from the 3rd to 2nd harmonic as the dominant overtone after blending. Notice that this results in a shift from a perfect fifth to an octave as the dominant overtone, indicating a purer tone.

In the blended state, all vowel Harmonic Group Amplitude Ratios for harmonics 3-6 are less than 0.4, less than 0.2 for harmonics 7-15, and close to 0.01 for harmonics 16-40. This implies that in the blended state, the fundamental and 2nd harmonic are at least 2.5 times more prevalent as harmonics 3-6, at least 5 times more prevalent than harmonics 7-15, and 100 times more prevalent than harmonics 16-40.

Overtone suppression can be expressed as a reduction in overall overtone amplitude, as defined by the difference in summed FFT amplitudes for harmonics 3-40 for unblended and blended vowels, divided by the summed FFT amplitudes for unblended vowels. These percentages are depicted in Fig. 4.1.5.

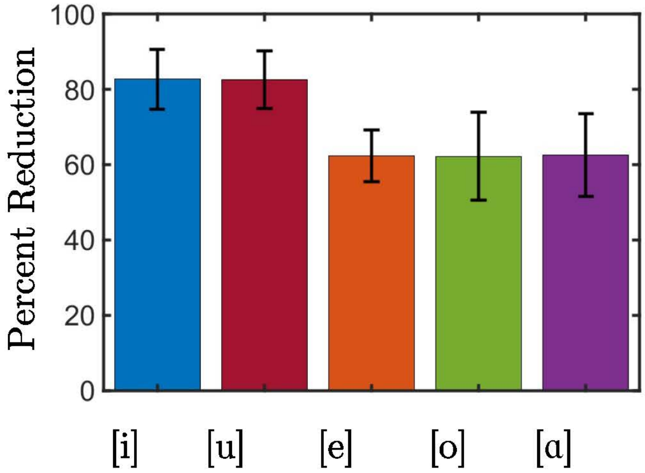


Figure 4.1.5 – Percent overtone reduction, expressed as the difference of summed FFT amplitudes for harmonics 3-40 for unblended and blended vowels, divided by the summed FFT amplitudes for unblended vowels. There is an 82% overtone reduction for vowels [i] and [u], and a 62% reduction for vowels [e], [o], and [ɑ].

For vowels [i] and [u], there is an (82 ± 8)% overtone reduction. This can be attributed to the comparatively higher presence of overtones in these vowels before blending, as indicated by Fig. 4.1.1. For vowels [e], [o], and [ɑ], there is a $(62 \pm 12)\%$ overtone reduction. This can be attributed to the comparatively higher presence of overtones in the vowels after blending, as shown in Fig. 4.1.4.

The low Harmonic Group Amplitude Ratios across all vowels as well as the 60 – 80% overtone reduction across all vowels supports the hypothesis of overtone suppression as a key factor in blending.

4.2 Vowel Adjustment and Lowering Formants

The frequency of the four vowel formants were averaged and tracked over time for all 5 trials, based on the methods described in Sec. 3.3. The first formant is represented by the bottom line. The dark lines represent the average formant frequency for the given time range, while the lighter lines represent the instantaneous frequency for each formant.

The normalized FFT amplitude data for the unblended and blended time states were plotted on a log scale in the ‘‘wing” plots on either end of the formant tracking plot to infer a relationship between the change in formant frequencies with the intensity of higher harmonics. This relationship is depicted in Figure 4.2.1.

(The rest of this page is left intentionally blank.)

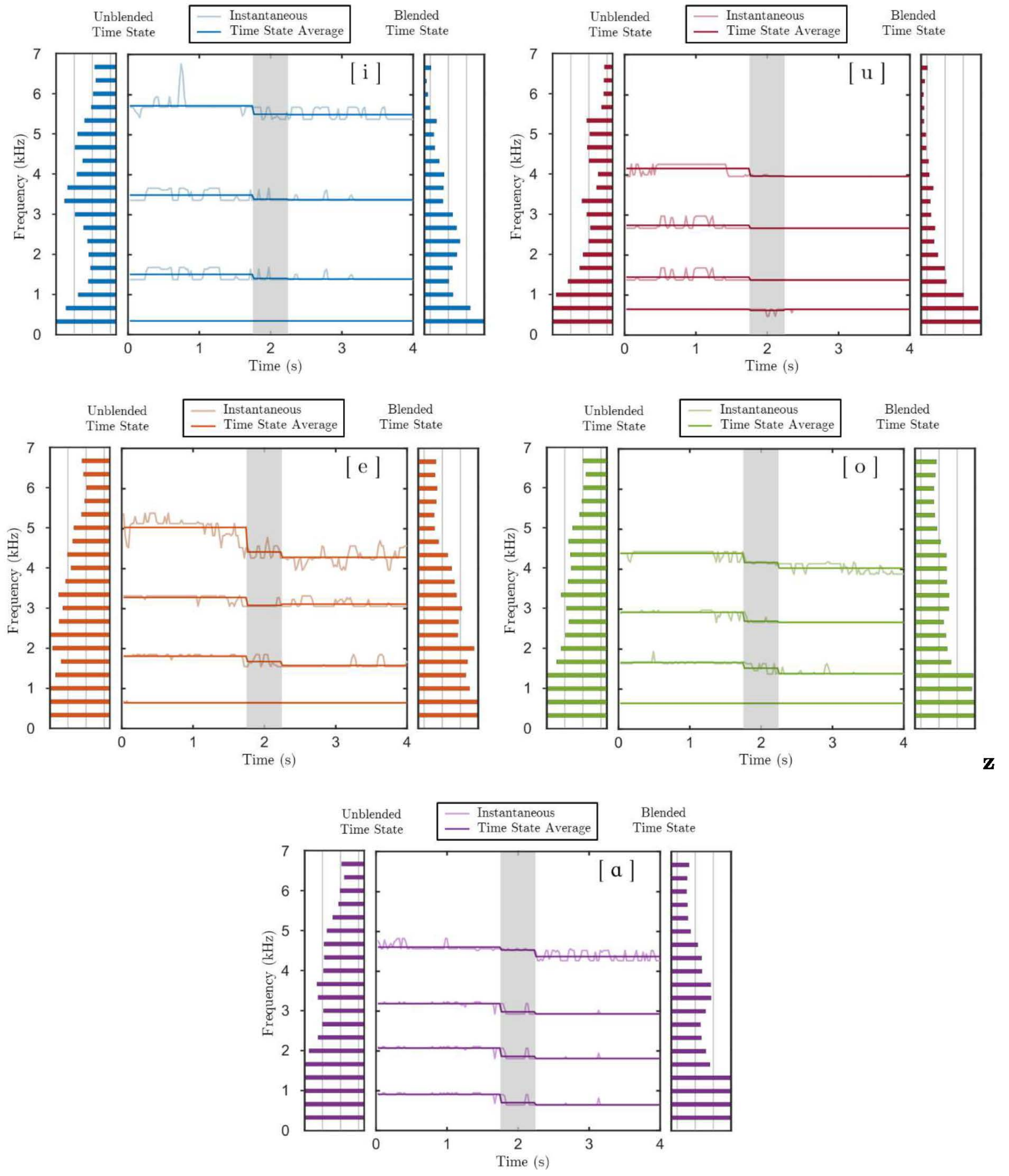


Figure 4.2.1 – Formant frequencies as a function of time, where the first formant is represented by the bottom line (lowest frequency) and the fourth formant is represented by the top line (highest frequency). The FFT amplitudes of each harmonic for the unblended and blended states plotted on a log scale from 10^{-4} to 1 are included as ‘‘wing” plots. All formants above the first decrease in frequency after blending, and there is a qualitative decrease in the intensity of higher harmonics. The first formant only decreases for [ɑ], as discussed in the text.

Across all vowels, there is a clear drop in the formant frequency between the unblended state (left of the grey) and blended state (right of the grey). The averaged data for [e], [o], and [ɑ] display an in-between step during the blending period that opens the possibility of further exploration of vowel shape as a determining factor of blend time.

For the current analysis, the start of blending was assumed to be 1.75 s for all vowels. However, as can be seen in particular for [e], [o], and [u], subconscious blending appears to have begun before this time, as indicated by increased deviation of the instantaneous frequency from the average for that time period. Future analysis could determine the “start of blending” from an increase in the standard deviation of the formant frequency in the time preceding the 2 sec “blending” time, and the end of blending as the time when the standard deviation again decreases.

Qualitatively, there is a dramatic decrease in the FFT magnitude for the higher harmonics in [i] and [u] and a downward shift of the local maxima. This qualitative shift less apparent but still present in vowels [e], [o], and [ɑ], which correspond to the 82% and 62% reduction in overtones for vowels [i] and [u] vs. vowels [e], [o], and [ɑ] as discussed in Sect. 4.1.

In Fig. 4.2.1, the first formant aligns to the highest amplitude harmonic in both the unblended and blended stages. For vowel [i], the first harmonic at around 293 Hz is the strongest, corresponding to the first formant. Similarly, for vowels [e], [o], and [u], the second harmonic around 586 Hz is the strongest, and the frequency is plotted as the second formant. Finally, for vowel [ɑ], the third harmonic and second harmonic are shown as strongest for the unblended and blended states, respectively.

For all vowels except [ɑ], the first formant does not change. For [ɑ], the drop of approximately 293 Hz can be attributed to the first formant dropping from the 3rd harmonic to the 2nd harmonic. The first formant of [i] aligns with the 1st harmonic, and for the other vowels aligns with the 2nd harmonic. From Fig. 4.2.1, there is a similar trend where the highest amplitude for an unblended [ɑ] is the at the third harmonic, while the highest amplitude drops to the second harmonic for a blended [ɑ].

These results point to the alignment of the location of the first formant – for vowels [u], [e], and [o], studies show that the first formant is somewhere between 300 and 600 Hz. Based on experimental data, the first formant appears to be closer to 600 Hz, the 2nd harmonic of D4, for the two altos sampled. As mentioned in Sec. 3.2, the dominance of the second harmonic does not significantly affect the timbre of the note since it is an octave above the fundamental.

For vowel [ɑ], the proximity of the first formant to the 3rd harmonic, a P5 interval, implies that there are overtones that dominate the sound and introduce a stronger note that is not the perceived pitch. The drop of the [ɑ] formant from the 3rd to 2nd harmonic, a P8 interval (octave), implies a purifying of the sound, where the perceived pitch and respective timbre becomes closer to that expected of the fundamental.

To quantify the shift in harmonics as a result of the lowering of the formants, the difference between the frequencies of the unblended and blended formants were calculated and divided by the fundamental frequency for D^4 (293 Hz) to quantify the approximate harmonic shift and are displayed in Figure 4.2.2.

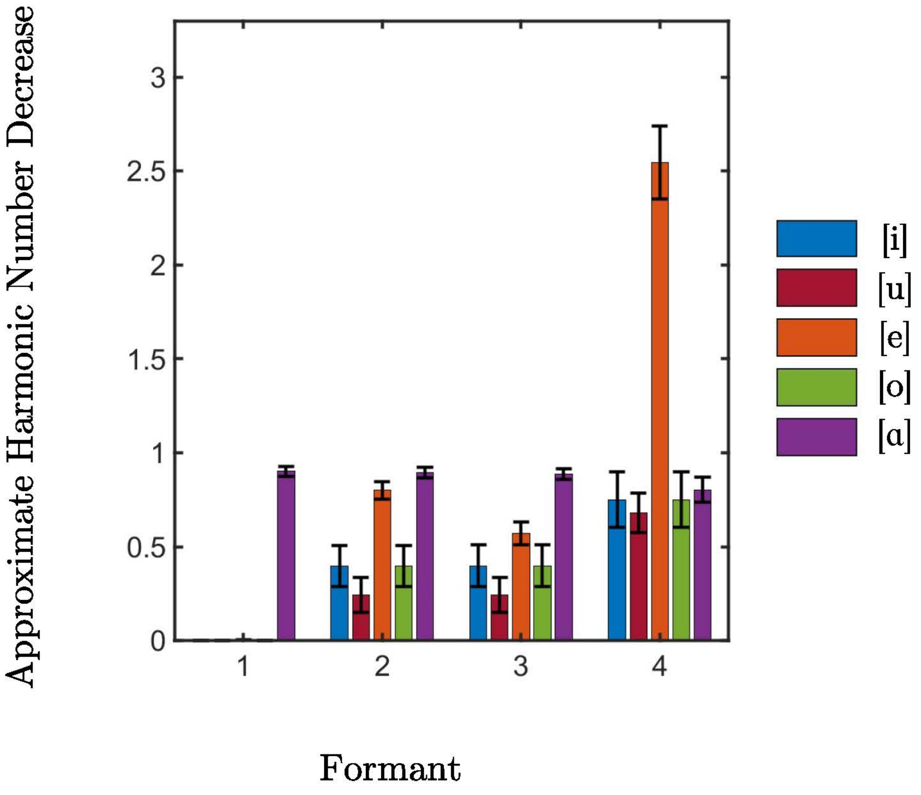


Figure 4.2.2 – The difference between the frequencies of the unblended and blended formants divided by the fundamental frequency (293 Hz). The decrease in blended formant frequencies for the 2nd – 4th formants is clearly evident. Non-integral changes in harmonic number are possible because of the definition of formant frequency relative to the median power of the cumulative PSD, rather than relative to the overtone series.

For the second, third, and fourth formants, the exact shift and alignment of formant frequencies to higher amplitude overtones is not an integer because of the way the formant frequency has been defined as the median power point, which does not have to line up with a single harmonic, especially if two adjacent harmonics have similar amplitude. It is still helpful to discuss the shift in formant frequency in units of the fundamental in order to attribute the change in formants to timbre darkening by overtone suppression.

There is a clear and consistent decrease in formant frequencies as indicated by Fig. 4.2.1, which has been quantified and expressed as a shift in amplified harmonics in Fig. 4.2.2. In the second formant, decreases in formant frequency correspond to approximately 0.3 - 0.4 overtones (88 – 117 Hz) for vowels [i], [u] and [o], and close to 1 full overtone (293 Hz) for vowels [e] and [ɑ]. The third formant follows a similar trend, with the exception of [e], where the third formant drop corresponds to approximately 0.5 overtones (147 Hz).

This is consistent with similar findings where vowel rounding is associated with a decrease in the second and third formants. This validates the choral practice of rounding vowels to achieve blend; the data from Fig. 4.2.2 that indicate vowel rounding also point to a shift toward lower, less dissonant harmonics.

For all vowels except [e], the fourth formant drops approximately 0.8 overtones (234 Hz); for [e], there is a decrease in 2.5 overtones (732 Hz). This large frequency drop is presently unexplained and merits further study.

In summary, all vowels except [ɑ] exhibited no change in the frequency of the first formant for the subjects tested. The 2nd and 3rd formants shifted by 0.3 – 1 harmonics, and the 4th formant shifted by about 0.5 harmonics for all vowels except [e], which exhibits an unexpected drop of 2.5 harmonics.

As formant frequencies decrease, amplified harmonics also move toward lower harmonics, which are often more consonant intervals to the fundamental and create a purer, more pleasant sound. Results point to a clear connection between the adjustment of vowels and the harmonics which the corresponding formants resonate with. This study has demonstrated a clear and consistent relationship between physical vowel adjustment, as evidenced by changes in formant frequency, and the timbre of a note, as evidenced by overtone suppression.

5. Conclusions and Recommendations

Blending can be characterized by overtone suppression in the frequency domain as well as a quantitative analysis of vowel formant changes. In vocal pedagogy, there is a well-established emphasis on blending by matching pitch, darkening tone, and matching and adjusting vowel shapes. This is achieved through overtone suppression and the lowering of vowel formants, which amplify lower harmonics at more consonant intervals to the fundamental. Comparison of the frequency spectra of unblended and blended vocalists in this study agrees with traditional vocal teachings and music theory, where blended tones correspond to fewer clashing and dissonant higher harmonics.

The summed Fourier amplitudes for higher harmonics as a fraction of the sum of the fundamental frequency and second harmonic decreases by (82 ± 8)% for vowels [i] and [u], and (62 ± 12)% for vowels [e], [o], and [ɑ] after transitioning to a blended sound.

Across all vowels, the 2nd, 3rd, and 4th formant drop 0.5 – 2.5 harmonics between the unblended and blended time states, indicating that vowel modification reduces the intensity of dissonant intervals in higher harmonics. Furthermore, vowel modification can be attributed as a key component of blending technique.

These results contribute a quantitative supplement to the qualitative instructions for blend, where vocalists are often taught to drop their jaws and round their vowels to create a qualitatively "darker" tone. Looking at the practice of vocal blend through a linguistic perspective opens up a different pedagogical approach to teaching vocal blend. Rather than instructing vocalists to match a certain shape by a vague description of a dropped jaw or pursed lips, vocalists can learn to sing blended tones by mimicking the specific sounds of the rounder, more back counterparts of the English vowels, emulating vowels from other languages.

The clear connection between vowel shape and the mechanisms of vocal blend is a powerful tool that can be used to identify the ideal vowel shapes for amplifying and suppressing different harmonics for a given pitch, thereby modifying the timbre. The methods practiced in this study can be applied beyond choral blend on the same pitch and voice range to choral blend on multi-note chords, where different vowel formations and adjustments can achieve a uniform manipulation of tone across a choir to carry out cohesive but dynamic performances.

References

[1]Lee, S.-H., Kwon, H.-J., Choi, H.-J., Lee, N.-H., Lee, S.-J., and Jin, S.-M., 2008, 'The Singer's Formant and Speaker's Ring Resonance: A Long-Term Average Spectrum Analysis," Clin Exp Otorhinolaryngol, **1**(2), p. 92.

[2] Oxenham, A. J., 2012, 'Pitch Perception," J Neurosci, **32**(39), pp. 13335–13338.

[3] Alm, J. F., and Walker, J. S., 2002, 'Time-Frequency Analysis of Musical Instruments," SIAM Rev., **44**(3), pp. 457–476.

[4] Deutsch, D., 2013, *Psychology of Music*, Elsevier.

[5] Terhardt, E., 1984, 'The Concept of Musical Consonance: A Link between Music and Psychoacoustics," Music Perception: An Interdisciplinary Journal, **1**(3), pp. 276–295.

[6] 'Power Spectral Density - an Overview (Pdf) | ScienceDirect Topics" [Online]. Available: https://www.sciencedirect.com/topics/engineering/power-spectral-density/pdf. [Accessed: 08-May-2020].

[7] Fillebrown, T., 1911, *Resonance in Singing and Speaking*, Oliver Ditson Company.

[8] Ladefoged, P., and Disner, S. F., 2012, *Vowels and Consonants*, John Wiley & Sons.

[9] Sundberg, J., 1988, 'Vocal Tract Resonance In Singing," p. 10.

[10] 'Handbook International Phonetic Association Guide Use International Phonetic Alphabet | Phonetics and Phonology," Cambridge University Press [Online]. Available: https://www.cambridge.org/gb/academic/subjects/languages-linguistics/phonetics-and-phonology/handbook-international-phonetic-association-guide-use-international-phonetic-alphabet. [Accessed: 14-Apr-2020].

[11] 'Scandinavian Languages - Phonology | Britannica" [Online]. Available: https://www.britannica.com/topic/Scandinavian-languages/Phonology. [Accessed: 08-May-2020].

[12] 'IPA Chart" [Online]. Available: https://www.ipachart.com/. [Accessed: 08-May-2020].

[13] Hillenbrand, J., Getty, L. A., Wheeler, K., and Clark, M. J., 1994, "Acoustic Characteristics of American English Vowels," The Journal of the Acoustical Society of America, **95**(5), pp. 2875–2875.

[14] 'Identification and Blend of Timbres as a Basis for Orchestration: Contemporary Music Review: Vol 9, No 1-2" [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/07494469300640341. [Accessed: 06-May-2020].

[15] Aspaas, C., Mccrea, C. R., Morris, R. J., and Fowler, L., 'Select Acoustic and Perceptual Measures of Choral Formation," Int J Res Choral Singing, pp. 11–21.

[16] Atkinson, D. S., 2010, 'THE EFFECTS OF CHORAL FORMATION ON THE SINGING VOICE," p. 11.

[17] "An Acoustical Study of Individual Voices in Choral Blend - Allen W. Goodwin, 1980" [Online]. Available: https://journals.sagepub.com/doi/10.1177/002242948002800205. [Accessed: 14-Apr-2020].

[18] Mitsuya, T., Samson, F., Ménard, L., and Munhall, K. G., 2013, 'Language Dependent Vowel Representation in Speech Production," J Acoust Soc Am, **133**(5), pp. 2993–3003.

[19] Ternström, S., and Sundberg, J., 1989, 'Formant Frequencies of Choir Singers," The Journal of the Acoustical Society of America, **86**(2), pp. 517–522.