

**PARSING THE SPECTRAL ENVELOPE:
TOWARD A GENERAL THEORY OF VOCAL TONE COLOR**

Ian Howell

Doctor of Musical Arts Thesis

The New England Conservatory of Music

Submitted: 18 May 2016

Corrected: 26 October 2016

Advisors: Katarina Markovic & Alan Karass

DMA Committee Readers: Matthias Truniger & Thomas Novak

Contents

Abstract.....	4
Acknowledgements.....	5
Foreword	6
1. How We Draw Vowels: An Introduction to Current Models	8
2. What are Timbre and Tone Color?.....	20
3. Exploring the Special Psychoacoustics of Sung Vowels.....	27
<i>Absolute Spectral Tone Color</i>	<i>29</i>
<i>The Multiple Missing Fundamentals</i>	<i>32</i>
<i>Local Spectral Coherence</i>	<i>35</i>
<i>Weak Tone Color Bridging.....</i>	<i>40</i>
<i>The Obvious True Fundamental.....</i>	<i>41</i>
4. Analysis.....	44
<i>Parsing the Spectral Envelope: A Model for Aurally Locating Vocal Tone Colors.....</i>	<i>46</i>
<i>The Classical Female (and Countertenor) Voice above the Treble Staff.....</i>	<i>50</i>
<i>The Classical Female (and Countertenor) Voice on the Treble Staff.....</i>	<i>54</i>
Conclusions.....	57
Appendices.....	58
<i>Appendix A: International Phonetics Alphabet Symbols.....</i>	<i>58</i>
Figures	59
Bibliography.....	97

Except where noted, Copyright © 2016 Ian Howell, all rights reserved

Abstract

The purpose of this study is to identify issues with the visual models of sung vowels currently used within singing voice pedagogy and voice science texts, and to propose a conceptual framework and new visual models that may more accurately characterize objective elements of timbre present in the singing voice. The timbre of the classical female (and countertenor) voice exposes blind spots in these spectrographic and schematic models, notably that they accommodate certain ambiguities present in speech, but problematic when applied to singing (especially high-pitched, melismatic singing). Essentially, above the treble staff, vowel clarity disappears entirely not because of vowel substitutions or modifications, but because the simplicity of the listener's percept is too distant from the timbral complexity of speech. The manner in which this vowel clarity changes as pitch ascends informs a meaningful discussion of the psychoacoustics of sung pitches throughout the range of both male and female voices, and suggests locating the source of timbre not within the singing body (at the point of production), but rather within the listener's paradoxically limited hearing mechanism (at the point of perception). The author hopes to point toward a still elusive general theory of vocal tone color by proposing the following five principles of singing voice perception currently absent in the voice science and vocal pedagogy literature: absolute spectral tone color, the multiple missing fundamentals, local spectral coherence, weak tone color bridging, and the obvious true fundamental. This thesis explores both immediate pedagogical implications of this framework for singers and voice teachers, and also points to substantial revisions to the models used in the singing voice pedagogy and voice science literature.

Acknowledgements

This thesis would not have been possible without the support (both material and intellectual) and guidance of many people within the New England Conservatory of Music community, including Tom Novak, Karen Holvik, Bradley Williams, Carole Haber, Robert Cogan, Matthias Truniger, Helen Greenwald, Katarina Markovic, Alan Karass, and Jason Coleman. I have received encouragement along the way from many within the field of singing voice pedagogy and voice science, including Kenneth Bozeman, Margaret Baroody, Linda Carroll, Karen Brunssen, Allen Henderson, and Scott McCoy. I offer additional thanks to Dann Mitton at the University of Toronto for a close, final read; Bodo Maass of Sygyt Software for both authoring the powerful spectrograph *Overtone Analyzer*, and also quickly responding to my numerous questions and feature requests; and to all the students who have gone through my vocal pedagogy classes at the New England Conservatory. My greatest thanks is reserved for my wonderful and supportive wife and partner, Virginia Pierce, and for my daughter Leah, whose remarkably high frequency spectral peaks have been fascinating to study over the first two years of her life.

Foreword

This thesis grew out of an attempt to find meaningful connections between the spectrographic voice analysis techniques currently used by voice scientists and pedagogues, the spectrographic music analysis approach pioneered by Professors Robert Cogan and Pozzi Escot at the New England Conservatory of Music in Boston, and the work of numerous psychoacousticians and linguists dating back to the 19th century.¹ These fields frequently borrow ideas and models from one another. At times this brings clarity. At other times the most useful approach for one field is problematic when directly applied to another. The manner in which singing voice science and pedagogy have embraced the visual models of vowels derived from speech science is just such an issue. Linking the objective information captured in a spectrogram (or derivative graph) to the perceived tone color of a sung vowel, especially as pitch rises, is challenging. Singers frequently vocalize in pitch ranges that produce a far less complex percept (the mental impression of the sound perceived) than speech, and the models that describe speech lack the specificity needed to describe the timbre of a resonant vowel sung at a high pitch. Those of us who study elite singing—and especially those who attempt to teach voice science and pedagogy to musicians—would benefit from a singing-specific model devoid of the tolerance for timbral variation found in speech. Although my primary audience is those working in singing voice science and pedagogy, I hope this thesis inspires further interdisciplinary collaborations between these equally valuable approaches.

¹ Spectrographic analysis finds a place in almost every modern vocal pedagogy and voice science text, e.g. the works of Vennard (1967), Doscher (1994), Coffin (1980, 1987), Donald Miller (2008), Titze (2000, 2009), Sundberg (1987), Bozeman (2013), and McCoy (2012). Robert Cogan and Pozzi Escot (1976) and Cogan (1984, 1998) lean heavily on spectrographic analysis to capture sonic elements of musical structure that elude notation, and the work of psychoacousticians like Helmholtz (1877), Mach (1885), Engel (1886), Stumpf (1890), Winckel (1967), and Plomp (1966, 2002), and speech scientists like Peterson (1952) to explain the vowel-like registers of all sounds. Indeed, Cogan has made the most significant connections between these fields to date.

A few terms to be used throughout bear clarification:

1. I will use scientific pitch notation, e.g. C4 is middle C, C5 an octave higher, B3 a half-step lower.
2. The terms *overtone* and *partial* are often used to describe the whole number multiples of the fundamental frequency that form the harmonics of a sound with pitch. The former refers to those harmonics higher than the fundamental; the latter includes the fundamental. I will use the term *harmonic(s)* (essentially equivalent to *partial*) to refer to these. The lowest frequency harmonic (with the frequency equal to the pitch) will be the first harmonic. The second harmonic is an octave higher, the third a further fifth higher, and onward through the harmonic series. At times I will also use the term *fundamental*. This is equivalent to the first harmonic.
3. I will use the International Phonetics Alphabet (IPA) and at times phonetic notation.²
4. I will use both linear and logarithmic displays of frequency when graphing the harmonics in figures, depending on which is more appropriate to the task at hand.
5. I will avoid the term *formant* as it currently means contradictory things to different parties. Two bedrock concepts of voice science inform the terms I will use. The vocal tract has changeable, pitched resonances that filter the harmonics produced by the vocal folds. The sound wave that reaches a listener will have high amplitude peaks of these harmonics in certain frequency ranges and low amplitude troughs in others (see **Figure 1**). Both of these phenomena, vocal tract resonances and the spectral peaks in the radiated sound, are currently called formants. I will endeavor to bring clarity by referring to them as vocal tract resonances and spectral peaks respectively. This includes clarifying the use of the term formant in quoted material. As the voice generally has multiple vocal tract resonances and spectral peaks for any one vowel, they will be numbered from lowest to highest in frequency, starting with the number one.
6. I will use the term *spectral segment* (or just *segment* if the context makes it clear) to mean a notch-filtered band of frequencies that is a contiguous portion of a sound's total spectral envelope. Linguists frequently use the term *segment* to mean a defined period of time of the total spectral envelope. My use of this term implies no temporal aspect.
7. When appropriate, videos of the samples illustrated in the figures, and special videos explaining specific concepts may be found here: <<http://goo.gl/Agt4kp>> or <<http://vocped.ianhowell.net/figures>>. In some cases, especially in chapter three, the reader will benefit significantly from experiencing the phenomena described in the text.

² Please see **Appendix A** for a list of common IPA symbols (shown by their [] brackets), which indicate a reasonably specific sound. If a symbol appears between slashes / /, it is a phoneme, a label applied to a group of similar sounds that can have a particular function within a word.

1. How We Draw Vowels: An Introduction to Current Models

Timbre is a slippery concept and a slippery percept, perceptually malleable and difficult to define in precisely arranged units.

Cornelia Fales³

More than a generation ago William Vennard (1967) wrote a compendious book on the physiology and acoustical science of singing voice technique.⁴ After outlining the best speech science of the day, and graphing the average distribution of spectral energy for the five Italian vowels,⁵ he notes that certain vowels share spectral peaks: “...when one sings Ay [e], he is really singing Oh [o] plus a high partial which is not heard in the Oh [o]; and when one sings Ee [i], he is really singing Oo [u], plus a still more ringing overtone.”⁶ By calling attention to the actual *tone color* of an isolated vowel’s separate spectral peaks, rather than their objectively measurable frequencies, Vennard threw down a gauntlet that remains on the ground today. Why this may be so, why it goes unnoticed, and how one may rise to his challenge are important questions to contemplate.

³ Cornelia Fales, “The Paradox of Timbre,” *Ethnomusicology* 46/1 (2002): 56-95, here 58.

⁴ William Vennard, *Singing: The Mechanism and the Technic* (New York: Carl Fischer, 1967).

⁵ Vennard, *Singing*, 127.

⁶ Vennard, *Singing*, 130. International Phonetics Alphabet (IPA) symbols inserted by author. See **Appendix A** for a list of common IPA symbols and examples.

Vocal pedagogy and voice science texts use several visual models to characterize the differences between vowels, notably variations of the spectral envelope and schematics (or images) of the vocal tract.⁷ These types of models display objective, rather than perceptual measurements. In the former, spectral peaks—usually representing average frequency centers for a large population of speakers—are indicated on an X/Y graph, simultaneously displaying either frequency and amplitude or the frequency centers of the two to five lowest spectral peaks. The information conveyed in these graphs may be represented on an interval scale by either frequency in Hertz (Hz) or musical pitch, with specificity limited only by the resolution of the measuring device (examples shown in **Figure 2**). The latter demonstrates the average position of the vocal tract articulators for a given vowel, notably the tongue, jaw, soft palate, and lips. While precision is reduced, the general position and range of motion for each articulator are easily shown (examples shown in **Figure 3**).

These models do clarify the spectral and physiological elements that differentiate various spoken vowel phonemes; however, they accommodate certain ambiguities that are problematic when applied to singing. While one might assume that a given vowel would be spectrally identical from speaker to speaker, the spectral peak frequency centers for a spoken vowel vary across a large population. Two somewhat similar vowels may have near identical peak frequencies, and drastically different vowels may share peaks in common. This variability suggests that a vowel percept is the result of the total shape (the spectral envelope) rather than discrete content of its spectrum. However, intelligibility of familiar words is not significantly degraded by filtering only a portion of the spectrum of speech and, with some limitations, it does

⁷ The spectral envelope is a visualization of the spectral content of a sound, broken up into the distribution of energy by frequency. In a sound with a pitch—such as a sung vowel—the spectral energy is generally concentrated most strongly in harmonics (whole number multiples of the fundamental frequency). The vocal tract consists of the contiguous air containing spaces between the vocal folds and the opening of the mouth and/or nose.

not matter which portion of the spectrum is filtered.⁸ **Figure 4** spectrographically displays the sentence fragment, "...my four little children will one day live in a nation where they will not be judged by the color of their skin but by the content of their character, I have a dream today," from Dr. Martin Luther King's well known speech.⁹ This text is presented as originally recorded, and then filtered three ways: the spectrum below 2,000Hz, above 2,000Hz, and between 300 and 3400Hz (the approximate bandwidth of the analog telephone system).¹⁰ The text is understandable in all four versions, which supports the idea that speech sounds are simultaneously encoded across the entire spectrum. Anyone who has listened to speech through an analog telephone or music through a tiny speaker incapable of reproducing the full range of human hearing knows that with even a limited amount of the spectrum present, information may certainly be conveyed. While spectrograms, power spectrums, and spectral envelope graphs (perhaps the most technically accurate models) display the spectral content of a complex sound, they offer little in the way of an obvious key to link what one sees to how one perceives it. One does not hear a voice as a chord of separate harmonics, though a spectrogram and a power spectrum display it as such. Finally, the spectral peak frequency locations of spoken vowels conveyed by these models become progressively less useful for sung vowels as pitch rises due to the inevitable misalignment of changing voice source harmonics and static vocal tract resonances.

Models of vocal tract shape and articulator position are perhaps more directly helpful for speakers and singers as they describe in concrete terms what to *do* in order to produce a specific

⁸ See Peter B. Denes and Elliot N. Pinson, *The Speech Chain: The Physics and Biology of Spoken Language 2nd Edition* (New York: W.H. Freeman, 1993, reissued 2015), 140-183 for a thorough description of these paradoxes.

⁹ The reader is reminded that videos demonstrating most figures are available online at either <<http://goo.gl/Agt4kp>> or <<http://vocped.ianhowell.net/figures>>.

¹⁰ See www.telecomabc.com/b/bandwidth.html

sound. Government Phonology, a linguistic theory proposed by Kaye, Lowenstamm, & Vergnaud (1985), describes vowels by lists of prominent features of the vocal tract. Vowels with lists containing only one important feature are called *elements*, and include [a], [u], and [i], which are found in almost all human languages. These vowels are assigned a status not unlike primary colors in that they are both autonomous and the building blocks of other vowels. Kaye, et al suggest, “an...[a] element may combine with an... [i] element to form a compound vowel [ε]. In like manner...[a] and...[u] combine to form [ɔ].”¹¹ Story, Titze, and Hoffman (1996) tested this approach, “...with magnetic resonance imaging (MRI) of the vocal tract to obtain volumetric shape functions of the airway... [and] found that the airway shape for any vowel in the English language can be constructed by taking percentages of /ɑ/, /u/, and /i/ (and perhaps /æ/) shapes and adding them to the neutral shape.”¹² However, when Kaye, et al suggest that [ε] is a combination of [a] and [i], they mean that the vocal tract for [ε] is between the shapes for [a] and [i], not that the sounds actually combine. If Vennard is correct that [i] contains [u], perhaps Government Phonology misses a more autonomous division of these primary vowels. The vocal tract shape is very different for [u] and [i] (see **Figure 5**), and nothing visually suggests their common spectral feature. What they share is timbral, not physiological.

As with spectral models, vocal tract shapes for spoken vowels break down as pitch rises and the harmonics present fall out of alignment with the vocal tract resonances. In order to continue singing resonantly as pitch ascends, the singer may have to actively modify their vocal tract, generally by opening the jaw, laterally spreading the lips, or raising the larynx (frequently

¹¹ Jonathan Kaye, Jean Lowenstamm, and Jean-Roger Vergnaud, “The Internal Structure of Phonological Elements: A Theory of Charm and Government,” *Phonology Yearbook* 2 (1985): 305-28, here 306.

¹² Ingo Titze and Katherine Verdolini Abbott, *Vocology: The Science and Practice of Voice Habilitation*, (Salt Lake City: National Center for Voice and Speech, 2012): 347, referencing Brad H. Story, Ingo Titze, and Eric A. Hoffman, “Vocal tract area functions from magnetic resonance imaging,” *Journal of the Acoustical Society of America* 100/1 (July, 1996): 537-54.

discouraged in classical training).¹³ Doing so changes the vowel's timbre. Berton Coffin (1980) offers fifteen degrees of opening (deviations from the vocal tract shape for speech) for any one vowel. However, he does acknowledge that, "...[i] with the 6th degree of opening sounds like [ɛ] and [i] with the 11th degree of opening sounds like [æ]."¹⁴ This calls into question the utility of continuing to conceive of the [i] as an 'i shape' as pitch rises and the mouth opens. Indeed, Coffin suggests that some students will prefer to think of the changing vowel color over the degree of opening.¹⁵ As a third possibility, perhaps opening vowels in this manner forms vocal tract shapes unlike those found in speech, but effective in singing.

As Kenneth Bozeman (2013) explores, the converse is also true: changing the pitch while retaining vocal tract shape sometimes passively modifies the vowel. As pitch rises while the vocal tract remains the same (meaning the pitches of the vocal tract resonances remain the same), harmonics of the voice cross the resonances. Each time this occurs, but especially when the second harmonic crosses the lowest pitched vocal tract resonance, the vowel color shifts in a predictable way. This is the basis of the "pitch of turning" found in Bozeman's (2013) registration framework.¹⁶ These two approaches are complementary and solve resonance issues in different pitch ranges for different vowels. Both push back against the idea that a single vocal tract shape will consistently produce the same vowel as pitch changes.

If I may generalize, spectrographic and vocal tract shape models work well to describe speech specifically because of qualities found in speech, yet frequently absent in singing. The

¹³ The term *resonance*, when applied to singing, implies not just that the sound wave created by the vocal folds finds a compliant response in some portion of the spaces comprising the vocal tract; in practice a resonant voice also exhibits a degree of perceptible ease. So resonant singing implies a high gain for a low effort.

¹⁴ Berton Coffin, *Overtones of Bel Canto* (Metuchen: Scarecrow Press, 1980), 25.

¹⁵ Coffin, *Overtones*, 25.

¹⁶ See Kenneth Bozeman, *Practical Vocal Acoustics: Pedagogical Applications for Teachers and Singers* (Hillsdale, New York: Pendragon, 2013), 26, for both a discussion of the pitch of turning, and the concept of passive versus active vowel modification.

relatively low, constantly fluctuating pitch of speech creates a rich spectrum of harmonics in an otherwise healthy voice, obviating the need to fine tune vocal tract resonances to match voice source harmonics. As Titze (2000) writes,

Regulation of intensity by... [vocal tract resonance] tuning is not relevant for conversational speech and for low-pitched singing. For fundamental frequency on the order of 100-200 Hz [about G2 to G3], the harmonics of the source are spaced close enough so that the... [vocal tract resonances] are energized at all times. Usually more than one harmonic resides in a... [vocal tract resonance] region, making the total output less dependent on a single harmonic of the source.¹⁷

Slight variations in the frequency centers of the spectral peaks of speech rarely cause significant changes in the density or general timbre of harmonic information. Keeping in mind that spectral peaks are caused by harmonics profitably interacting with vocal tract resonances, by raising the pitch and removing the bulk of the harmonics, one also removes the ability to create the subtle spectral variations that characterize speech. **Figure 6** demonstrates how quickly harmonics shed this quality as pitch rises (here a comparison of the linear spacing of harmonics at five pitches from C3 to F6).

Indeed, ongoing timbral variation is characteristic of, rather than incidental to speech. Spoken phonemes are not static, isolated units.¹⁸ Imagine the difference between the /k/ sounds in *keep* and *cool* to get a sense of how the vocal tract shape of the vowels [i] and [u] change the tone color of the initial consonant. Now compare the [i] in *tea* and *peep* to notice how the /i/ is subtly changed by the initial consonant. In speech, the vocal tract shape of a given phoneme affects the motion and position of the articulators for the phonemes that precede and follow, a

¹⁷ Ingo Titze, *Principles of Voice Production*, 2nd Edition (Iowa City: National Center for Voice and Speech, 2000), 258.

¹⁸ See Denes, 143, "...the speech wave has very few segments whose principal features remain even approximately static."

process called *coarticulation*.¹⁹ See **Figure 7** for an example of a single female subject speaking the text “Oh what a lovely day” in three pitch ranges. Note that the higher the pitch, the less dense the harmonic information available to create subtle variations in the spectral envelope; e.g. the diminishing detail of the glide from /l/ to /i/ in “lovely” from top to bottom. If timbre encompasses changes in the spectrum over time, there is simply less potential for timbral variation at higher pitches. I have carried out informal experiments playing the bottom (around the pitch D6) and top (a comfortable speaking pitch) samples of **Figure 7** for subjects with no knowledge of the text. Subjects cannot comprehend the words in the high-pitched sample the first time. However, after hearing the text as regular speech, they can understand the bottom sample the second time.

For musicians, the most troubling issue may be that these models do not describe aspects of tone color, timbre, or the harmonic complexity²⁰ of vowels with any degree of specificity. Indeed, Robert Cogan and Pozzi Escot (1976) lament that, “A theory of musical tone color has yet to be created.”²¹ Though the voice science and pedagogy communities know that the lowest two (or three) spectral peaks play a significant role in vowel identification,²² the models explored above offer no mechanism to confront the timbral role of an individual singer’s specific spectral peaks. If a vowel’s spectral peaks each have multiple harmonics in speech, but fewer when sung at a higher pitch, how might this change be characterized? If a singer wants to sing a pitch higher than the spectral envelope for the vowel suggests is possible, what is the resulting tone color (see

¹⁹ See Mark H. Ashcraft, *Cognition* (Upper Saddle River, NJ: Pearson Prentice Hall, 2006), 382-3, for a discussion of coarticulation. See also Terrance M. Nearey, “Static, dynamic, and relational properties in vowel perception,” *Journal of the Acoustical Society of America* 85/5 (May 1989): 2088-2113, for a discussion of conflicts between the *context effect* and *inherent quality* in current speech research.

²⁰ In this context, harmonic complexity refers to the number of harmonics that constitute each spectral peak.

²¹ Robert Cogan and Pozzi Escot, *Sonic Design: The Nature of Sound and Music* (Englewood Cliffs, New Jersey: Prentice-Hall, 1976), 328.

²² See Bozeman, 12 and Denes, 142-3.

again **Figure 6**)? If the ideally resonant vocal tract shape for a pitch and vowel produces a different vowel as pitch changes (Bozeman's passive modification), how should one explain the difference? Nor do these models represent the energy of higher harmonics—which certainly shape timbre, but also define the pitch and complexity of the sound wave—in a manner consistent with human perception. As I will demonstrate, the pitch one perceives is not governed solely by the energy of the harmonic that appears to be the fundamental on a spectrogram (furthest to the left in **Figure 6**).

The psychoacoustics literature that addresses the question is unambiguous that simple tones, "...are characterized by a typical frequency-dependent timbre."²³ Reinier Plomp (1966) argues that the prevailing view of late 19th and early 20th century researchers was that not only did simple sounds, "...have some resemblance, depending upon frequency, with particular speech vowels," but that the timbre of complex sounds can only be understood based on this assumption.²⁴ Psychoacoustics labels this quality (somewhat unimaginatively) *brightness*.²⁵ Much later in his career, Plomp (2002) warns against the bias of describing, "...the timbre of a complex tone...simply as the sum of the timbre of its sinusoidal components."²⁶ Indeed, one does not hear a sung vowel as a chord of sinusoidal tones; it resolves into a single percept. It is logical that the timbre of each sinusoidal harmonic also resolves in some manner.

It is not a conceptual leap for singers to notice that a vowel has multiple colors. Classical singers frequently discuss the *chiaroscuro*—or bright/dark quality—of the ideal voice, recognize

²³ Reinier Plomp, *Experiments on Tone Perception* (Soesterberg: Institute for Perception RVO-TNO, 1966), 131-2.

²⁴ Plomp, *Experiments*, 132. See also Robert Cogan, *Music Seen, Music Heard: a picture book of musical design*. (Cambridge: Publication Contact International, 1998), 110, for another history of this line of thought.

²⁵ Plomp, *Experiments*, 132.

²⁶ Reinier Plomp, *The Intelligent Ear: On the Nature of Sound Perception* (London: Lawrence Erlbaum, 2002), 133.

that the ring of the singer's formant²⁷ is distinct from the vowel (i.e. a vowel can have or lack the singer's formant), and do not expect that a soprano's C6 will exhibit the same richness and complexity as a bass' G2. However, our teaching models currently lack objective language to describe this, and discourse usually devolves into metaphors. Critically, a new model must accommodate the difference between the inherent experience of tone color and the phonetic context that allows one to understand words when spoken or sung. Plomp (2002) terms this *audition* ("...the ways in which the sound stimulus is processed... so that its specific characteristics are preserved,") and *cognition* ("...the way in which our previous experience with speech is used to interpret the new signals").²⁸ Clearly speech engages both phenomena in a manner that wordless, high-pitched, melismatic singing does not.

Returning to Vennard's challenge, I will demonstrate that the spectral region shared by [u] and [i] (the first spectral peak of both vowels) indeed sounds like [u]. These vowels do not just share an objectively measurable portion of the spectrum, they share a *tone color*. The [i] vowel is at minimum timbrally dualistic. However, there is almost always spectral energy in the range characteristic of [i] present in the [u] as well, albeit at a much lower relative amplitude.

Figure 8 demonstrates how quickly this can become confusing. In these examples of a synthesized sung voice, [u] shares a second spectral peak with [ɔ], [i]'s second peak is the same as [y]'s third, and [a] and [ʌ] differ only in first peaks. A chart like this—that excludes both the relative amplitudes and spectral complexity of each peak—leaves one with few clues to decipher how these vowels objectively differ. Given that speech is comprehensible with only a narrow

²⁷ The singer's formant is a clustering of vocal tract resonances in the region centered on 3kHz that powerfully amplifies harmonics. Its frequency center varies with voice type (generally falling between 2kHz and 4kHz), and its presence (especially in male voices and in female voices on and below the staff) is generally accepted as an indication of a cultivated classical singing voice. See Sundberg, *The Science of the Singing Voice*, 123-4, Donald Miller, *Resonance in Singing*, 120 and Koichi Omori, et al, "Singing power ration: Quantitative evaluation of singing voice quality," *Journal of Voice* 10/3 (1996): 228-235 for further information.

²⁸ Plomp, *The Intelligent Ear*, 104.

part of the spectrum present, and that vowels apparently share spectral characteristics in a yet unclear way, while one can say that this chart illustrates something true about the physical properties of these sound waves, it describes neither how one perceives them, nor what would be lost as specific peaks change due to a rise in pitch.

Perhaps a modification to Vennard's analogy is appropriate: all tone colors exist in varying proportions in all vowels. A given spectral peak will always exhibit a dependable timbre related to the frequencies of its constituent harmonics; vowels are the complex combination of the timbres of these peaks. Indeed, all tone colors (and thus all vowels) could be said to be present in the buzzy sound that the vocal folds make prior to the spectral transformation of the vocal tract (see again **Figure 1**). One does not notice this because they are all happening at the same time! For that matter, they exist in every sound, regardless of its source; perhaps devoid of the additional sonic structures that differentiate a voice from a chainsaw, but present nonetheless. Imagine the low 'ooooooo oooooo' of a foghorn or brilliant 'eeeeeeeeeee' of squealing car brakes.²⁹ These onomatopoeic labels are both intuitive and inseparable from a human's perception of the physical properties of each sound wave.

The singing voice (especially in the classical style) is in many ways the opposite of speech. Singing frequently demands a higher fundamental pitch, vowel durations augmented beyond the cadence of speech (and often extended over changes in pitch), a minimization of resonance-inhibiting coarticulation, and a reduced palette of vowel colors as pitch ascends. Coarticulation certainly exists in classical singing, however singers frequently modify consonants to be as non-disruptive to vowel resonance (and modify vowels to remain as constantly resonant) as possible. This aesthetic result is driven by the acoustical demands of the

²⁹ As addressed in Robert Cogan's register based system of tone color, found in Robert Cogan, *New Images of Musical Sounds* (Cambridge: Harvard University Press, 1984), 7 and 12, (see **Figure 18**).

workplace, which values voices capable of being heard unamplified over orchestral instruments. In contrast to speech, a singing-specific analytical model must order tone color on an objective scale, rather than a subjective scale of relative brightness. It must also explain how a given segment of the spectrum³⁰ (whether a peak or not) contributes its tone color to the vowel. This will depend on the shape, complexity, and amplitude of the segment relative to the rest of the spectrum. Any model capable of accommodating this must explain the way in which each sung spectral peak comes to have its specific tone color, and how the tone colors of the multiple spectral peaks present in almost all vowels relate to one another.

I heed Plomp's (2002) warnings, who in criticizing the limitations of the microscopic, sonically sterile study of sound that characterizes much 20th century research, instead pushes us to observe sound in its natural (chaotic) environment.³¹ Cogan and Escot similarly suggest that tone color analysis must focus on the relationships between subsequent sounds, and that, "...analysis cannot limit itself merely to the description of *single* sounds, no matter how technically sophisticated that description may be."³² François-Joseph Fétis perhaps best summed up these concerns in 1878: that Helmholtz wanted to, "...annul the delicate sensations of the artistic ear for the benefit of essentially brutal calculations."³³ However, one must recognize that an analytical middle ground exists, especially for an instrument capable of both cognition-dependent speech and audition-dependent melismatic singing. One should strive to find ever more meaningful ways to understand the tone colors of the sounds that combine in music; indeed, the timbre of a voice transmits significant information regarding vocal technique. To say

³⁰ Again, here the term 'segment' refers to a frequency range of the spectrum, not a period of time of the total spectrum.

³¹ Plomp, *The Intelligent Ear*, 132-137.

³² Cogan and Escot, *Sonic Design*, 328.

³³ Benjamin Steege, *Helmholtz and the Modern Listener* (Cambridge: Cambridge University Press, 2012), 80.

that nothing remains to be learned about an individual sound's timbre is to discount our ability to learn, not to discount the potentially meaningful hidden properties of that sound or its intrinsic characteristics relative to its musical context.

To best understand the whole of a sung vowel's timbre, and to engage and analyze it according to its inherent properties, I believe we must break it apart into conceptual and perceptual units smaller than common sense suggests exist. Graphical timbre analysis of the singing voice is currently problematic because of the shortcomings of the models themselves. New models ought be created and incorporated into the vocal pedagogy literature. Such models stand to significantly improve our understanding of singing registration choices (regardless of genre), and clarify the actual *sounds* of formant tuning and acoustic resonance strategies commonly discussed in the vocal pedagogy literature.³⁴ This thesis is an exploration of natural phenomena as they arise in music and an attempt to offer new language to bring clarity to the percept of singing. What follows is both a challenge to our assumptions about sound, timbre, and vowels, and also a prescriptive framework for thinking about which vowel timbres the human singing voice is capable of healthily producing as pitch changes. When one considers that this material has the potential to fundamentally change the way singers and voice teachers *hear* technical deficiencies, the practical applications are near limitless. While the complexity of speech introduces elements not present in singing—and is thus incompletely described by a singing-centric model—my hope is that this exploration will favorably impact the study of all sound perception, speech included.

³⁴ Here *formant tuning* refers to the practice of adjusting the vocal tract to raise or lower the pitch of the vocal tract resonances to better agree with the frequencies of certain voice source harmonics. See Donald Miller, *Resonance in Singing* (Princeton: Inside View Press, 2008), Bozeman, *Practical Vocal Acoustics*, and Scott McCoy, *Your Voice: An Inside View* (Delaware: Inside View Press, 2012), Barbara M. Doscher, *The Functional Unity of the Singing Voice* (London: Scarecrow, 1994), Berton Coffin, *Coffin's Sounds of Singing* (Lanham: Scarecrow Press, 1987), and Coffin, *Overtones of Bel Canto* (Metuchen: Scarecrow Press, 1980) as examples of commonly referenced texts suggesting this approach.

2. What are Timbre and Tone Color?

Timbre, of all the parameters of music, is the one least considered. It lacks not only an adequate theory, but even an inadequate one.

Robert Cogan³⁵

Those who study timbre frequently criticize its definition: “Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.”³⁶ According to this definition, timbre is a way to characterize the *differences* between the spectro-temporal character of one sound and another (e.g. from different sources or the same source but different mechanical or acoustical adjustments). Patterson (2010) writes,

Informally, the standard definition of timbre is regarded with considerable amusement. You might expect the definition of timbre to tell you something about what timbre is, but all the definition tells you is that there are a few things that timbre is not. It is not pitch, it is not loudness, and it is not duration. It is everything else.³⁷

³⁵ Robert Cogan, “Toward a Theory of Timbre: Verbal Timbre and Musical Line in Purcell, Sessions, and Stravinsky,” *Perspectives of New Music* 8/1 (Autumn - Winter, 1969): 75-81, here 75.

³⁶ The 1999 American National Standards Institute’s definition of timbre as found in Roy D. Patterson, et al, “Reviewing the Definition of Timbre as it Pertains to the Perception of Speech and Musical Sounds,” In *Neurophysiological Bases of Auditory Perception*, edited by Enrique A. Lopez-Poveda, Alan R. Palmer, and Ray Meddis, 223-233, (New York: Springer, 2010), here 223.

³⁷ Patterson, 223.

Musical timbre has no fixed form of notation. Cogan and Escot (1976) suggest that composers do not, "...notate a tone color," rather they notate, "...the instrumental means by which it is produced."³⁸ This certainly describes the way a vocal tract shape represents the physical means of vowel production. The catch-all nature of this definition allows researchers to follow different lines of thought according to their goals. In linguistics the role of timbre may relate to the creation of meaning in a listener's mind. In music analysis the study of timbre may center on difficult to notate, yet sonically crucial elements of musical structure. In audiology timbre may be an interesting byproduct of the mechanical function of the ear. In artificial intelligence, timbre may be a means to allow computers to identify meaningful patterns in a sound wave.³⁹ In classical singing, timbre may relate to overall brightness, darkness, and ring (essentially aesthetic concerns), as well as effective registration, carrying power, and amplitude (workplace requirements). All of these approaches explore timbre for different reasons.

Perhaps a general theory of vocal tone color does not exist precisely because, as De Poli and Prandoni (2008) note, "unlike other features of musical sounds, such as pitch or loudness, timbre cannot be linked directly to one physical dimension; its perception is the outcome of the presence and of the absence of many different properties of the sound, the perceptual weight of which is still in many ways unclear."⁴⁰ Near limitless by definition, timbre cannot be restricted to prioritize any one point of view. How one chooses to define and engage the idea of timbre changes the questions asked and the relevance of the various associated parameters: attack,

³⁸ Cogan and Escot, *Sonic Design*, 328.

³⁹ The automatic identification of a recording's musical genre has practical applications for streaming services like Spotify and Apple Music. See Kamelia Aryafar and Ali Shokoufandeh, "Music Genre Classification Using Explicit Semantic Analysis," *MIRUM'11* (Nov. 30, 2011): 33-37, for one example of this approach.

⁴⁰ Giovanni De Poli and Paolo Prandoni, "Sonological models for timbre characterization," *Journal of New Music Research* 26/2 (2008): 170-197, here 170.

decay, and release time, distribution of spectral energy, changes in the spectrum over time, presence of noise or harmonics out of the harmonic series of the fundamental, etc. In many ways, classical singing is so simple a timbral proposition that it is not served by any of the many complex analytical methods available.

I believe the terminology underpinning current vocal pedagogy teaching models must invite singers to notice the difference between the aspects of the singing voice that communicate language (as in speech) and those that simply transmit sound (as does an instrument). In the former, the complex spectro-temporal flux caused by coarticulation renders the specific tone color of a vowel less important than its context. However, in the latter, as frequently found in vocalises and melismatic passages, the listener must process the inherent timbre of the voice with few (if any) contextual clues. Terrance M. Nearey (1989) points to these two perceptual processes:

...concerning context effects, isolated vowels are *not* by their nature impoverished stimuli; rather, in many conditions they are well identified. Therefore, extreme theories of cospecification of vowels by consonantal context must be rejected...[meaning that one does not differentiate vowels in speech based solely on the surrounding consonants]. On the other hand, as Strange *et al* (1983) point out, there are *never* any large disadvantages for vowels in consonantal context as might have been expected from some "target" theories ...[meaning that one has a good deal of leeway regarding the specific spectral peaks that will convey a given vowel in context].⁴¹

Essentially, when listening to speech, one hears vowels for their inherent quality and for their phonetic context simultaneously. As pitch rises, especially above the treble staff, and certain spoken phonemes become more physically difficult to resonantly produce (e.g. an [u] or the /k/ or /p/ that precedes the [i] in *keep* or *peep*), a singer gradually loses the ability to engage the latter and leans more heavily on the former. While some singers are physiologically able to

⁴¹ Nearey, 2089.

preserve the vocal tract shape of specific speech level phonemes at higher pitches than other singers, it is both logical and intuitive that there is an upper pitch limit for intelligible speech (see again **Figure 7**). When pursuing a singer-centric model of timbre analysis, the question of inherent quality appears to be the more fruitful of the two paths. Up to this point I have used the terms *timbre* and *tone color* fairly interchangeably. In order to tease out the qualities of timbre most relevant to singing, I would like to draw an important distinction between them.

Building on the work of Fritz Winckle (1967), who believed that the spectral peaks of a sound wave are, “the most significant earmarks of sound...” despite the relevance of other factors to the nature of timbre, I would like to appropriate the term *tone color* to describe a subset of the characteristics currently lumped into the broad definition of timbre.⁴² As Plomp (1966) and others have noted, simple sounds have an inherent color tied to frequency. This means that two simple sounds of an identical frequency, but from different sound sources (perhaps matching harmonics isolated from a trumpet and a recorder), will exhibit the same tone color. Other aspects of timbre related to the mechanical creation of the sound may differ (e.g. attack, decay, release, ongoing variations in intensity or vibrato, and additional noise surrounding the harmonic) but the tone color itself will not (see **Figure 9**). Given a narrow enough frequency band, any moment of sound (with or without pitch) has a specific tone color. If the spectral envelope changes over time, and the relative contributions of the simple sounds change, the overall tone color also changes. Timbre characterizes this change. This may seem like an academic distinction, but it allows for a revision of the problematic definition of timbre. If timbre

⁴² Fritz Winckel, *Music, Sound and Sensation*, trans. Thomas Binkley (New York: Dover, 1967), 24. This line of thought also finds agreement with the use of the term *sound color* in Wayne Slawson, *Sound Color* (Berkeley: University of California Press, 1985), 20, however Slawson quickly pivots to the application of this concept to structural transformations of the total spectral content of a sound.

is everything that differentiates two sounds of the same pitch, loudness, and duration, tone color invites one to consider what two different sounds may have in common. **Figure 10** and **Figure 11** demonstrate that the tone color of one instrument may be imposed onto the remaining timbral elements of another. In **Figure 10** the tone color of a tuba is grafted onto the other timbral characteristics of a bass-baritone. In **Figure 11** the tone color of a bass-baritone is grafted onto a viola da gamba. In both cases the timbral qualities (especially the qualities of attack, decay, and release) of the original instrument shine through the change in tone color. While I do not expect universal acceptance of this definition of *tone color*, I believe it is important to make the conceptual distinction.

What then, to reformulate Nearey's question, is the *unimpoverished nature of an isolated vowel*? Is it the tone color created by the harmonics of the voice, other elements of timbre idiomatic to the voice (such as the attack, decay, and release), or the layers of meaning imbued by coarticulation? It is most certainly a combination of the first two, as the recognizable, but otherwise unsatisfying transformation of the viola da gamba into an [a] in **Figure 11** suggests. However, below the complexity of the timbral elements that differentiate a viola da gamba from a voice lie common aspects of tone color. This is why the viola da gamba may be modified to create a vowel without otherwise losing its sonic identity. **Figure 12** compares the long term average spectrum (LTAS) of a synthesized singing voice and a cello at the same pitch.⁴³ Note the manner in which the voice produces a smooth slope of harmonics, while the cello produces jagged peaks and valleys. A specific harmonic exhibits the same tone color in both voice and

⁴³ From Donald Miller, *Resonance in Singing* (Princeton: Inside View Press, 2008), 116, a long term average spectrum (LTAS), "...accumulates spectral measurements over a specified duration, displaying them lumped together in a single power spectrum." While this reduces the resolution of spectro-temporal flux within the model, it allows one to generally characterize the spectral qualities of a longer sample.

cello, however, the way in which they are combined and fluctuate over time (their timbres) helps distinguish one from the other.

Returning briefly to the vowels in **Figure 8**, **Figure 13 (bottom)** displays them spectrographically. Note that [u] and [i]’s lowest five harmonics are nearly identical. [u] and [ɔ] both have energy in harmonics 6-9 (though at higher amplitude in [ɔ]), that harmonics 13-16 are the bottom of a cluster in [i] and the top of a cluster in [y], and that harmonics 8-10 are similar in [a] and [ʌ] (though at a higher amplitude in [a]). When heard in isolation, the tone color of the spectral segment common to [u] and [i] are basically indistinguishable and sound like [u]. The tone color of the segment common to [u] and [ɔ] is nearly the same (like [ɔ]), but has a higher amplitude in [ɔ] than [u]. Raising the amplitude of this segment in [u] would tip the vowel toward [ɔ]; it is a part of [u], but not the primary tone color. The tone color of the segment in common between [i] and [y] is similar (a harsh sound, like an [i]), but is closer to [e] in the [y]. These peaks are visually similar, but not identical in tone color. The tone color of the common segment in [a] and [ʌ] sounds like the bright edge of [a]. It is present, but quieter in [ʌ]. These examples begin to cast light on the gulf between what a visual model conveys and what a vowel actually sounds like, as defining qualities of certain vowels reside passively within the percepts of others. These examples outline three common relationships between the same spectral segments occurring in different vowels on the same pitch: 1) identical ([u] and [i]), 2) similar in tone color but differing in amplitude ([u] and [ɔ]; [a] and [ʌ]), and 3) similar but differing somewhat in tone color ([i] and [y]). Other possibilities include drastically differing in spectral shape (e.g. a peak in one vowel and a trough in another) or, returning to the voice and cello in **Figure 12**, a slope that is smooth in one and jagged in the other. As soon as pitch changes, the harmonic complexity (the number of harmonics present) of each segment also changes, creating

a new set of internal spectral relationships. Most importantly, I believe these spectral segments each have a specific tone color that can be heard both in isolation and also when the remainder of the spectral envelope is reintroduced. A schematic of this concept is found in **Figure 14**.

These realizations suggest fundamental questions about vowel perception: If a spectral envelope has several peaks, what do those peaks sound like in isolation? Does the timbre of a spectral peak in isolation persist when reintroduced to the overall sound wave? If one spectral peak contains one harmonic and another has several, with what level of detail may one characterize the difference between them? If a singer no longer produces a given spectral peak (due to a change in vowel or pitch), may one characterize the difference in terms of variation in that part, rather than the whole timbre? As explored above, may one reconcile that different vowels may share a common spectral peak but differ profoundly in overall quality? Due in large part to the acknowledged poverty of the vocabulary used to characterize timbre, one is not generally equipped to address these questions. Into this ambiguously labeled space I will offer a multi-layered theory of tone color and vowel perception. The framework that follows cultivates specific language and suggests specific models to explore these questions.

3. Exploring the Special Psychoacoustics of Sung Vowels

Every time we introduce a new tool, it always leads to new and unexpected discoveries, because Nature's imagination is richer than ours.

Freeman Dyson⁴⁴

Having established the need to explore the inherent tone colors of vowels, their individual spectral peaks, and the harmonics that compose these peaks, I will now present and demonstrate an analytical method based on five under-discussed or unlabeled, yet apprehensible principles of perception derived from studying the singing voice. These principles are currently absent in the voice science and vocal pedagogy literature. They are:

1. Absolute spectral tone color
2. The multiple missing fundamentals
3. Local spectral coherence
4. Weak tone color bridging
5. The obvious true fundamental

Together, these principles form a conceptual framework that chips away at the model of the spectral envelope of speech as an event horizon beyond which no meaningful understanding of vowel timbre may be extracted. I caution that while this framework encourages one to become

⁴⁴ Freeman Dyson, *The Scientist as Rebel* (New York: New York Review of Books, 1995), 33, found in Robert Cogan, *Music Seen, Music Heard: a picture book of musical design* (Cambridge: PCI, 1998), 6.

aware of aspects of tone color that I believe to be objectively present in a singer's sound, that such detailed listening is neither obligatory nor necessarily common. The value of this framework likely lies in the objective descriptive language it may bring to those who teach and study the singing voice, an instrument that produces a sound that may be perceived in a variety of ways based on both musical context and the focus of the listener's attention. While this is not the only way to perceive the singing voice, it is both helpful and consistent with less detailed frameworks.

This approach runs counter to the prevailing implementation of acoustic (formant) theory in voice science and vocal pedagogy texts, which engages vowels at the level of the totality of the spectral envelope. Donald Miller (2008) writes,

...consider the... [vocal tract resonance] frequencies of the [i] sung on the pitch A4... The quality of the vowel is unremarkable perceptually, yet the first... [vocal tract resonance], which dominates in the sound through the strong first harmonic, is about 50% higher than the typical speech value.⁴⁵

Although everything Miller writes is correct, it misses a deeper truth. Return to the [u]/[i] pair in **Figure 13 (bottom)** and recall that a single harmonic in the range of the pitch A4 (regardless of the vowel) sounds like [u]. Were it to truly dominate the spectrum, the vowel would change.

Scott McCoy (2012) similarly explores issues surrounding the production of [i] ascending to the pitch D5, "...an octave higher than the typical first...[vocal tract resonance] of the vowel."⁴⁶

McCoy offers a solution based on the idea that tracking the pitch of the lowest vocal tract resonance to match the first harmonic means moving through a series of vowels that, in speech, contain the necessarily adjusted first resonance. Returning to **Figure 2 (b)**, one might map those vowels in order of ascending first vocal tract resonance from either [i] → [I] → [e] → [ε] → [a] or

⁴⁵ Donald Miller, *Resonance in Singing* (Princeton: Inside View Press, 2008), 25.

⁴⁶ Scott McCoy, *Your Voice: An Inside View* (Delaware: Inside View Press, 2012), 46.

[u]→[o]→[ɔ]→[a]. As pitch ascends higher than the frequency of the first vocal tract resonance of a vowel, move along these vowel paths to preserve the intensity boost gained by tuning the first vocal tract resonance close to the first harmonic. This makes a high degree of intuitive sense, and finds agreement with Berton Coffin's Chromatic Vowel Chart, which is based on managing the pitch of the first vocal tract resonance as sung pitch rises or falls.⁴⁷ However, Coffin and McCoy offer heuristic solutions. As pitch rises, I argue that the basic character of what is considered to be a vowel percept fundamentally changes. Our tonal models must similarly change.

These special psychoacoustic principles of singing suggest a novel analytical method based on parsing the spectral envelope of a sung vowel into perceptually coherent segments.⁴⁸ Central to this proposition, I will define and demonstrate a perceptual scale of *absolute spectral tone color* (see below). While significantly informed by Cogan (1998) and Winckel (1967), this scale is based on the absolute, rather than relative tone colors of simple sounds.⁴⁹

Absolute Spectral Tone Color

Definition

absolute spectral tone color: Any two or more simple sounds (e.g. a sine wave, single harmonic of a complex tone, or narrowly notch filtered band of noise) of identical frequency, regardless of their sources, will produce an identical tone color percept independent of other spectral fluctuations considered aspects of timbre. If these simple sounds are

⁴⁷ See Berton Coffin, *Coffin's Sounds of Singing* (Lanham: Scarecrow Press, 1987), and *Overtones of Bel Canto* (Metuchen: Scarecrow Press, 1980).

⁴⁸ Again, here the term *segment* means a notch-filtered band of frequencies that is a contiguous portion of a sound's total spectral envelope.

⁴⁹ See Robert Cogan, *Music Seen, Music Heard: a picture book of musical design*. (Cambridge: Publication Contact International, 1998), 110, for a history of this line of thought, and Fritz Winckel, *Music, Sound and Sensation: A Modern Exposition*, trans. by Thomas Binkley (New York: Dover, 1967), 14, for additional information.

located within a complex sound, their inherent absolute spectral tone color is never lost or changed, only expressed or masked. These tone colors may be placed on a continuum, and bear a meaningful similarity to several vowels.

Discussion

If one describes the tone color of simple sounds along a continuum of brightness (as Plomp (1966) suggests is accepted in psychoacoustics),⁵⁰ the principle of absolute spectral tone color requires that we label any two simple sounds of the same frequency with an identical brightness value. Similarly, if we adopt Cogan's register-based system and divide timbre into grave, neutral, and acute regions, two simple sounds of the same frequency exhibit exactly the same quality of graveness, neutrality, or acuteness.⁵¹ It matters less that we use a specific scale than that we make the conceptual leap that there are absolute values along the continuum. In speech, the nuanced differences between near similar absolute spectral tone color values (reflected in the ranges of spectral peaks possible for the same vowel over a population of speakers)⁵² obviate the need for fine gradations. Driven by questions specific to the higher pitch ranges inhabited by singers, the scale I propose here goes into greater detail and makes more specific claims than previous efforts. See **Figure 15** for my scale based on the closest vowel-like quality and further tested while exploring the principles of *local spectral coherence* and *weak tone color bridging*. I must emphasize that these simple sounds are only *similar* to the defining quality of the chosen vowel; they are *vowel-like*. A vowel is a complex combination of many tone colors. Simple sounds, by definition, are not capable of fully representing the vowels they may play a strong role in defining. Additionally, though ranges of simple sounds elicit a similar

⁵⁰ Plomp, *Experiments*, 132.

⁵¹ Cogan, *Music Seen Music Heard*, 110.

⁵² Gordon E. Peterson and Harold L. Barney, "Control Methods Used in a Study of the Vowels," *Journal of the Acoustical Society of America* 24/175 (1952): 175-184, here 182.

vowel-like percept, even the tone color within a single vowel-like range lies on a continuous spectrum. Thus the tone color at the border between two different ranges may be perceptually fuzzy and exhibit qualities of both, much like the borders between spectral colors in a rainbow. The shorthand I use is the tilde ‘~’ plus the letter corresponding to the IPA symbol for the appropriate vowel. For example, a simple sound with a frequency of the pitch D4 is ~u. This means it is *like* the defining tone color of [u].⁵³

I also acknowledge two biases that permeate this work. The first is anthropomorphic in that I assign values preferentially based on the way that humans voice sound. Meaningful variations of tone color certainly exist below the frequency range I have labeled ~u and above the range labeled ~bright i. For physiological reasons, the human voice does not create strong spectral peaks below and above those ranges, so less perceptual awareness of variation exists for those tone colors. I am sure that subtle gradations exist within ~u and ~bright i that are simply irrelevant to both this inquiry, and also human experience more generally. The second is my bias toward the sounds of English, which certainly informs the order in which I have attempted to study vowels. My use of IPA, that groups sounds common to many languages, hopefully alleviates most of this. Even if a non-English speaker prefers slightly different borders for these IPA symbols, I am confident that they will consistently do so.

The vowel-like tone colors presented in **Figure 15** likely represent a class of vowels (that I tentatively term ‘spectral vowels’) characterized by a strong peak in a dominant tone color. Notably absent are some common English vowels such as [ʌ] (this and other variations of /ə/ appear to arise when multiple tone colors occur simultaneously at more subdued amplitudes than

⁵³ I hope linguists will accept the visually simpler language of ~u rather than ~[u]. I am hesitant to use the IPA brackets in this context for fear of implying that ~u expresses something more than the most narrowly vowel-defining portion of the full spectrum of [u].

found in the spectral vowels—this allows the neutral background harmonics to come closer to the foreground) and [ɜ], and common German and French vowels like [y], [Y], [ø], and [œ] (that appear to arise when the third vocal tract resonance lowers to give a boost to—and effectively widen the area of effect of—the second spectral peak). Also missing are continuants (sustained, voiced consonants like [l], [m], [n], and [ŋ] (like “sing”)), sounds characterized almost exclusively by the buzzy sound of what would be background harmonics in a true vowel. Further research is warranted.

The Multiple Missing Fundamentals

Definition

the multiple missing fundamentals: Provided it contains sufficient harmonics to trigger the missing fundamental phenomenon, each spectral peak of a sung vowel gives rise to a separate missing fundamental characterized by the absolute spectral tone colors of its constituent harmonics. Therefore, multiple, separately tone-colored missing fundamentals may coexist with the true fundamental in the perceptual space of the pitch.

Discussion

The missing fundamental phenomenon has been well documented since before Helmholtz, who called them *differential tones*.⁵⁴ A listener will hear the pitch of a complex periodic wave stripped of its first harmonic, the only harmonic with the actual frequency of the pitch (see **Figure 16 (top)**). This is, in most cases, a purely cognitive experience; the actual fundamental does not exist in the sound wave. This is such a dependable phenomenon that the analog telephone system was designed to transmit no information below about 300Hz (around

⁵⁴ Hermann L. F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, Fourth edition (1877), translated by Alexander J. Ellis (New York: Longmans, Green, and Co., 1912), 153.

E ♭4), well above the speaking pitch for most males and some females. Indeed, very few harmonics from a periodic wave are needed to elicit this phenomenon, provided they are contiguous and from the same harmonic series (see **Video Example 3.1**), though the fewer the number of harmonics, and the higher within the series they fall, the weaker the missing fundamental. Citing Houtsma and Smurzynski (1990), Plomp (2002) declares that:

...the pitch problem can be considered as settled. That is, the pitch of tones occurring in music and speech is primarily determined by the lower harmonics resolved by the ear ...[the lowest eight harmonics according to Norman-Haignere, et al (2013)]. The periodicity of the unresolved higher harmonics may also contribute, but to a lesser extent.⁵⁵

The lowest eight harmonics span a pitch range of three octaves. Thus below the pitch C2, all eight fall exclusively within the ~u range; however, within the usable pitch range of most singers, these resolvable harmonics cover a wide range of absolute spectral tone colors.

As the fundamental has a fixed absolute spectral tone color based on frequency, the pitch of almost every possible vowel is actually an aggregate of that true fundamental and a group of distinctly tone-colored missing fundamentals. The sample in **Figure 16 (top)** ought not be thought of as the fundamental plus ‘everything else’ that remains when the fundamental is removed. The ‘everything else’ is comprised of multiple sub units, each capable of producing a missing fundamental with a unique, immutable tone color. This is profoundly strange, and forces one to reconsider whether a spectrogram conveys information in a perceptually relevant manner at all. **Figure 16 (bottom)** is a schematic that illustrates this perceptual phenomenon.

According to D. Robert Ladd, et al (2013), while recent experimental evidence suggests that only a portion of the population actually hears the missing fundamental (implying that those who do not instead hear spectrally—they hear the frequencies of the actual harmonics present

⁵⁵ Plomp, *The Intelligent Ear*, 28.

instead), methodological issues may have contributed to such conclusions.⁵⁶ Ladd, et al suggest instead that most people are able to switch between these listening modes depending on a variety of circumstances, which better aligns with my empirical evidence: most listeners notice the missing fundamental when you teach them what to listen for.⁵⁷ The research reviewed by Ladd does not appear to address the question of how the tone color of a cluster of harmonics does or does not change when that cluster is reintegrated into the complete periodic wave.

I argue that the missing fundamental occurs separately for each spectral peak of the voice, provided that peak contains sufficient harmonics to trigger such a percept (usually a minimum of two, but this is somewhat dependent on frequency range and position within the harmonic series). Additionally, contiguous attenuated harmonics (appearing as troughs in the spectral envelope) also contribute a neutral, buzzing missing fundamental. The important concept to consider is that each of these missing fundamentals has a tone color related to the absolute spectral tone colors of its constituent harmonics (e.g. a spectral peak in the ~u absolute spectral tone color frequency range will never express the inherent tone color of ~i).⁵⁸ This suggests that multiple missing fundamentals, separable not by their pitch, but rather by their tone color, coexist in the perceptual space of a vowel's fundamental pitch. This is another way to say that the pitch itself has multiple aspects, perceptually separable by tone color. One of the dangers of using spectrographic models is the potential to assume that the intensity of the displayed first harmonic is the same as the intensity of the perceived pitch, and that independently perceptible harmonics are represented by the intensity of the displayed higher harmonics. For the most part,

⁵⁶ D. Robert Ladd, et al, "Patterns of individual differences in the perception of missing-fundamental tones," *Journal of Experimental Psychology: Human Perception and Performance* 39/5 (Oct 2013): 1386-1397, here 1386.

⁵⁷ For a review of previous missing fundamental studies, and new research suggesting the fluidity of *missing fundamental* versus *spectral hearing* modes, see Ladd, et al, "Patterns of individual differences in the perception of missing-fundamental tones."

⁵⁸ This will be further explored through the principles of local spectral coherence and weak tone color bridging.

the harmonics of a voice displayed on a spectrogram are *literally aspects of the pitch*. In some cases, one may also notice a harmonic as a distinctly heard simple sound separate from the pitch. Overtone singers cultivate this skill. Vocal registration and vowel modification may be characterized by both the manner in which these missing fundamentals come to dominate our attention for a given pitch/vowel combination, and also the way in which they change, combine, or disappear entirely as pitch changes and the harmonic complexity of each spectral peak is altered.

Local Spectral Coherence

Definition

local spectral coherence: A given spectral peak of a periodic sound wave may contain harmonics that represent more than one absolute spectral tone color range, but the spectral peak will come to be defined by the absolute spectral tone color of the amplitude-weighted average frequency (spectral centroid) of its harmonic(s). In this manner, each of the multiple missing fundamentals contributes a single sub-tone color to the vowel's composite tone color.

Discussion

Closely linked to the multiple missing fundamentals, local spectral coherence helps to explain the perceptual process connecting the absolute spectral tone colors of harmonics and the tone color(s) of a complex sound. E.g. if we consider harmonics 2-4 of the pitch A3 (frequencies equal to the fundamentals of the pitches A4, E5, and A5) within this framework, they will elicit the respective absolute spectral tone colors of ~u, ~o, and ~ɔ in isolation. The amplitudes of these three harmonics may be individually manipulated to produce a missing fundamental at A3 that sounds similar to each of these three tone colors (see **Figure 17**). I label this phenomenon

with a ‘less than’ sign (<) followed by the IPA symbol of the absolute spectral tone color corresponding to the amplitude-weighted average frequency of the group of harmonics.⁵⁹ These tone colors are still *less than* the total tone color of the vowels they emulate, but more complex than a simple sound. Again, these < percepts exist primarily as missing fundamentals located in the perceptual space of the pitch.

One may not simply isolate any set of contiguous harmonics and expect to find the tone color of that spectral segment strongly present in the overall vowel percept; a few conditions appear vital. First, with the exception of the first harmonic, there must be a sufficient number of harmonics preset to elicit a missing fundamental. If not, the segment boundaries must expand; at times this means encompassing more than one absolute spectral tone color range. Second, the slope of the segment must generally be bound or bordered by harmonics of lower intensity than the segment’s peak. This means that at times the first and second spectral peaks of a vowel like [a] will create separate spectral segments, each with its own < percept and missing fundamental. At other times, if there is no harmonic of lesser amplitude between them, and especially if each peak contains one or no harmonics, what would have been the first and second spectral peak at a lower pitch perceptually coheres into a single < percept. As mentioned above, this also means that a series of harmonics of relatively similar amplitudes will likely cohere into a single < percept, as is found in the buzz of the vocal fold source sound (see again **Figure 1**) or the background troughs of attenuated harmonics that lie between spectral peaks.

⁵⁹ As with the use of the ~ to label absolute spectral tone color, I hope linguists will accept the visually simpler language of <u rather than <[u].

The idea of local spectral coherence finds significant agreement with the “subband spectral centroid based features,” explored by Phu Ngoc Le, et al (2011).⁶⁰ Their approach measures the average amplitude and frequency center for a band of harmonics, though they do not characterize tone color, or dynamically adjust their bands to follow the spectral peaks as vowels change. One may imagine that in **Figure 17**, parts **(b)**, **(c)**, and **(d)**, the tone color of the percept of the three harmonics tips toward the absolute spectral tone color of the highest amplitude harmonic because the amplitude-weighted average frequency of those three harmonics skews toward the highest amplitude harmonic. And this is true. However, while those same three harmonics in **Figure 17 (a)** form a different envelope shape (a subtle decreasing slope from harmonic 2 through 4) than in **Figure 17 (c)**, they also sound < o; albeit a different quality than found in **Figure 17 (c)**, but < o nonetheless. Again though, in both **Figure 17 (a)** and **(c)** the average central frequency based on the amplitude of each harmonic falls toward the center, which here (probably around the harmonic at E5) has the absolute spectral tone color ~o.⁶¹ In this way, slightly differently-shaped spectral segment slopes may elicit a similar < percept.

Other recent research also supports the idea that each spectral peak of a vowel contributes a different tone color. Jennifer Bizley, et al (2013), study vowel discrimination in a population of trained ferrets. Based on the idea that,

...human listeners rely on the **relationship** between...[spectral peak] frequencies in order to correctly identify vowels, ...[s]ince the ferrets in this study were required to discriminate between two vowels only, they could potentially base

⁶⁰ Phu Ngoc Le, et al, “Investigation of spectral centroid features for cognitive load classification,” *Speech Communication* 53/4 (April 2011): 540-51, here 542.

⁶¹ Current experimental data suggest that higher frequency harmonics within a slope have a subtly disproportionate effect on the < percept at the same amplitude, though this is not uniformly true. In certain ranges, this may be consistent with the effect of the equal loudness curve, a measure of the ear’s sensitivity based on frequency range. Further research is warranted.

their judgments on the frequency of either the first or second...[spectral peak] or a combination of the two.⁶² (emphasis added)

However, the ferrets responded in an unexpected way:

While we cannot be sure what cues the ferrets used when discriminating the full four-...[spectral peak] vowels, these observations suggest that the ferrets can identify these two vowels using just the location of the first... [spectral peak] of /u/ and the second... [spectral peak] of /ε/...⁶³

As discussed in **Figure 13**, the first spectral peak (Bizley uses the term *formant*) of [u] is the segment of the spectrum with the tone color that characterizes [u] (<u). It is logical that her ferrets were able to identify [u] based on that peak, as the second spectral peak sounds <ɔ. [ε], however, relies heavily on the tone color of the second spectral peak, which the ferrets needed for proper identification. **Figure 19** explores this idea in greater detail. The isolated vowel [ε] is followed by the vowel in the context of the words *bed*, *fed*, *said*, and *red*. When only the second spectral peak is presented, the [ε] sounds like [ε], albeit missing other elements of tone color usually present. The words sound similarly impoverished, but are understandable. When only the first spectral peak is presented, the isolated [ε] vowel sounds <o. Again, impoverished, but clearly not [ε]. The words are still understandable and sound like [o] and a muted [ε] simultaneously. This is the strange effect of coarticulation layered onto the perception of tone color. Wonderfully, the ferrets were able to recognize the intrinsic [ε]-like tone color (<ε) of the second spectral peak and the *not* [ε] quality (instead <o) of the first spectral peak.⁶⁴

⁶² Jennifer Bizley, Kerry M. M. Walker, Andrew J. King, and Jan W. H. Schnupp, "Spectral timbre perception in ferrets: Discrimination of artificial vowels under different listening conditions," *Journal of the Acoustical Society of America* 33/1, (January 2013): 365-76, here 368. In this quote, *formant* aligns with my use of *spectral peaks*.

⁶³ Bizley, 370.

⁶⁴ Dr. Bizley was generous enough to share her "single spectral peak" audio samples, synthesized in Matlab, with the author. They are a full spectrum of harmonics, filtered through a single virtual vocal tract resonance. In contrast to my examples here, her samples also featured low amplitude harmonics above and below her single spectral peaks. The effect, however, is a sound dominated by the tone color of the single peak.

The principle of local spectral coherence may be used to explore the frequency boundaries of each range of absolute spectral tone color. By synthesizing a segment of the spectrum, one may shape it to elicit the < percept of the spectral centroid (see **Video Example 3.2**), and then change the < tone color by reshaping the spectrum. As the fundamental of the series of harmonics rises, one may notice the < percept change as the amplitude-weighted average frequency (spectral centroid) crosses absolute spectral tone color boundaries.

Finally, local spectral coherence offers a way to explain the harmonic complexity (literally the number of harmonics) of one spectral peak versus another. As pitch rises, the number of harmonics occupying a given absolute spectral tone color band necessarily diminish. This means that the character of the missing fundamental triggered by those harmonics simplifies. As pitch falls, not only do the number of harmonics in a given spectral peak increase, more of them fall into what is known as the critical band of hearing.⁶⁵ This is a frequency band of about a minor third within which the ear cannot well resolve the individual frequencies of two simple sounds. All harmonics ascending from the fifth harmonic fall within the critical band of a neighbor. If multiple harmonics fall within a critical band, and especially if they are of near equally high amplitude, one perceives *roughness*, “the buzzing, rattling auditory sensation accompanying narrow harmonic intervals,”⁶⁶ in the tone. All harmonics but the first five of a baritone’s A3 fall within the critical band of a neighbor. Depending on how intensely the baritone sings, there is much potential for complex and buzzy roughness. Contrast this with a soprano singing a high F6 (1396.9 Hz). Her first five harmonics reach to nearly 7,000 Hz, well

⁶⁵ Johan Sundberg, *The Science of the Singing Voice* (Dekalb: Northern Illinois University Press, 1987), 108.

⁶⁶ See the abstract to Pantelis N. Vassilakis and Roger A. Kendall, “Psychoacoustic and cognitive aspects of auditory roughness: definitions, models, and applications,” *Proc. SPIE 7527*, Human Vision and Electronic Imaging XV, 75270O (February 17, 2010): doi:10.1117/12.845457.

above the frequency range generally energized by classical female voices. She will likely have no rough, buzzy complexity at all. These local spectral segments are characterized by at least three important parameters: < tone color (the absolute spectral tone color of the average amplitude-weighted frequency), complexity (number of harmonics), and roughness (number of harmonics within a critical band of its neighbors and their relative amplitudes). As alluded to above, the resolvability of the harmonics must also play a role in characterizing the sound of a spectral peak, though further research is warranted.

Weak Tone Color Bridging

Definition

weak tone color bridging: According to the principle of absolute spectral tone color, certain vowel sounds ought be impossible if the pitch places the first and second harmonics (separated by an octave) or the second and third harmonics (separated by a fifth) below and above the absolute spectral tone color range of the target vowel (see **Figure 15** for these ranges of absolute spectral tone color). A conceptual extension of local spectral coherence, which suggests a spectral segment will have the tone color of its amplitude-weighted central frequency (spectral centroid), one may perceive the bridged tone color (where no harmonic is present), but it will be qualitatively weaker in comparison to a pitch/vowel combination that places a harmonic within the bridged tone color's frequency range (see **Video Example 3.4**).

Discussion

Assuming equal amplitude, the spectral tone color center (spectral centroid) of the first and second harmonics is roughly a fifth above the first harmonic. The spectral tone color center of the second and third harmonics is roughly a major third above the second harmonic. E.g. if the

first harmonic is A4 (440 Hz), the second harmonic is A5 (880 Hz), and both harmonics are of equal amplitude, the spectral tone color center will be E5 (660 Hz). In this example, A4 sounds ~u, A5 sounds ~ɔ, and E5's tone color (~o), though not present, will be weakly expressed.

Increasing the amplitude of the upper or lower harmonic will shift the spectral centroid toward its absolute spectral tone color instead (either ~u or ~ɔ). In theory this principle applies to pairs of higher harmonics, but as the intervals become smaller, so do the odds of bridging an entire absolute spectral tone color.⁶⁷ A weakly bridged tone color will be notated with a superscript or parenthetical "wb," e.g. <a^{wb} or <a(wb).

The Obvious True Fundamental

Definition

the obvious true fundamental: Under certain circumstances, the first harmonic (the only harmonic with the frequency of the pitch) may be easily perceived as distinct in tone color from the rest of the spectrum.

Discussion

The first harmonic occupies a unique place within the harmonic series. It is the only harmonic with the frequency of the pitch, it will almost never fall within the critical band of another harmonic (thus it will almost never be associated with roughness),⁶⁸ and ascending from approximately C#4, it will always be the only harmonic within its absolute spectral tone color range. Though at low pitches it is often grouped into the <u percept of its local spectral segment, these unique characteristics mean that the first harmonic may be simultaneously perceived as a

⁶⁷ This may help to explain Bozeman's "pitch of turning," a passive vowel modification that takes place as the second harmonic (and to a lesser degree higher harmonics as well) crosses the first formant. What was a strong first spectral peak becomes a weakly bridged tone color. For more, see Bozeman, *Practical Vocal Acoustics*, 26.

⁶⁸ Below about 100 Hz (approximately G2) the critical band is wider than the frequency difference of the first and second harmonics. Above that pitch, the first harmonic will never fall into the critical band of another harmonic. See Sundberg, *The Science of the Singing Voice*, 108 for a discussion of the critical band of hearing.

part of and distinct from the total percept of the spectrum. No other harmonic exhibits this quality so dependably.

Because classical female singers (and countertenors) often shape their vocal tract to align their lowest vocal tract resonance to couple with the first harmonic as pitch ascends (a resonance strategy called *whoop* or *hoot*) this harmonic is often significantly louder than the rest of the spectrum.⁶⁹ As Richard Miller describes, “...among prominent female artists, when they are singing in [the] upper range, the first... [spectral peak] and the fundamental are often enhanced and exhibit increased acoustic energy in the lower portion of the spectrum.”⁷⁰ This causes a portion of the total vowel percept to be significantly defined not by the < tone colors of several spectral segments simultaneously (as found in the lower pitch range of the voice), but rather by the simple absolute spectral tone color of the obvious true fundamental. Additionally, on and above the treble staff, the vibrato-associated amplitude fluctuations often found in higher harmonics appear to be less present in the first harmonic. This means that certain technical competencies (such as rate and extent of vibrato, and intonation) may be best aurally assessed via the obvious true fundamental. Further research is warranted.

As pitch rises, the absolute spectral tone color of the first harmonic will predictably pass through several tone color ranges (see again **Figure 15**). The common range of the female and countertenor classical singing voice encompasses the absolute spectral tone colors of ~u, ~o, ~ɔ, ~ɑ, and (at the upper extreme) ~a. Traditional male voice types sing with their fundamental almost exclusively in the ~u range. This means that as pitch rises in a voice trained to *whoop*, the vowel percept will come to be defined in part by the absolute spectral tone color of the first

⁶⁹ See Bozeman, *Practical Vocal Acoustics*, 23, and Donald Miller, *Resonance in Singing*, 52.

⁷⁰ Miller, Richard. *Solutions for Singers: Tools for Performers and Teachers*. Oxford: Oxford University Press, 2004, 75.

harmonic, with significant tone color changes in the fundamental from ~u to ~o around C5/D5, ~o to ~ɔ around F#5/G5, ~ɔ to ~ɑ around B5/C#6, and to ~a by F#6. These changes in tone color align with the transitions found in a well-balanced classical female singing voice; since the absolute spectral tone color of the first harmonic at the same pitch produced by two different singers will be the same, this explains the near uniform aspects of good registration across all singers. Richard Miller (2000) places a soprano's acoustic registration transitions at approximately C#5, F#5, and C#6.⁷¹ Miller gives mezzo sopranos and contraltos slightly lower pitch points, implying that some aspect of this registration is physiological; i.e. for an equal amount of physical ease, the pitch range of the first vocal tract resonance may lie lower for lower voice types, thus making specific resonance tuning transitions easier at slightly lower pitches. However, the similarity of these pitches to the obligatory shifts in the absolute spectral tone color of the first harmonic (independent of vowel) may mean that these registration points have a significant psychoacoustic component as well. This is to say that the expectation of how a vowel percept will change as pitch rises may also need to align with perceptual, rather than strictly physical limitations.

⁷¹ Richard Miller, *Training Soprano Voices* (New York: Oxford University Press, 2000), 117.

4. Analysis

...let us explore freely in this vast realm, not uncritically, but with an open eye and ear for the unexpected connection, the so far unrecognized pattern, the previously unfelt flash of sensation and expression.

Robert Cogan⁷²

Models of vowels exclude information by design, presenting a narrow view of just a few of the many measurable parameters. I believe meaningful revisions to such models must capture how one perceives sound, not simply display factual information about a sound wave or vocal tract shape. In the short term, I encourage those in the singing community who use our current models to add the scale of absolute spectral tone color to the appropriate frequency axes of the models in **Figure 2**.⁷³ This will quickly allow students to notice that the first spectral peak has a dependable < percept that is in most cases different from the primary tone color of the vowel; [u] is a notable exception. See **Figure 20** for an example of such a revision that helps to explain the change in tone color of the soprano voice above the pitch G5. Note that Vennard's chart indicates that above the pitch G5, only those vowels that fall outside of (toward the upper right

⁷² Robert Cogan, *The Sounds of Song: a picture book of music for voice* (Cambridge: PCI, 1999), 12.

⁷³ Perhaps someone will offer a three dimensional revision of the vowel quadrilateral (**Figure 3 (e)**) to indicate the anticipated vowel for the same vocal tract shape as pitch rises.

corner of the chart) the thick, shaded bands may be sung. Consider though, that in this range a soprano's first harmonic may be the highest amplitude part of the spectrum. This means that for a given pitch in this range, the vocal tract shape of the vowel on the chart would be strongly characterized by the absolute spectral tone color of the value indicated on the horizontal axis (that of the first harmonic).

What this simple addition misses, however, is a method to display the complexity of (number of harmonics constituting) each spectral peak and the relationships between the several spectral peaks (the internal relationships of tone color within a vowel). For example, this revised chart (see again **Figure 20**) suggests that the neutral [ʌ] vowel (like “up”) is a combination of a first spectral peak that sounds either <u or <o and a second that sounds <a. Without a third axis of amplitude, this fails to elegantly capture the difference between [ɑ] and [ʌ], despite the apparent qualities they share (notably the frequency of the second spectral peak). [ɑ] is meaningfully characterized as a strong first spectral peak that sounds <ɔ and a strong second peak that sounds <a (and defines the vowel). One may imagine these tone colors as the *depth* and *clarity* of [ɑ] respectively.⁷⁴ [ʌ] arises when lower intensity <u/<o and <a peaks sound simultaneously. As Bozeman (2013) suggests, as two vocal tract resonances approach one another in frequency they “mutually reinforce” one another, raising the amplitudes of both resulting spectral peaks.⁷⁵ As the first vocal tract resonance of [ɑ] lowers (relative to the second resonance) to create [ʌ], both spectral peaks (especially the second) diminish in amplitude, and the character of the overall percept becomes more neutral. So our model must also accommodate that these vowel-like building blocks of tone color sometimes combine in a manner that doesn't

⁷⁴ This aligns well with Bozeman's characterization of the first and second spectral peaks of vowels, Bozeman, *Practical Vocal Acoustics*, 13-16.

⁷⁵ Bozeman, *Practical Vocal Acoustics*, 16.

strongly express the < tone color of either the first or second spectral peak. Depending on the pitch and vowel, each of these < percepts may be rich and complex (many harmonics in each spectral peak), simple and pure (few or one harmonic), or somewhere between. Importantly, as pitch and vowel change, each spectral peak's < percept may change qualities independent of the remaining spectral peaks. In the analyses that follow, I must direct attention to an obvious blind spot within my own models: the use of long term average spectrums. While they limit one's awareness of subtle spectro-temporal shifts present in the voice (indeed, something as ubiquitous as vibrato ought be understood in terms of the subtle change in each spectral peak's < percept as pitch rises and falls and the relative amplitude of each harmonic non-uniformly changes), I hope to consider this simplification a starting point for exploring deeper issues, rather than a limitation.

Parsing the Spectral Envelope: A Model for Aurally Locating Vocal Tone Colors

To begin to explore a logical visual language that captures the principles introduced in chapter 3, first consider an analysis of a single [a] sung by a baritone on the pitches B2 and B3, and a synthesized B4 based on the same vocal tract resonances. In **Figure 21 (a), (b), and (e)**, notice that the complexity of (number of harmonics present within the same frequency range) the entire spectral envelope decreases as pitch rises. Almost the entirety of the first two spectral peaks of the B2 **(c)** fall within the critical band (indicated by the gray box), lending a buzzy quality to the segment of the spectrum that sounds most like the vowel. At the pitch B3 **(d)**, most of the remaining harmonics in this range fall outside of the critical band, further reducing the complexity. An octave higher at B4 **(e)**, the (synthesized) sample has only three harmonics remaining in the entire frequency range occupied by two uniquely tone colored, buzzy spectral peaks at the pitch B2. Instead of a missing fundamental from each of two spectral peaks in this

range, one finds only the purity of the obvious true fundamental plus a single potential missing fundamental from the second and third harmonics.

The way in which this vowel passively modifies as pitch rises cannot adequately be characterized as a shift from one vowel to another.⁷⁶ The perceptual differences between these three samples have to do with the complexity and character of the spectral peaks themselves, not simply their frequency centers or the associated vocal tract shape. The vocal tract resonances that produce an [a] at B2 produce something qualitatively different at B4. One could argue that the vowel percept of B4 (and certainly at even higher pitches) falls short of presenting the basic elements of a speech level vowel: a warm quality of the first spectral peak, a clear quality of the second spectral peak, and (at least for spoken pitches below about C5) a warm ~u or <u *below*, and a bright <i (that becomes the singer's formant in singing) *above* the tone colors of the vowel. Given that a sung vowel at a high pitch loses many of these basic qualities, perhaps such singing should be objectively analyzed based on its inherent tone colors rather than in the context of those expected from speech. Since much of classical voice training aims to balance registration such that the singer may negotiate a large range of pitches with a consistently low level of effort, a discussion of the way in which these basic elements change, simplify, and disappear gradually as pitch ascends is both relevant and helpful. To visualize this, first see **Figure 22**. This is the B2 pitch from **Figure 21** as a traditional graph of the spectrum (top) and broken into individual spectral peaks with separate < percepts (bottom). Each one of these spectral peak segments creates a missing fundamental that adds its < tone color to the aggregate tone color of the pitch. The segment of harmonics labeled *background* sounds buzzy and indistinct, like the unfiltered sound of the vocal folds.

⁷⁶ As discussed above, passive vowel modification refers to shifts in the vowel that result from changing the sung pitch without changing the frequencies of the vocal tract resonances.

Figure 23 aligns the B2, B3, and B4 from **Figure 21** as broken spectrum graphs. Follow any one spectral peak from low to high pitch and notice how it simplifies. The <u present in the B2, for example, reduces to a ~u (single harmonic) at the B4. The buzzy background harmonics in the B2 shrink in the B3 and disappear entirely in the B4. The <o and <a present as complex separate peaks (four and five harmonics respectively) predominantly within the critical band reduce in complexity as pitch rises until they merge into one <a completely outside of the critical band.

I would like to further clarify this model by discarding the dimension of frequency, a radical departure from one of the core assumptions inherent to spectrographic models: that the frequency of each harmonic is relevant to our perception of the vowel. It is, of course, but in context I believe one generally perceives the tone color, not the frequency of that harmonic.⁷⁷ If harmonics of the voice are primarily experienced as tone color aspects of the pitch (overtone singing is a notable exception), our model should compress these various ~, <, and <^{wb} percepts into the pitch space of the fundamental while representing aspects of their complexity, tone color, and amplitude. If one seeks to derive information relevant to resonant singing technique from the qualities present in a singer's vowels, the pitch is where to aurally locate that information.

Figure 24 demonstrates the ~ and < percepts from **Figure 23** aligned on a three dimensional graph. The vertical axis is amplitude, horizontal axis the harmonic complexity, and depth axis is tone color (which becomes shallower for the B4). Note that this model captures not only the number of tone color areas present in each sample (depth axis), but also the harmonic

⁷⁷ Yes, depending on style, pitch, and registration higher harmonics may rise to the listener's awareness; however, within the classical singing idiom, I generally argue that this is in addition to, rather than instead of the role of the harmonic as a part of a missing fundamental.

complexity of each spectral segment (horizontal axis), the relative prominence of each segment (vertical axis), and the differences in tone color grouping. For example, the fundamental becomes a perceptually separable simple sound with an absolute spectral tone color between the B2 and B3 (though because of its role in the spectral slope, it also coheres on B3 with the <0 segment), and the <0 segment is absorbed into the <a segment between the B3 and B4.

Figure 25 graphs the parsed spectra of a baritone's [a] and [i] on the same pitch (B2). The depth of tone color for [i] is less than [a], the harmonic complexity of the second and third spectral peaks of [i] is significantly wider than for any one peak of [a] (requiring additional data points along the horizontal axis), the background harmonics in [i] are both more complex and of lower amplitude than those in [a]. Finally, the <u percept, a warm tone color beneath the more relevant <0 and <a in [a] is a higher amplitude, more complex spectral peak in [i].

These two vowels allow one to consider the way in which the local spectral coherence effect perceptually glues harmonics (which have absolute spectral tone colors) together into < percepts. The <u segment of [a] in **Figure 25** contains three harmonics (with frequencies corresponding to the fundamentals of the pitches B2, B3, and F#4), each within the absolute spectral tone color of ~u. The fourth harmonic (B4) lies on the ~u side of the transition to ~o. The fifth harmonic (D#5) sounds ~o. The fourth and fifth harmonics are pulled into the <0 of [a], but the <u of [i]. They retain their absolute spectral tone colors in both vowels, but function as a part of two different < tone colors because of the shapes of their local spectral peaks. These are differences that can be heard when one draws his or her attention to them. The parsed spectrum graphs invite us to locate these < percepts as aspects of pitch, separable by tone color, amplitude, and complexity.

Methodologically, defining the spectral segment boundaries in the previous figures requires some trial and error. The process involves isolating the central frequencies of a peak and adding harmonics individually on either side until either an additional < percept emerges, or the addition of further harmonics ceases to change the < tone color of the segment (likely relegating those harmonics to a background segment). Returning to the concept of the subband spectral centroid, each time a high enough amplitude harmonic is added in this manner, the spectral centroid of that segment shifts. For example, in **Figure 24 (bottom)** the highest amplitude peak of the < i actually lies in the ~e range. However, the high amplitude harmonics to the right of that peak fall into the ~i range, raising the spectral centroid of that segment to < i, despite the strong ~e elements. This raises interesting questions about the utility of selecting the highest amplitude frequency of a spectral peak to represent a vowel on an x/y plot graph. Yet to be determined, and as far as I can tell currently absent from the psychoacoustics literature, are exact measurements for the threshold characteristics of these segment borders.

The Classical Female (and Countertenor) Voice above the Treble Staff

A similar graphical analysis of the classical female (and arguably countertenor) singing voice on and above the treble staff is interesting not for its complexity, but rather for its simplicity. Gone are the large number of harmonically complex spectral peaks found in the previous figures. Instead, one perceives a sound increasingly defined by the obvious true fundamental and additionally colored by a single (or perhaps two) contiguous spectral segment(s) spanning multiple absolute spectral tone color regions, likely devoid of vowel defining harmonics within the critical band. Depending on the singer and tonal model (bright and edgy, warm, muffled, etc...), the total percept may even consist of just the obvious true fundamental and a single, indistinct background of higher frequency harmonics. This means that

the number of missing fundamentals comprising the pitch literally decreases, and those that remain have a diluted < percept increasingly characterized by the weak timbre bridging principle. Additionally, few truly resonant pitch and vowel combinations remain, especially for vowels with lower second spectral peaks (e.g. [u], [o], [ɔ], [ɑ], and [a]), necessitating active vocal tract modifications that result in a weakly-defined sound that likely fails, I would argue, to meet the threshold of complexity found in spoken vowels. Consider, for example, the resonant vowels available at the pitch G5. The first harmonic sounds ~ɔ, so [u] and [o] are not possible. The second harmonic already sounds ~æ or ε, so resonant versions of [ɑ], and [a] are not possible. Almost all vowels sung on this pitch will actually be dominated by the ~ tone color of the first harmonic, perhaps characterized by a weak, bridged tone color defined by the amplitude relationship of the first and second harmonics, and additionally colored by an indistinct cluster of higher harmonics. Through contextual phonetic clues, one may imply a wide range of vowels, but the singer's tonal model must accommodate that the objectively producible sounds are rather limited.

Consider the pitches F5, G5, and A b5 from m. 35-36 of Renée Fleming's (1997) recording of Mozart's "Porgi amor" from *Le Nozze di Figaro*. Here Ms. Fleming chooses a resonance strategy that aligns the first vocal tract resonance with the first harmonic (see **Figure 26**), and brings the greatest intensity to the lower end of the spectrum, causing the absolute spectral tone color (F5=~o, G5=~ɔ, A b5=~ɔ) of the first harmonic to dominate the overall percept. The remaining harmonics create a < percept that sounds like variations of <i. However, these harmonics span tone colors from ~æ/ε to bright ~i, lack intensity, and fall outside of the critical band. Essentially, these higher harmonics add a sheen to the ~o and ~ɔ that dominate the percept, without changing the quality of the obvious true fundamental.

When comparing Ms. Fleming’s performance to that of a heavier, brighter voice, such as Ms. Lisa Della Casa, the non-universality of Ms. Fleming’s approach comes into focus. In Ms. Fleming’s A $\flat 5$, the second harmonic is significantly quieter (on average about 30dB) than the fundamental. The spectral centroid of a segment consisting of these two harmonics at these amplitudes still falls near the $\sim \text{c}$ range (just above the pitch A $\flat 5$). Consider this parsed spectrum graph (**Figure 27**) of Ms. Della Casa singing the same A $\flat 5$ from “Porgi amor.” While Ms. Della Casa’s second harmonic (on average only 20dB less than her first harmonic) does not change the $\sim \text{c}$ found in Ms. Fleming’s first harmonic to a weakly bridged $\langle \alpha^{\text{wb}}$ percept (it would have to have been at a much higher amplitude to do this), compared to Ms. Fleming’s A $\flat 5$, Ms. Della Casa has both a stronger peak in the *bright* $\langle i$ range, and also a peak in an extremely bright range of $\sim i$ (here covering about 6-10kHz, the pitches G8 through D9, labeled “Ext Bright $\langle i$ ”). Notably, this peak falls within the critical band of hearing, and sounds as buzzy and rough as you might find in a baritone’s singer’s formant, albeit with an extremely bright $\langle i$ tone color and some ambiguity of pitch. As **Figure 27** shows, Ms. Fleming is not simply singing a different vowel, or singing the same vowel more quietly. Each singer presents a slightly different sound across the axes of complexity and tone color. These differences may arise from training, aesthetics, or each singer’s inherent capabilities. In this example, Ms. Fleming produces a simpler sound than Ms. Della Casa. As a result, different tone colors come to prominence.

This analysis suggests that active vocal tract modifications above the treble staff do not substitute a workable vowel for a problematic one, and that a resonance strategy (like *whoop*) does not have a uniform sound; rather, as pitch ascends, a vocal tract shape that profitably matches vocal tract resonances and glottal source harmonics produces a small number of harmonics relative to speech, with obligatory and disparate absolute spectral tone colors. The

percept of the potential variations of these absolute spectral tone colors is significantly limited relative to speech (see again **Figure 7** and consider the objective tone colors of the high-pitched bottom sample). Exploring this idea further, **Figure 28** displays a soprano singing the pitches C5-C6 (which rise above the absolute spectral tone color range of ~u) first with a resonant vocal tract shape of her choice, and second with vocal tract strictly shaped like a dark [u], each harmonic exhibits the same absolute spectral tone color in each resonant/non-resonant pair, despite changing the vocal tract. What changes is the overall intensity of the spectrum, and the physical ease of production. This ease comes across in the sound as even/uneven vibrato and accurate/inaccurate intonation. Stripped of contextualizing consonants, from D5-C6 the second of each pair never actually sounds like [u], though it does contrast with the resonant version. This pitch range of the female singing voice is fascinating for this reason: the context effect becomes vital to vowel perception in exactly the range where consonants are most limited by the decreasing range of coarticulation of the available vocal tract shapes. Technical training of classical singers in this range becomes increasingly focused on setting tonal expectations such that the singer only attempts workable modifications. Another way to think about this is that voice teachers may better serve their students by suggesting that they make a pitch through a certain shape, and let go of the idea of a vowel entirely. Indeed, I suspect that in the higher female voice one does not infer the correct word in consonantal context despite the presence of the wrong vowel; rather, one infers the correct vowel because the sung sound is perceptually ambiguous at those pitches. Effectively, above the staff, vowel clarity disappears entirely because the simplicity of the percept is too distant from the timbral complexity of speech. Characterized by strong qualities of absolute spectral tone color, and weakly bridged tone colors, what remains is vague enough that multiple linguistic meanings may be overlaid.

The Classical Female (and Countertenor) Voice on the Treble Staff

The female (and countertenor) voice on the treble staff presents a different set of issues as this range transitions from the complex qualities of speech to the dissolution of speech like vowels that occurs above the staff. As discussed in chapter 3, the principle of the obvious true fundamental suggests that beginning approximately with the pitch C#4, the fundamental is the only harmonic with its absolute spectral tone color. The second harmonic, an octave higher, is at least one absolute spectral tone color range away. Yet for many vowels on the staff (with the possible exceptions of [u] and [o]), the fundamental remains an additional warm quality unnecessary for vowel identification. A classical singer using a *whoop* resonance strategy will allow the absolute spectral tone color of the first harmonic to exert itself, but it remains perceptually separable. This is perhaps most noticeable for vowels with a high second spectral peak, like [i]. **Figure 29** shows a countertenor moving between [u] and [i] on the pitch E4. In both vowels the <u is strongly defined by the ~u absolute spectral tone color of the fundamental (highlighted with overlaid boxed text), which remains the same amplitude in both vowels. More dramatic in the [i] than [u] (yet still noticeable in the [u]), the fundamental (~u) becomes perceptually separable from the second and third harmonics. Heard in isolation, one is unable to discern whether the fundamental was extracted from the [i] or [u], and as Vennard predicted, the ~u percept is a noticeable common thread between these vowels.

I will conclude this glimpse into a psycho-acoustical analysis of the female (and countertenor) voice on the treble staff with an excerpt from the classical repertory that illustrates the perceptual shifts that occur in this pitch range. In m. 46-49 of Richard Strauss' "Beim Schlafengehen" from *Vier Letzte Lieder* (**Figure 30**) from Ms. Gundula Janowitz's 1971 recording, the continuous power of the fundamental relative to the remaining harmonics is

apparent in the spectrogram, and exists because Ms. Janowitz couples her first vocal tract resonance with her first harmonic (*whoop* resonance). The spectrogram in **Figure 31** displays the vertical frequency axis logarithmically to aid aligning the fundamental with the pitches on the piano keyboard. Additionally, I have delineated the ranges of absolute spectral tone color relevant to the compass of the fundamental. **Figure 31** suggests that the fundamental (at times significantly more intense than the rest of the spectrum) should be separable by tone color. Indeed, once the remainder of the spectrum is removed, the fundamental can be heard changing along a continuum of ~u, ~o and ~ɔ as pitch rises. Whether an untrained ear (or a trained ear distracted by something else) would hear this distinctly is not particularly relevant to the value of listening in this manner. Since whoop resonance is a technical competency for classical singers (especially, but not limited to female, countertenor, and unchanged voices), the ability to hear this harmonic separately, and thus determine if the coupled resonance is correctly managed by the student, is advantageous for singing teachers. **Figure 32** is a series of LTAS parsed spectrum graphs, one for each pitch of the melisma on the word “Flügen” in m. 46-47. All the tone colors across this entire sample are indicated on the depth axis, and the obvious true fundamental is indicated as boxed text. The spectral centroid (the amplitude-weighted average frequency that characterizes the < percept) of the first and second harmonics (in this excerpt rarely higher than the tone color of the first harmonic in isolation) is indicated in the lower text box. In this way, one may observe how the number of < percepts changes as pitch rises and falls (generally more for lower pitches), how the amplitude of the lowest < percept comes to dominate the vowel as pitch rises, and how the same pitch in different contexts might vary in subtle ways. What is certainly clear from these graphs is that each pitch has a different tone color quality (despite the same written vowel), strongly characterized by the < or <^{wb} percept of the first and second

harmonics and the ~percept of the first harmonic. Also of note, above the pitch F5, the lowest spectral peak absolutely dominates the overall sound. **Figure 33** illustrates that such an approach dictates that different vowels will reduce to a similar simple percept at high enough pitches. Compare the F5 from the word “Flügen” ([y]) m. 47 (top) with the F5 from the word “schweben” ([e]) (bottom) from m. 48. One may see that despite the difference in written vowel, the parsed spectrum graphs show that Ms. Janowitz sings almost identical sounds. Neither are true vowels, yet they create a very similar percept.

Returning to Miller’s transitional pitches for female registration, the graphs in **Figure 32** also show that below the pitch C5 the separate <percepts are more equally represented. The A ♭4, for example, features an <e of near equal amplitude with the <u. In contrast, the E ♭5 already exhibits characteristics of pitches above the staff: a more dominant low spectral peak, and a simpler, quieter, higher spectral peak. It is worth noting that E ♭5 is the highest pitch in this sample where Ms. Janowitz was able to energize harmonics in the <e range. The F5, G5, and A ♭5 all simplify to a second spectral peak of <i. So, elements of psychoacoustics are at play in the pitch regions that Miller defines as vital to understanding registration in the female voice. Simply describing these transitions as changes in vowels misses the rich variation apparent in the parsed spectrum graphs. One would expect a performance by a different singer, or by a singer who belted rather than whooped to also be meaningfully describable by this analytical method. Further research is warranted.

Conclusions

The immediate implications of this type of analysis are profound: rather than reduce the sound of an elite classical singer to a group of vowels, one may productively analyze its inherent tone color, describe obligate changes that deepen one's understanding of singing technique, and graph a singer's expressive choices with greater detail than previously available. I hope that a computer program will eventually be able to execute this sort of analysis in real time, though certain challenges (such as defining threshold parameters for the local spectral coherence effect) need to be addressed first. More broadly, a sung vowel may be divided into multiple unique percepts so long as one knows where to aurally locate them. Placed within a framework that explains the manner in which these percepts change, one may hear and predict qualitative differences in tone color as pitch and vowel vary. This is especially relevant for treble staff singers (e.g. females, children, countertenors), as the tessitura of their fundamentals encompasses several ranges of absolute spectral tone color, and the potential rich roughness of lower pitches gives way to weakly bridged tone colors as pitch ascends. I believe that the sound of optimal vocal registration may be described with greater specificity than the current visual models to describe vowels afford. Members of the singing voice science and pedagogy communities may be able to use this type of analytical approach to better listen, teach, and sing.

Appendices

Appendix A: International Phonetics Alphabet Symbols

IPA CHART FOR LYRIC DICTION

VOWELS	ENGLISH	ITALIAN	GERMAN	FRENCH
[i]	see [si]	ivi [ˈivi]	Lied [li:t]	ici [isi]
[ɪ]	wit [wɪt]		Bitte [ˈbɪtə]	
[e]	chaotic [keˈatɪk]	perché [perˈke]	Erde [ˈʔe:rdə]	été [ete]
[ɛ]	bell [bɛl]	bene [ˈbene]	Feld [fɛlt]	rêve [ʁevə]
[u]	blue [blu]	luna [ˈluna]	Ruhe [ˈru:ə]	jour [ʒur]
[ʊ]	look [lʊk]		jung [jʊŋ]	
[o]	obey [ʔoˈbeɪ]	solo [ˈsolo]	Mond [mo:nt]	pauvre [povrə]
[ɔ]	ought [ʔɔt]	core [ˈkɔre]	Sonne [ˈzɔnnə]	aurore [ɔrɔrə]
[æ]	glad [glæd]			
[a]	sigh [saɪ]	cara [ˈkara]	allein [ʔalˈlaen]	voilà [vwala]
[ɑ]	father [ˈfɑðə]		Abend [ˈʔa:bənt]	âme [amə]
[ʌ]	up [ʔʌp]			
[y]			Blüte [ˈbly:tə]	sûr [syʁ]
[ʏ]			Küsse [ˈkyssə]	
[ø]			schön [ʃø:n]	feu [fø]
[œ]			können [ˈkœnnən]	seul [sœl]
[ɜ]	bird [bɜd]			
[ʁ]			Vater [ˈfa:təʁ]	
[ɑ̃]				enfant [ɑ̃fɑ̃]
[ɛ̃]				bien [bjɛ̃]
[õ]				ombre [õbrə]
[œ̃]				humble [œ̃blə]

Lyric Diction Workbook Series

Audio Examples are on the listening page at
www.stmpublishers.com

Source: Cheri Montgomery, “IPA Charts for Lyric Diction,” STM Publishing (stmpublishing.com), <http://nebula.wsimg.com/0f288123bb2580bf33136ca9d158a8aa?AccessKeyId=2FDAD2B87450CB38B630&disposition=0&alloworigin=1>. Reprinted with the gracious permission of the publisher.

Figures

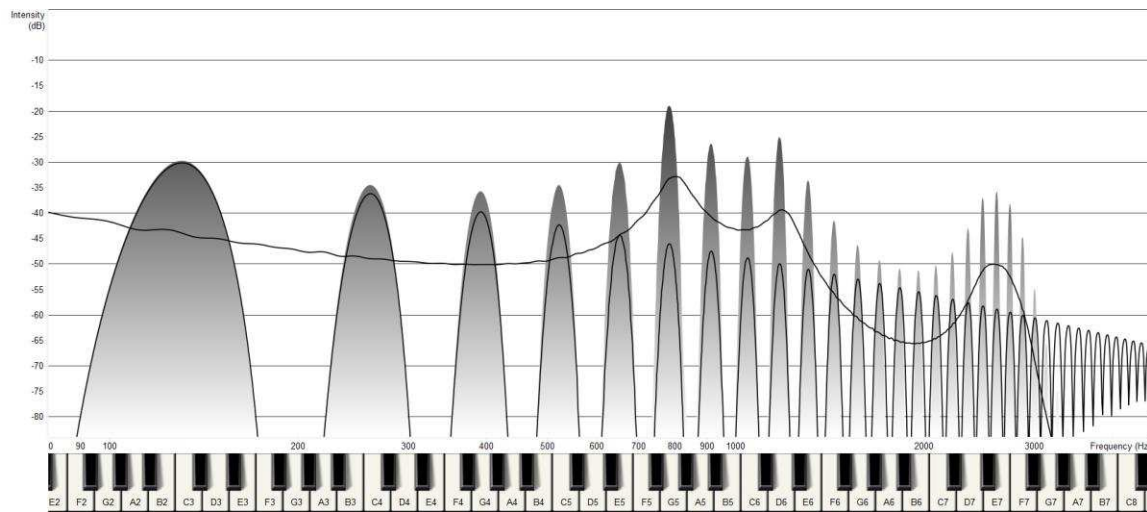


Figure 1: The vocal tract filter reshapes the spectral envelope of the glottal source sound. The black outlined harmonics represent the glottal source sound (here a synthesized example of the compression wave created by the vocal folds) that sounds like a buzz due to the steady roll off in amplitude of higher harmonics. The black line represents the disposition of a vocal tract shaped for [a] to reshape the source harmonics. The solid harmonics represent the spectrum of the reshaped compression wave. Note that the highest intensity harmonics fall into the octave from E5 to E6. **Source:** Author's analysis.

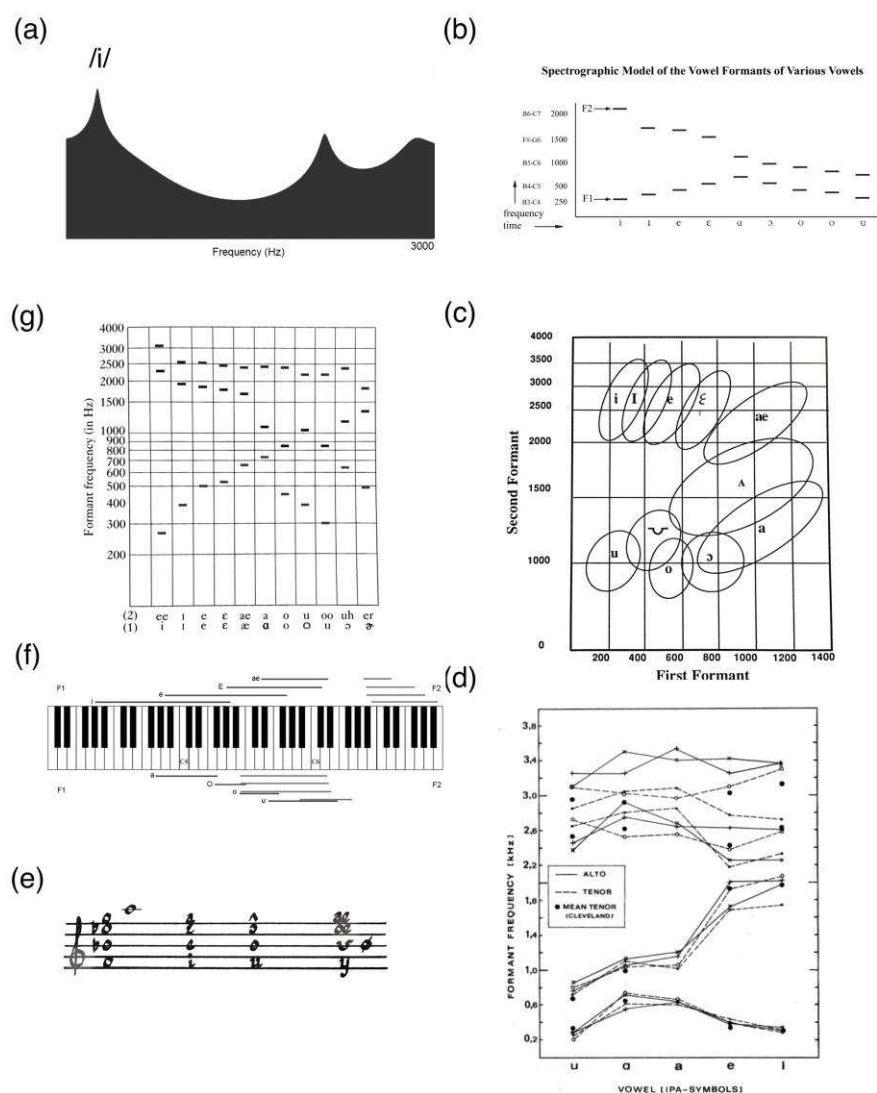


Figure 2: Charts representative of models used in voice science and vocal pedagogy texts to explain vowels in terms of vocal tract resonances and spectral peaks. Clockwise from top left: (a) average acoustic spectrum of [i] from McCoy, *Your Voice: An Inside View 2nd Edition*, 42 (horizontal axis is frequency in Hz, vertical axis is amplitude); (b) Graph of first two vocal tract formants [resonances] from Bozeman, *Practical Vocal Acoustics*, 13 (horizontal axis is vowels, vertical axis is frequency); (c) Plot graph of first and second formant (spectral peaks) locations; points falling within a circle represent simultaneous first and second spectral peak values that have been observed in speakers (horizontal axis is the first spectral peak frequency in Hz, vertical axis is the second spectral peak frequency in Hz), from Doscher, *The Functional Unity of the Singing Voice 2nd Edition*, 138; (d) Graph of four lowest vocal tract resonances of specific singers for five vowels, from Sundberg, *The Science of the Singing Voice*, 107 (horizontal axis is vowels, vertical axis is frequency in Hz); (e) Treble staff notation of the average pitch of the first spectral peak for the indicated vowels, from Doscher, *Functional Unity*, 152; (f) Piano keyboard of ranges of first and second vocal tract resonances across voice-types, from McCoy, *Inside View*, 45; (g) A graph of the average values of the three lowest spectral peaks for eleven English vowels as spoken by males, Reprinted by permission from Waveland Press, Inc. from Denes and Pinson, *The Speech Chain*, 143 (horizontal axis is vowels, vertical axis is frequency in Hz). Images a-f reprinted with the gracious permission of their respective copyright owners.

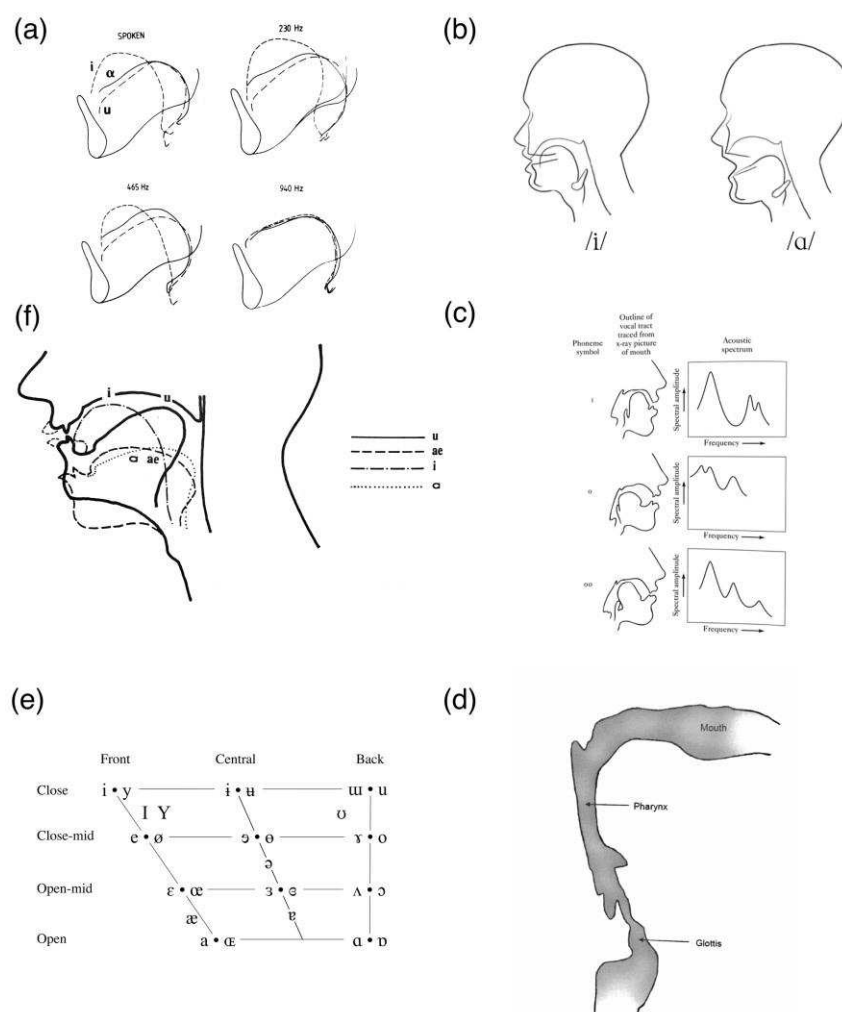


Figure 3: Charts representative of models used in voice science and vocal pedagogy texts to explain vowels in terms of vocal tract shapes. Clockwise from top left: (a) contours of the tongue body for three vowels in four pitch ranges from Sundberg, *The Science of the Singing Voice*, 128; (b) side view of the vowels [i] and [a] from Bozeman, *Practical Vocal Acoustics*, 62; (c) aligned vocal tract shapes and spectrum of three vowels reprinted by permission from Waveland Press, Inc. from Denes, *The Speech Chain*, 74; (d) sketch of the air-containing spaces of the vocal tract (not including the nasal cavity) for a neutral vowel from Titze, *Vocology*, 340; (e) The vowel quadrilateral from IPA Chart, <http://www.internationalphoneticassociation.org/content/ipa-chart>, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association. This is an organization of vowels along a horizontal axis representing the location of the point of greatest tongue constriction in the mouth (front, central, back), and a vertical axis representing the distance of the tongue from the roof of the mouth (close, close-mid, open-mid, or open). Vowel pairs to the right and left of the bullet points represent rounded and unrounded (lips) versions; (f) tongue position for four vowels from Doscher, *Functional Unity of the Singing Voice*, 112. Images a, b, d, and f reprinted with the gracious permission of their respective copyright owners.

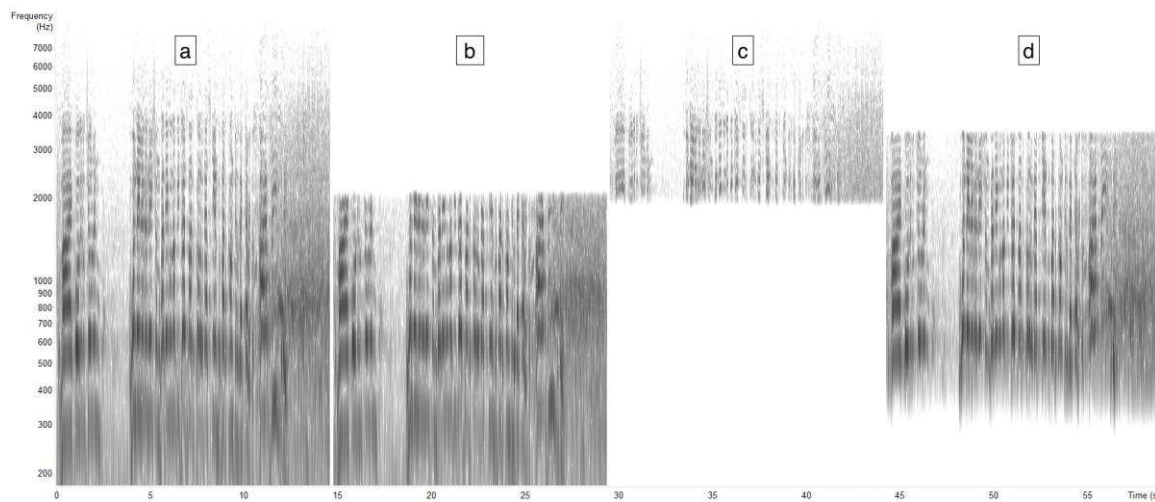


Figure 4: A portion of Dr. Martin Luther King’s “I Have a Dream” (a) with original spectrum of the recording, (b) only sound below 2,000 Hz, (c) only sound above 2,000 Hz, and (d) only sound between 300 Hz and 3400 Hz (similar to an analog telephone). The text is understandable in all four versions, which supports the idea that speech sounds are simultaneously encoded into the entire spectrum. **Source:** “Martin Luther King, Jr. I Have A Dream Speech,” <https://www.youtube.com/watch?v=3vDWWy4CMhE>

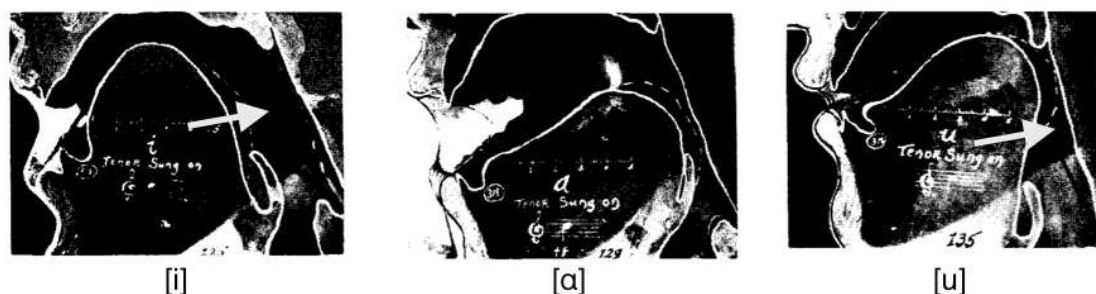


Figure 5: X-Ray images of Enrico Caruso's mouth and throat in profile, from left to right [i], [a], [u]. Note that the pharyngeal space (from glottis to the tongue hump, here indicated with a white arrow), a primary determinant of the pitch of the first vocal tract resonance, is significantly larger for [i] than [u]. However, in a given human these vowels share roughly the same first vocal tract resonance frequency, imparting a near identical vowel-like aspect of timbre to the overall timbre of each vowel. **Source:** G. Oscar Russell, *Speech and Voice, with X-rays of English, French, German, Italian, Spanish, Soprano, Tenor, and Baritone Subjects* (New York: Macmillan, 1931), found in Berton Coffin, *Overtones of Bel Canto* (Metuchen: Scarecrow Press, 1980), 183. Reprinted with the gracious permission of Scarecrow Press.

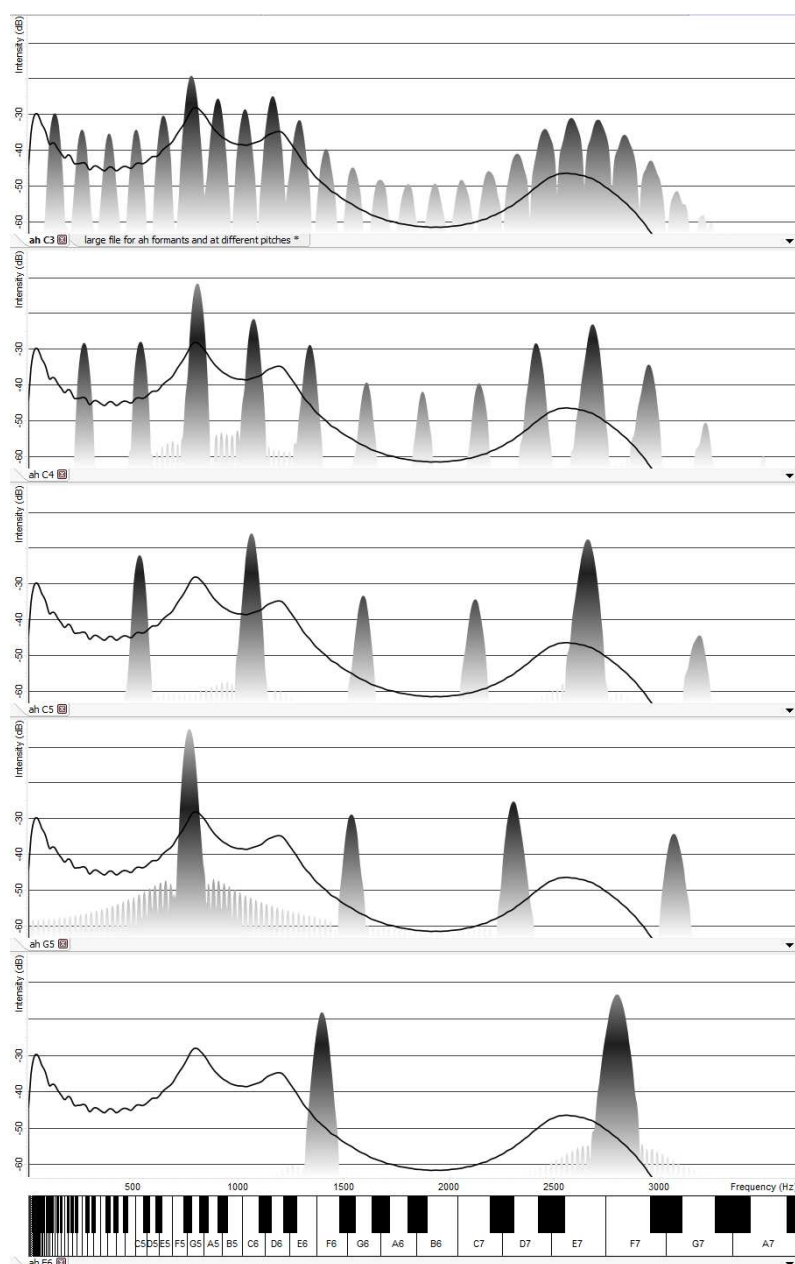


Figure 6: From top to bottom the density of harmonics (on a linear frequency axis) for the pitches C3, C4, C5, G5, and F6 as filtered by the resonances of an [a] shaped vocal tract (the black solid line). Note that as pitch rises all harmonics rise and become progressively more spread out. The harmonics of C3 and C4 outline a shape similar to the ideal envelope for [a]. C5, G5, and F6 form much simpler shapes. The change in envelope fundamentally changes the vowel percept. **Source:** Vowels synthesized by author in MADDE.

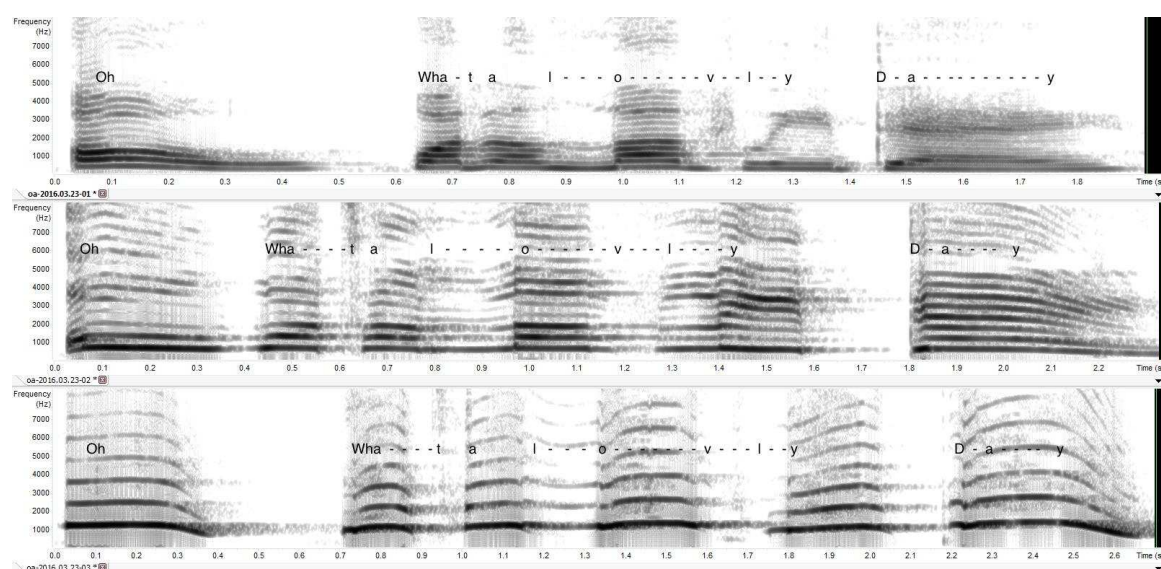


Figure 7: Spectrograms of a single female subject speaking the phrase, “Oh, what a lovely day” in three pitch ranges. A comfortable speaking pitch varying between approximately C4 and G4 (Top), centering around D5 (Middle), and centering around D6 (bottom). Note the diminishing density of harmonic information as the pitch rises. **Source:** Analysis by the author.

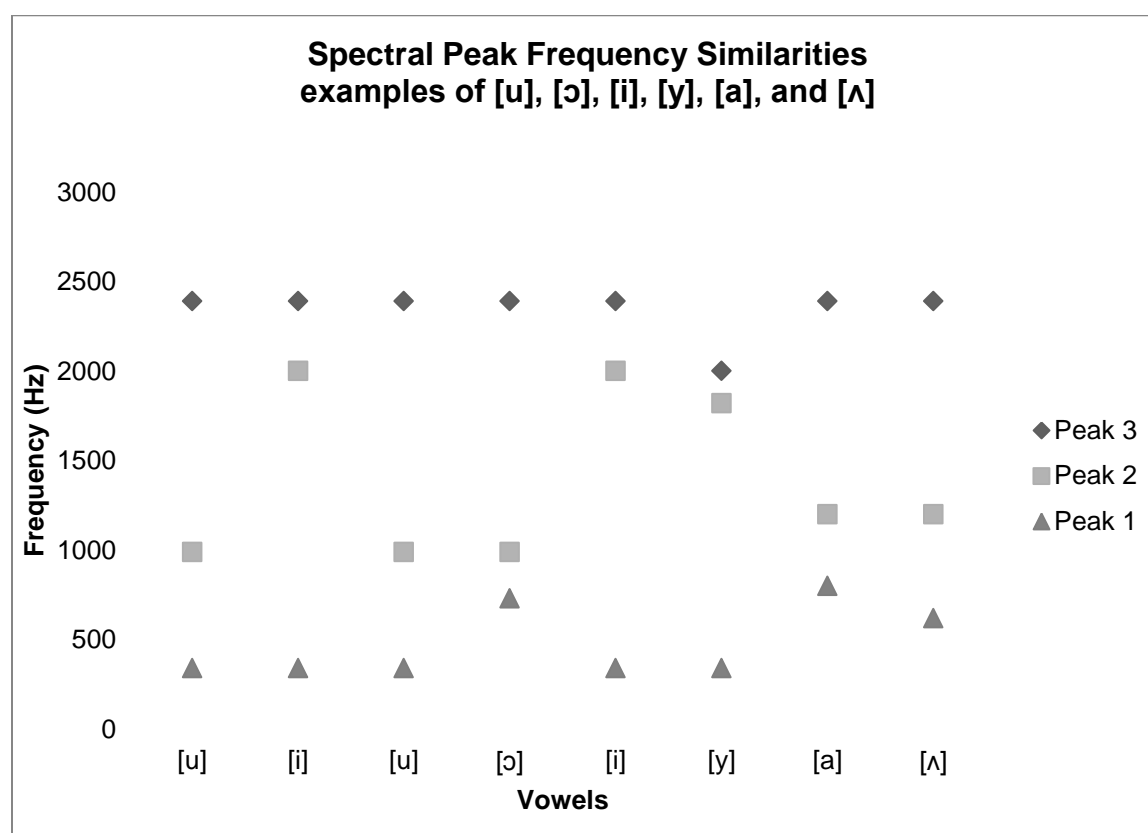


Figure 8: First, second, and third spectral peak frequency centers for the sung vowels [u], [ɔ], [i], [y], [a], and [ʌ]. Note that almost all these vowels share a third peak. Additionally, [u] and [i] share a first peak; [u] and [ɔ] share a second peak; [i] and [y] both share a first peak and also [i]'s second peak may overlap with [y]'s third peak; and [a] and [ʌ] differ only in the frequency of the first peak. **Source:** Vowels synthesized by author in MADDE.

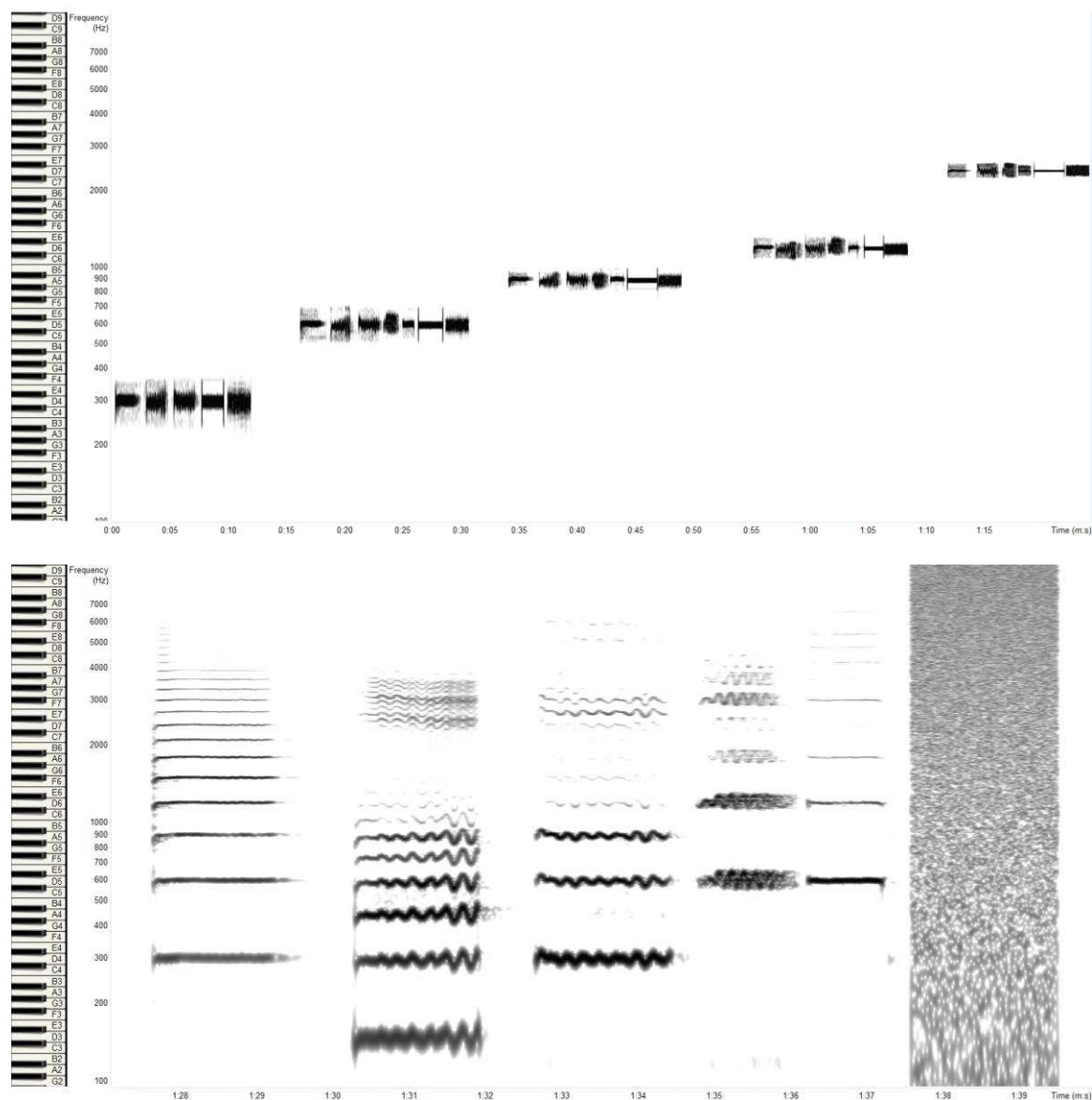


Figure 9: Comparing the tone colors of isolated harmonics from different sources. (Top) From left to right, normalized isolated harmonics with the frequency of the fundamental of D4, D5, A5, D6, and D7. Isolated harmonics of the same frequency all exhibit the same tone color. The sources (bottom) from left to right are a trumpet, male voice, same voice an octave higher, female voice, alto recorder, and white noise. Additionally, a sine tone has been added to the isolated samples (top). **Source:** All files recorded and analyzed by author.



Figure 10: From left to right a bass-baritone singing [a] (without vibrato) at the pitch C3, a tuba playing the pitch C3, and the singer's spectrum reshaped to imitate a long term average spectrum (LTAS) of the tuba sample. The tone color of the third sample is clearly that of a tuba, but the attack, decay, release, and overall spectro-temporal flux is that of a singer. **Source:** Author's analysis.

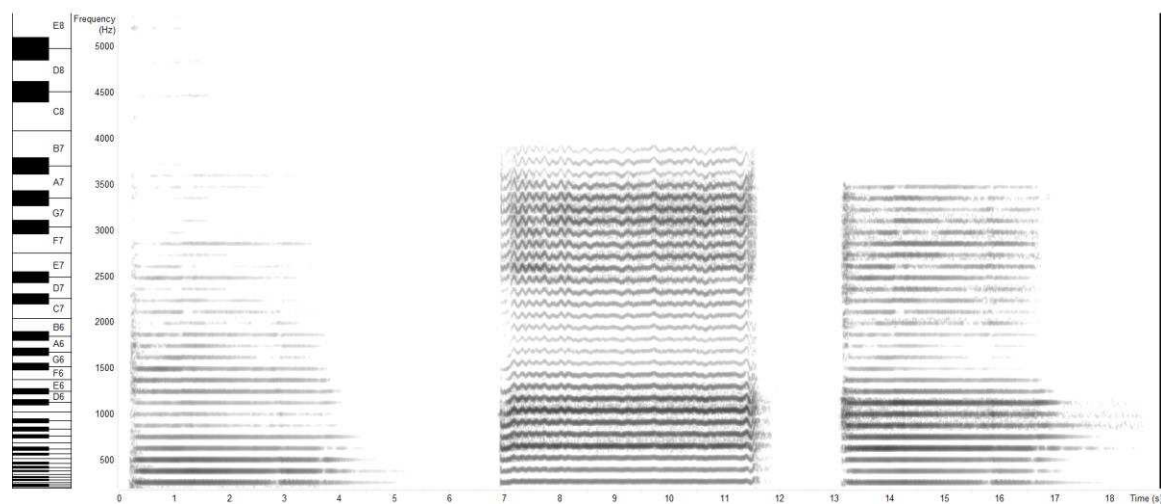


Figure 11: From left to right a bass viola da gamba playing the pitch B2, a bass-baritone singing [a] (without vibrato) at the pitch C3, and the viola da gamba's spectrum reshaped to imitate a long term average spectrum (LTAS) of the singer's sample. The tone color of the third sample is clearly that of an [a] vowel, but the attack, decay, release, and overall spectro-temporal flux is that of a viola da gamba. **Source:** Author's analysis.

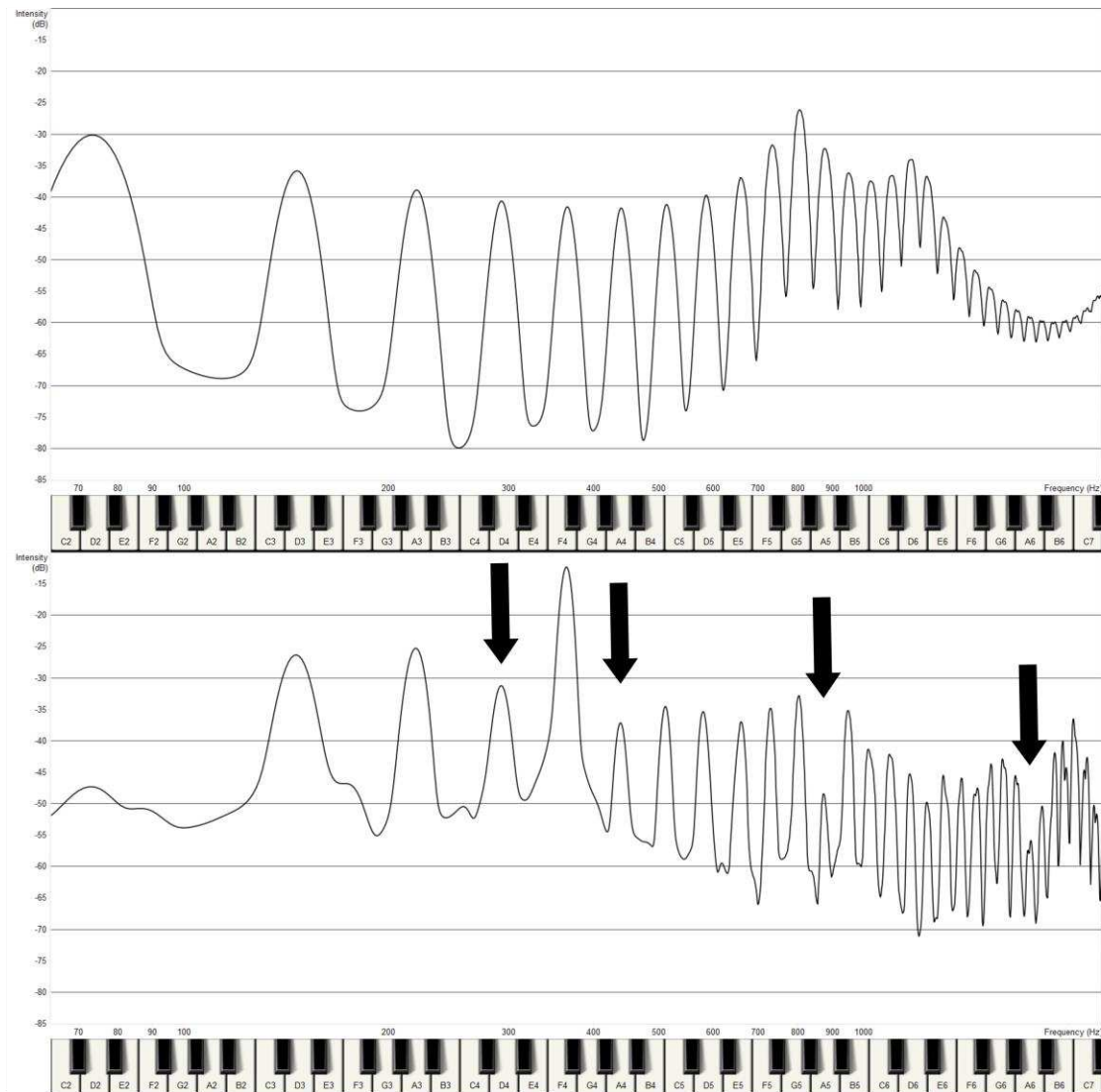


Figure 12: A long term average spectrum (LTAS) of a synthesized voice (top) and cello (bottom) on the pitch D2. Note that the slope formed by the peaks of the voice's harmonics is continuous. The cello's slope has several sudden drops in harmonic intensity (indicated with arrows). **Source:** Top synthesized in MADDE by author. Bottom from "Cris Cambell - Tone and Color of the Cello - Strings By Mail," https://www.youtube.com/watch?v=t1RsDTQ8_Bg.

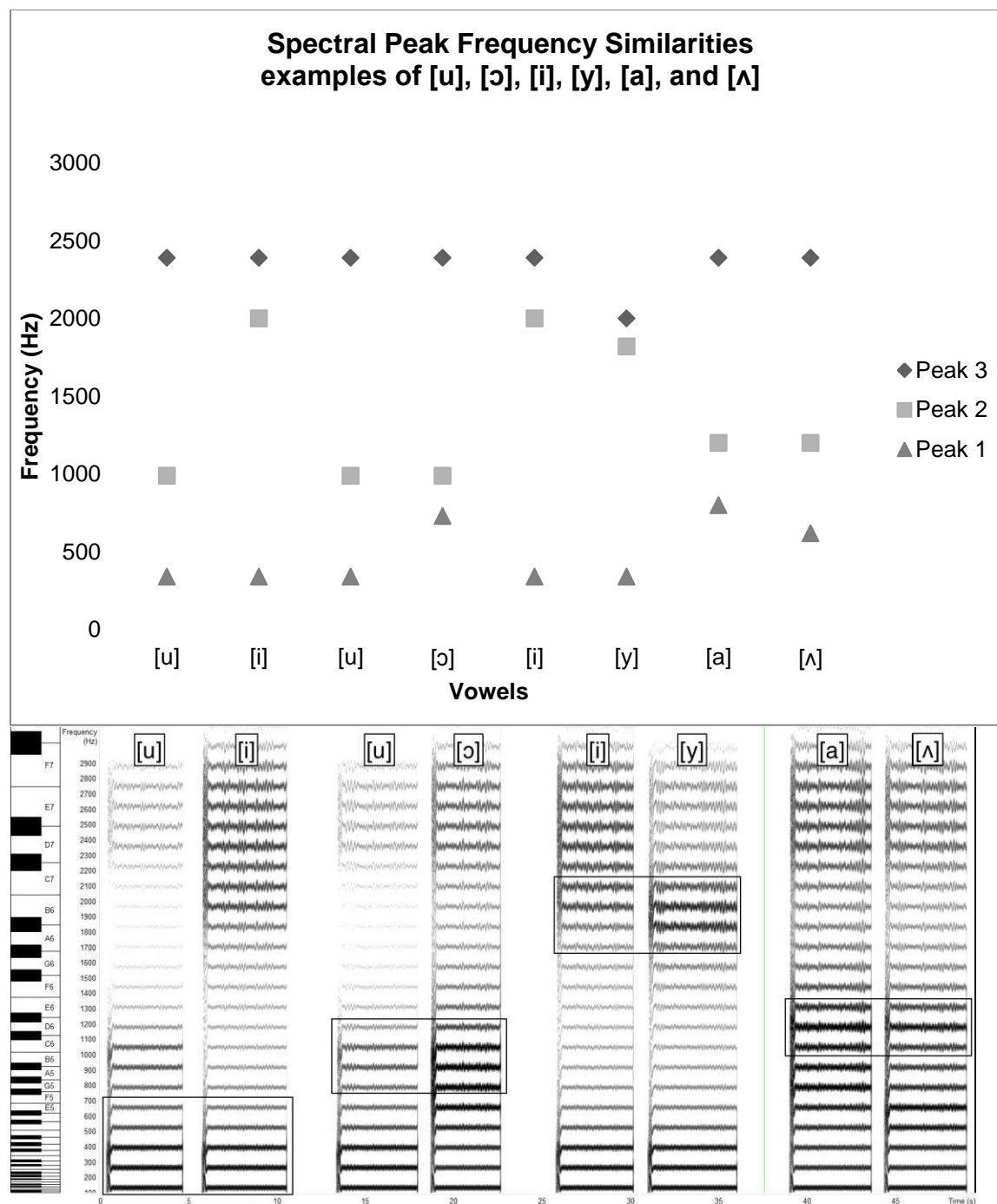


Figure 13: (Top) the graph from **Figure 8** showing the first, second, and third spectral peak frequency centers for the vowels [u], [ɔ], [i], [y], [a], and [ʌ]. (Bottom) the vowels graphed in **Figure 8** shown spectrographically. From left to right, [u], [i], [u], [ɔ], [i], [y], [a], [ʌ] all on the pitch C3. Common spectral peaks between vowel pairs (excepting the near uniform third spectral peak) are indicated here with boxes. **Source:** (Top) Synthesized by author in MADDE; (bottom) Author's analysis.

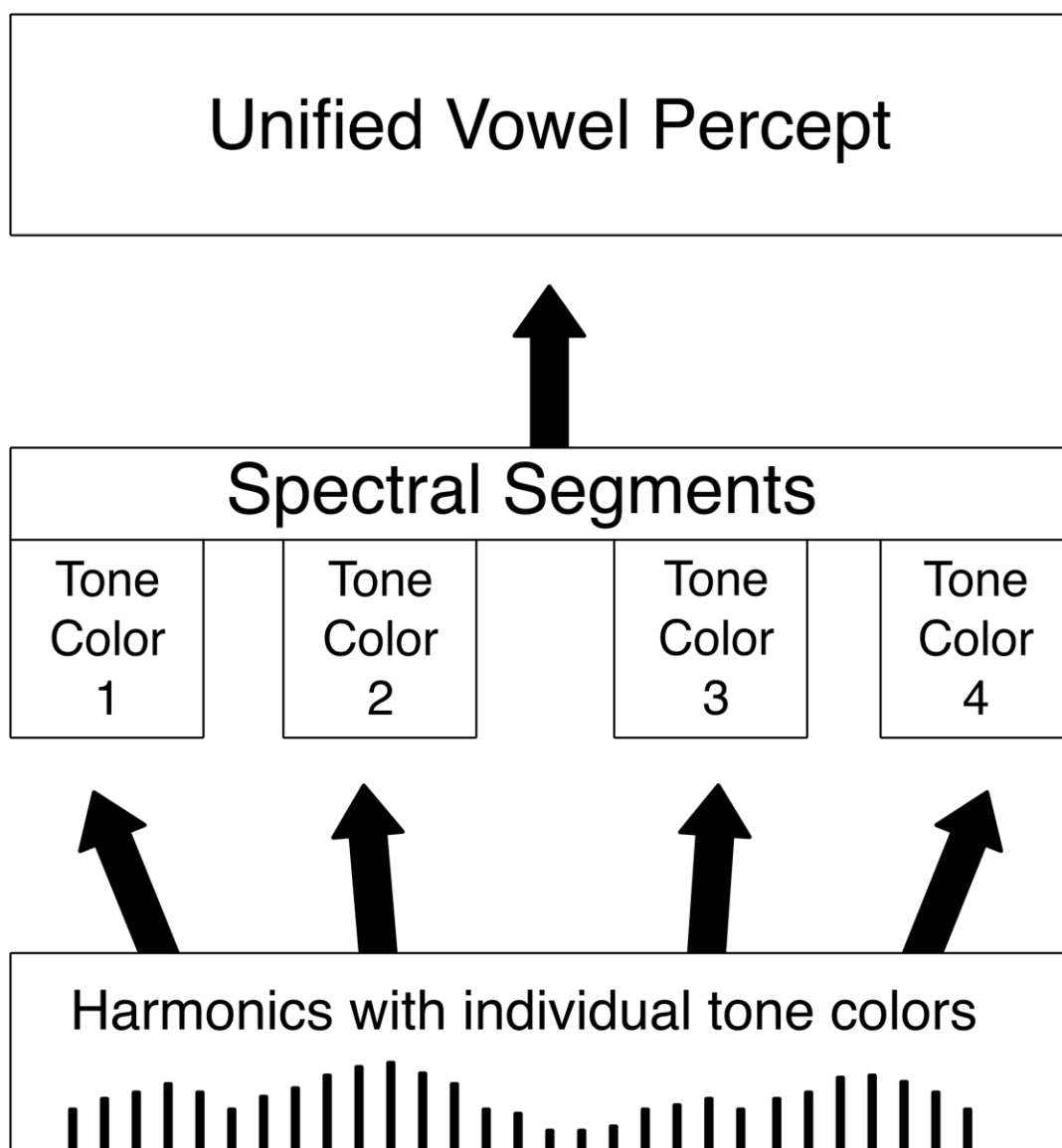


Figure 14: Perceptual flowchart of the manner in which harmonics with individual tone colors (bottom) cohere into spectral segments (frequency bands) with unique tone colors (middle), which in turn form the unified vowel percept (top). **Source:** Author.

Absolute Spectral Tone Color

Approximate Ranges of "Vowel-like Colors" for a Simple Sound

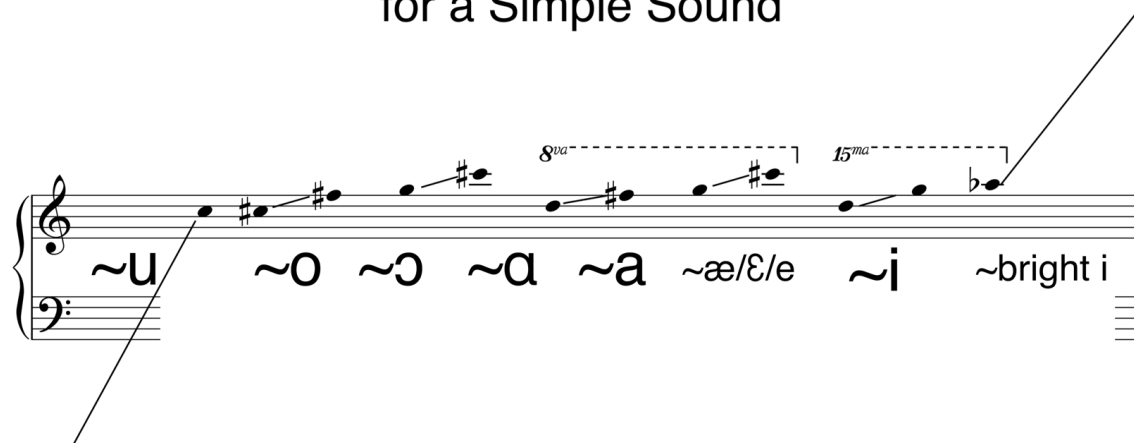


Figure 15: Absolute Spectral Tone Color: Approximate Ranges of vowel-like tone color of a simple sound. I use the convention of ‘~’ followed by the International Phonetics Alphabet (IPA) symbol of the vowel closest in timbre to the simple sound. These values are inspired by Robert Cogan’s register-based analysis of the “sonic qualities” of sine waves, found in **Figure 18**. See **Appendix A** for word-based examples of the vowel sounds associated with each IPA symbol. **Source:** Created by the author.

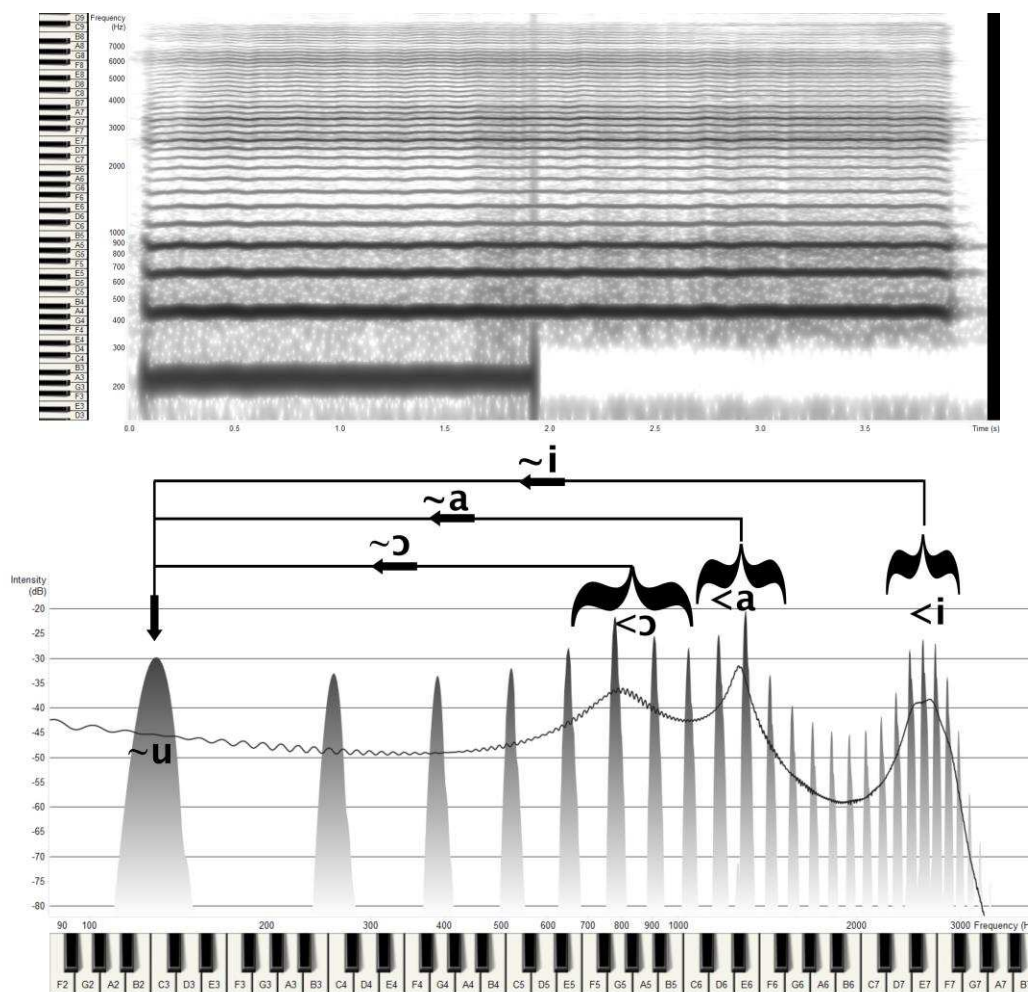


Figure 16: (Top) Spectrogram of male voice singing A3 with and without the fundamental. Perception of the pitch A3 is not changed when the fundamental is removed, just an aspect of the tone color. **Source:** Author's analysis. (Bottom) Author's schematic of the spectrum of a synthesized [a] demonstrating the manner in which spectral peaks are experienced as differently tone-colored aspects of the pitch via the multiple missing fundamentals property. The black line is an approximation of the vocal tract resonances (which correspond with the spectral peaks). Note that the fundamental (first harmonic) has the absolute spectral tone color ~u.

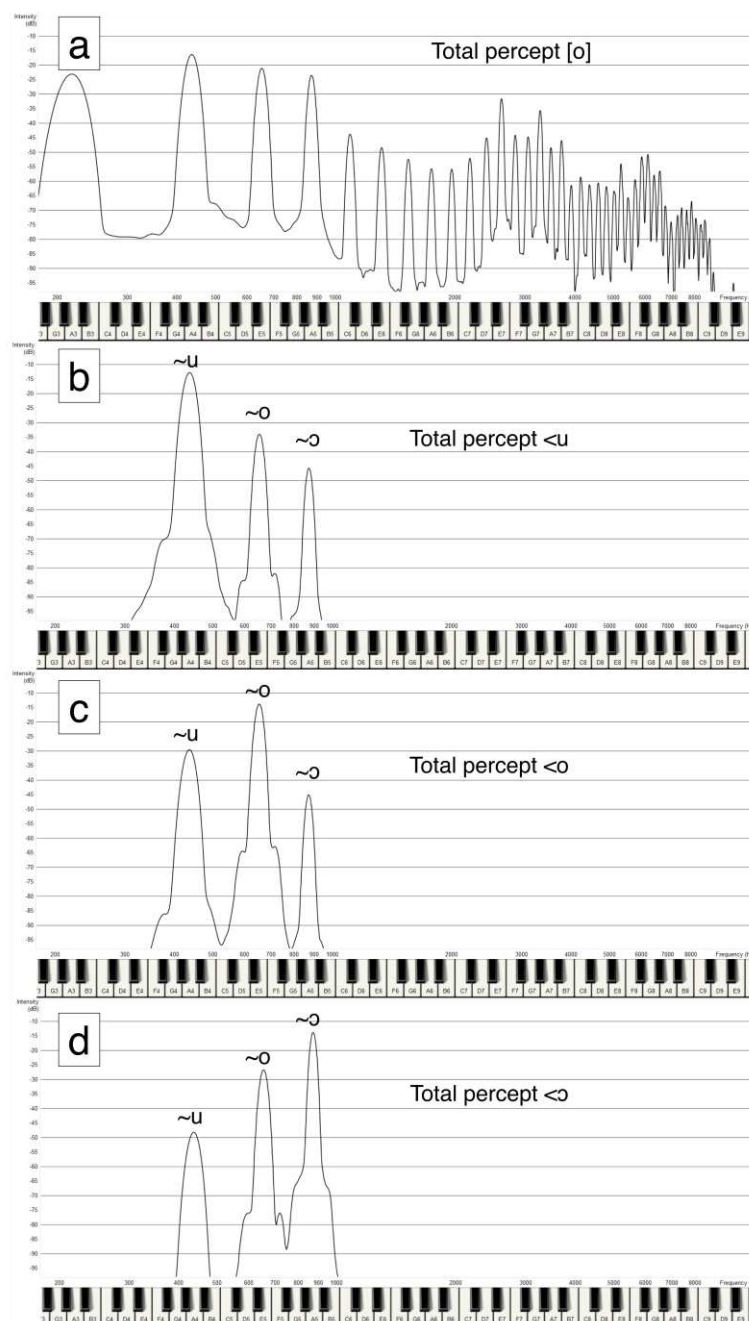


Figure 17: Exploring the local spectral coherence of harmonics 2-4 of the pitch A3. Isolating these three harmonics creates a missing fundamental pitch with a specific tone color. Reshaping this segment of the (a) spectral envelope for [o] results in three different percepts similar to, but less than (<) the vowels [u], [o], and [ɔ]. (b) <u, (c) <o, and (d) <ɔ. **Source:** Author's analysis.

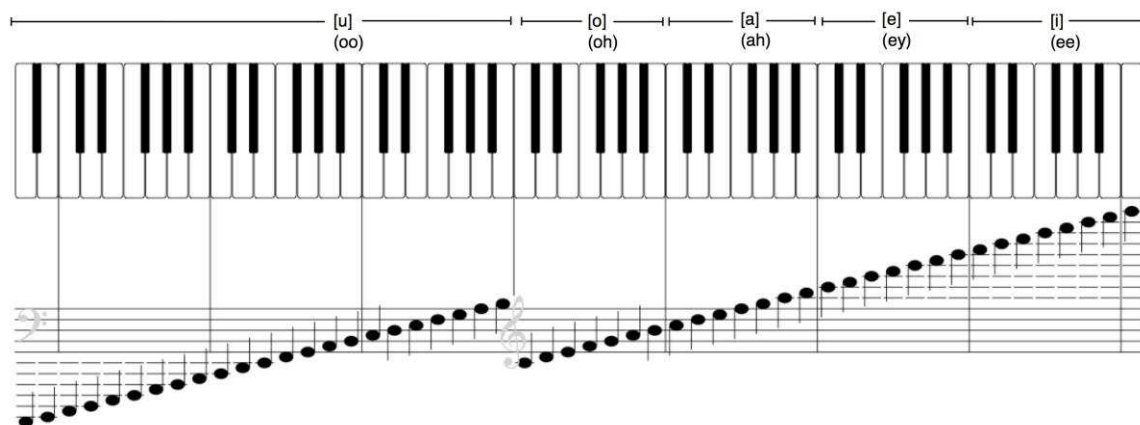


Figure 18: Robert Cogan's register-based analysis of the “sonic qualities” of sine waves. **Source:** A new schematic by the author based on Robert Cogan, *New Images of Musical Sounds* (Cambridge: Harvard University Press, 1984), 7, 12.

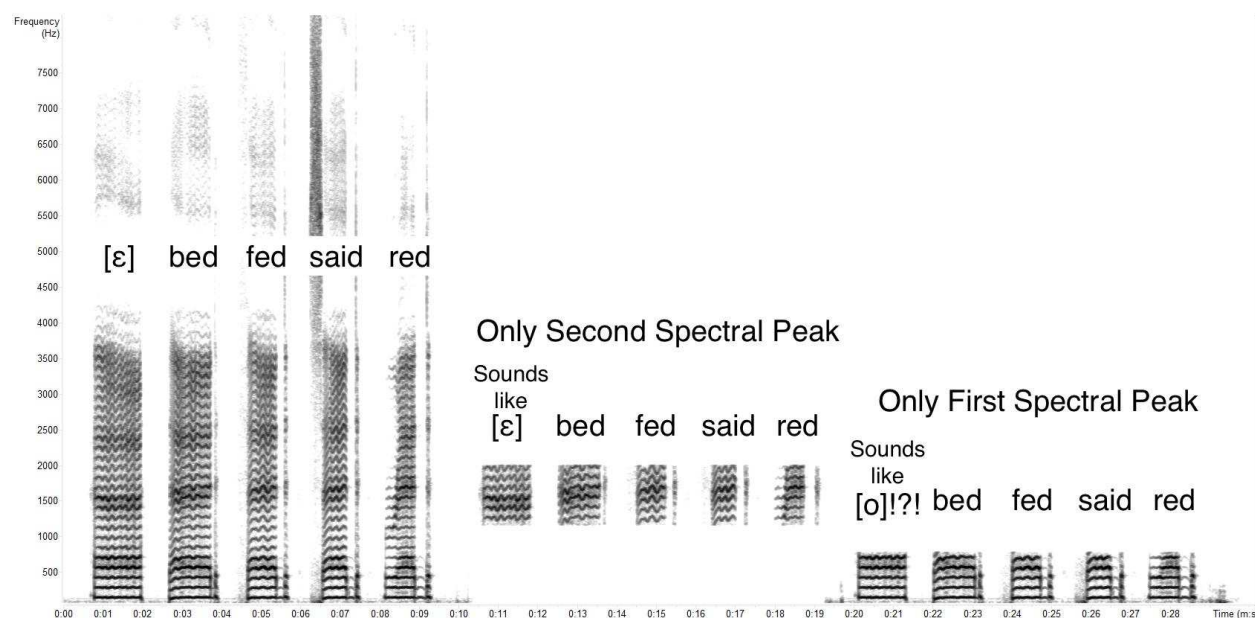


Figure 19: From left to right, “[ε], bed, fed, said, red” three ways. The full spectrum (left), only the second spectral peak (middle), and only the first spectral peak (right). According to Bizley, et al, ferrets can be trained to recognize the second spectral peak as [ε], however they failed to recognize the first spectral peak (here notated as sounding like [o]!?! as [ε]). Note that the context created by coarticulation allows all of the words to be recognized, if distorted.
Source: Analysis of author’s voice by author.

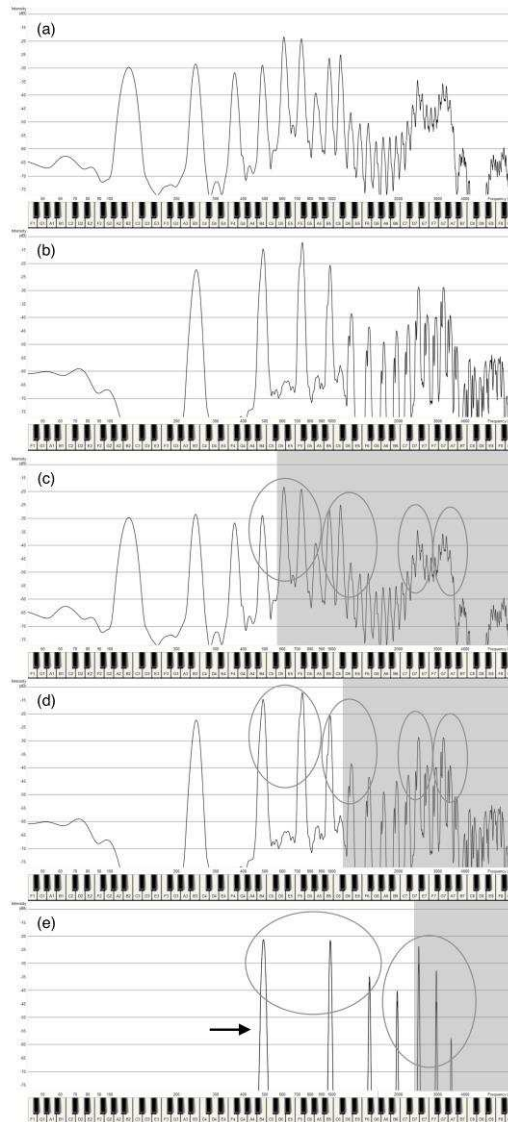


Figure 21: Long term average spectra (LTAS) of [a] sung by a baritone at the pitches (a and c) B2, (b and d) B3, and B4 (e). The horizontal axis is frequency, vertical axis is amplitude. The gray shaded areas indicate regions where harmonics fall within the critical band of their immediate neighbors. Note that the overall spectral shape is very similar (which suggests consistent vocal tract resonance frequencies between samples), and that the B2 has twice as many harmonics as the B3 within the same frequency range. (c) ovals indicate the first four spectral peaks of the B2; each gives rise to a separate missing fundamental with the < tone color of its amplitude weighted central frequency. (d) shows that for B3, the harmonic complexity of each peak has been reduced, and almost all of the harmonics forming the two lowest peaks fall outside of the critical band, further decreasing the complexity of these < percepts. The tone color shifts toward a more neutral vowel, and loses much of the edge and clarity present in the B2. (e) shows a synthesized B4 based on the vocal tract resonances of (a) and (b). This is a completely different percept. The two lowest spectral peaks have merged into a single percept, as have the third and fourth. The fundamental is now clearly audible as a separate tone color, only two missing fundamentals are present, and almost none of the harmonics are within the critical band of a neighbor. This changes the number of missing fundamentals, the complexity (number of harmonics constituting the peak) of each missing fundamental, and the overall quality of the vowel. **Source:** Analysis by author.

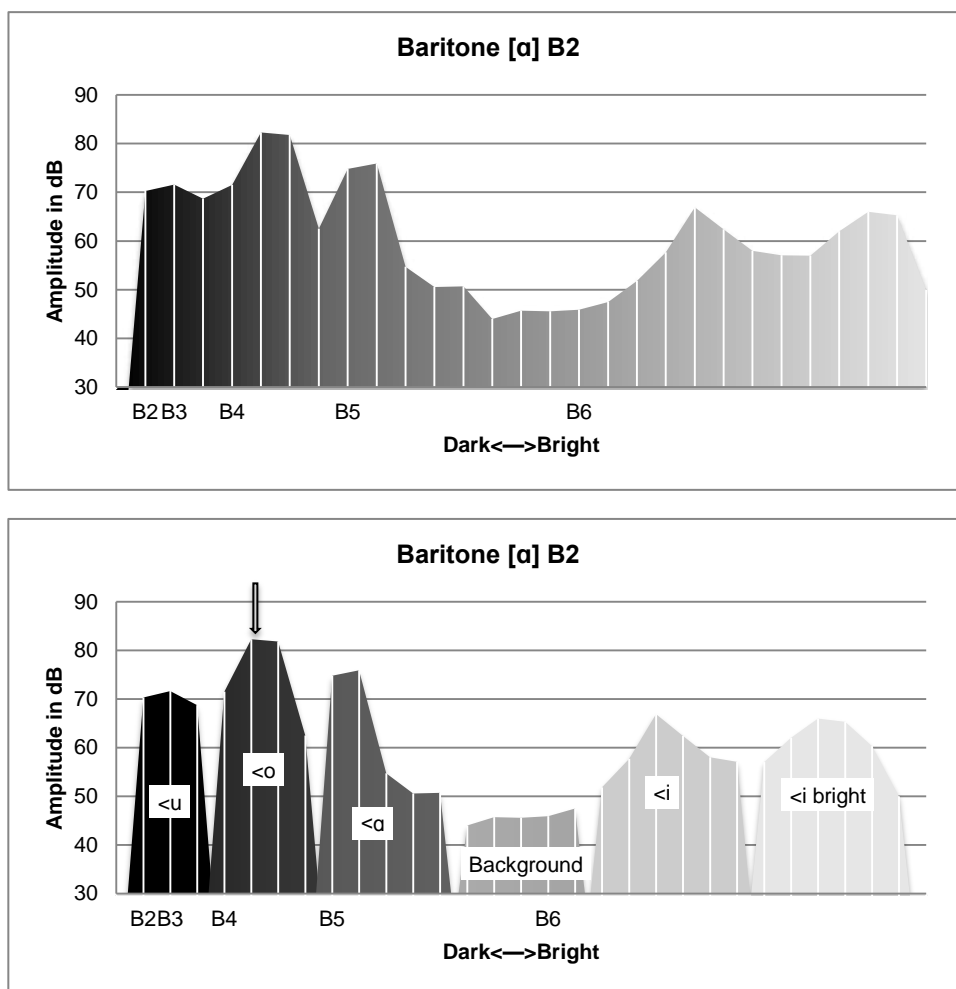


Figure 22: (Top) Traditional spectrum graph of audio from **Figure 21 (a)** with the addition of a scale of brightness from low to high frequency. (Bottom) The same graph broken into individual spectral peaks with the < percepts indicated. Each segment creates its own missing fundamental. Every harmonic to the right of the arrow falls within the critical band of its neighboring harmonics. **Source:** Analysis by the author. Amplitudes in dB are relative, not calibrated.

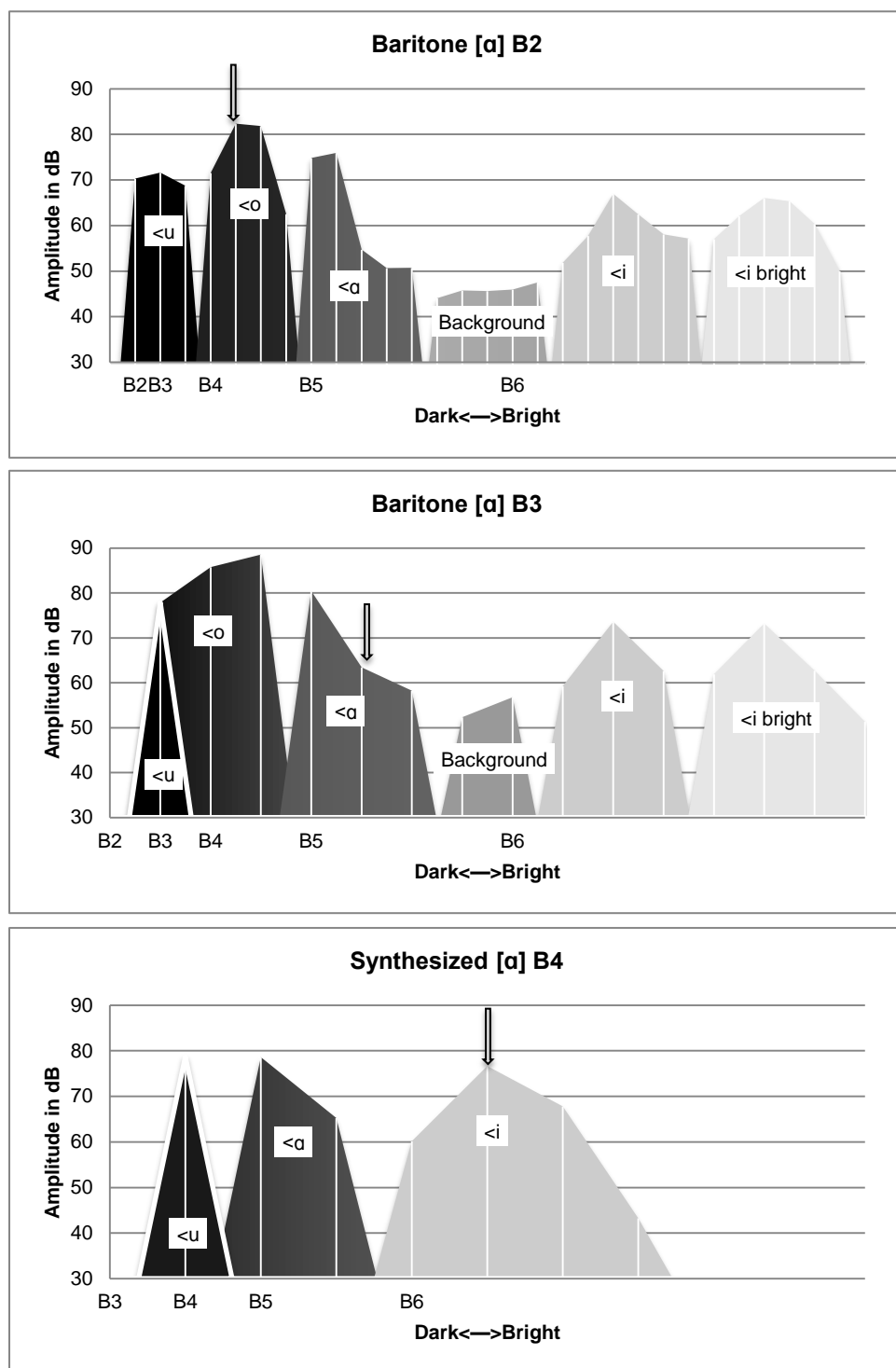


Figure 23: The audio from **Figure 21** (c) top, (d) middle, and (e) bottom, displayed as a graph broken apart by spectral peaks. Note the manner in which both the number of separate < percepts changes as pitch rises, the percentage of the total spectrum that falls above the beginning of the critical band (arrow), and the number of harmonics (vertical white lines) that form each spectral peak. **Source:** Analysis by author. Amplitudes in dB are relative, not calibrated.

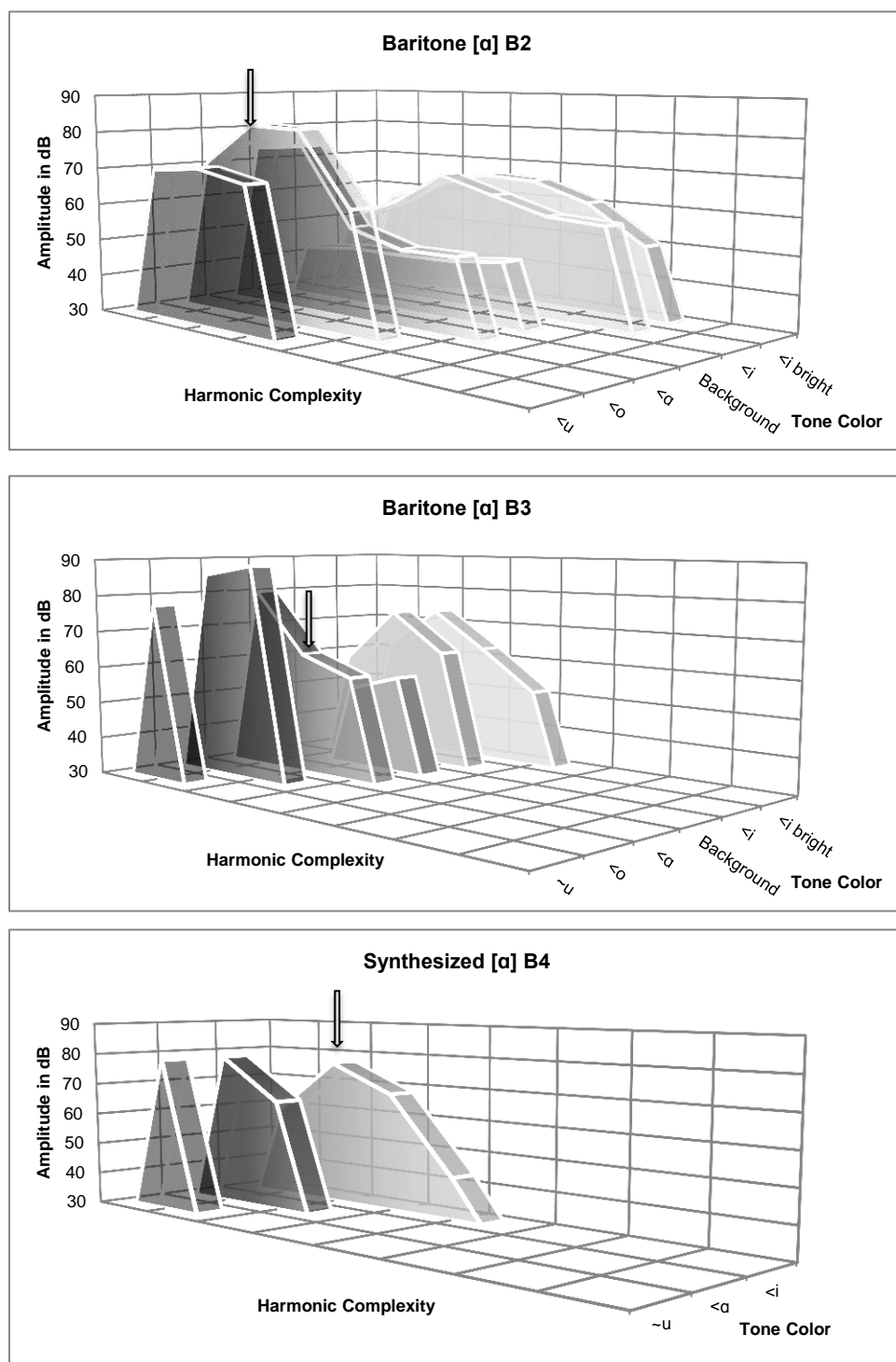


Figure 24: Parsed spectrum models of the B2 (top), B3 (middle), and B4 (bottom) from **Figure 21**. Vertical axis is amplitude, horizontal axis the harmonic complexity, and depth axis is tone color (which becomes shallower for the B4). Arrow indicates the area above which harmonics fall within the critical band. Drop lines indicating harmonics have been removed. Note that the B3 (middle) and B4 (bottom) graph the tone color contribution of the fundamental as separate from the <o and <a respectively. **Source:** Analysis by author. Amplitudes in dB are relative, not calibrated.

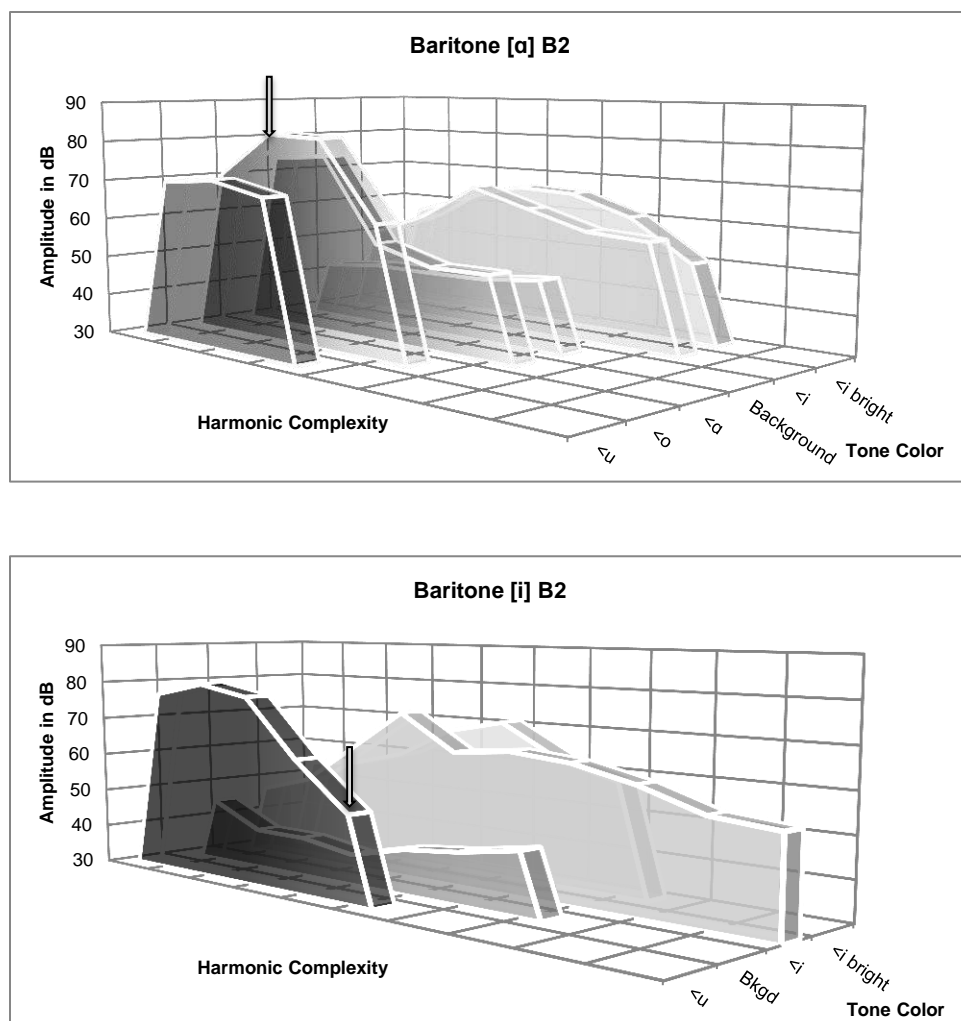


Figure 25: Parsed spectrum graphs of a baritone singing [a] (top) and [i] (bottom) at the pitch B2. Vertical axis is amplitude, horizontal axis the harmonic complexity, and depth axis is tone color (which becomes less complex for an [i]). Arrow indicates the area above which harmonics fall within the critical band. **Source:** Analysis by author. Amplitudes in dB are relative, not calibrated.

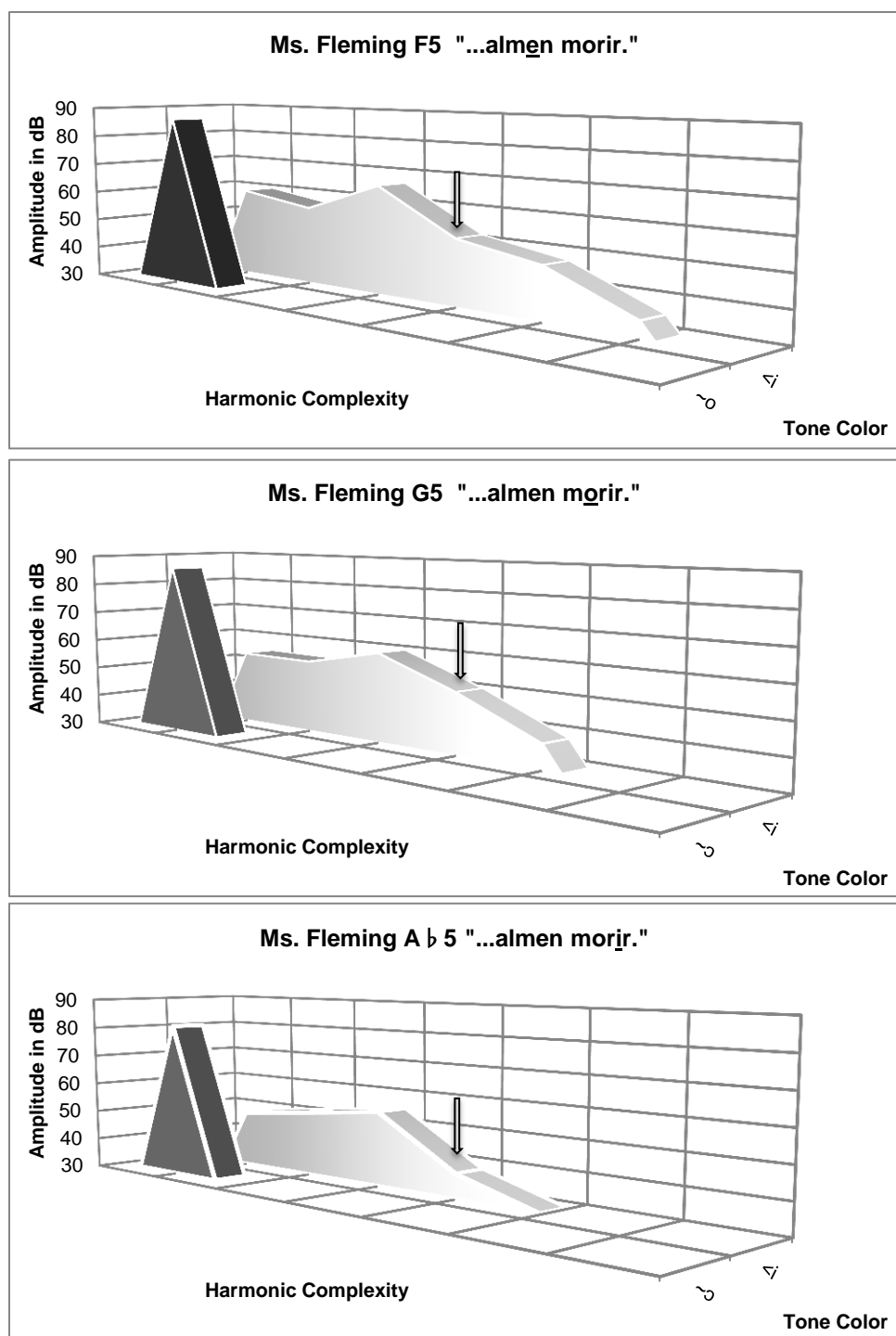


Figure 26: Parsed spectrum graphs of an excerpt from m. 35-36 of Renée Fleming's recording of Mozart's "Porgi amor" from *Le Nozze di Figaro*, F5 (top), G5 (middle), and A ♭ 5 (bottom). Arrow indicates the area above which harmonics fall within the critical band. **Source:** Author's analysis of excerpt from "Renée Fleming Great Opera Scenes," London Symphony Orchestra, George Solti, Conductor (1997). Analysis by author. Amplitudes in dB are relative, not calibrated.

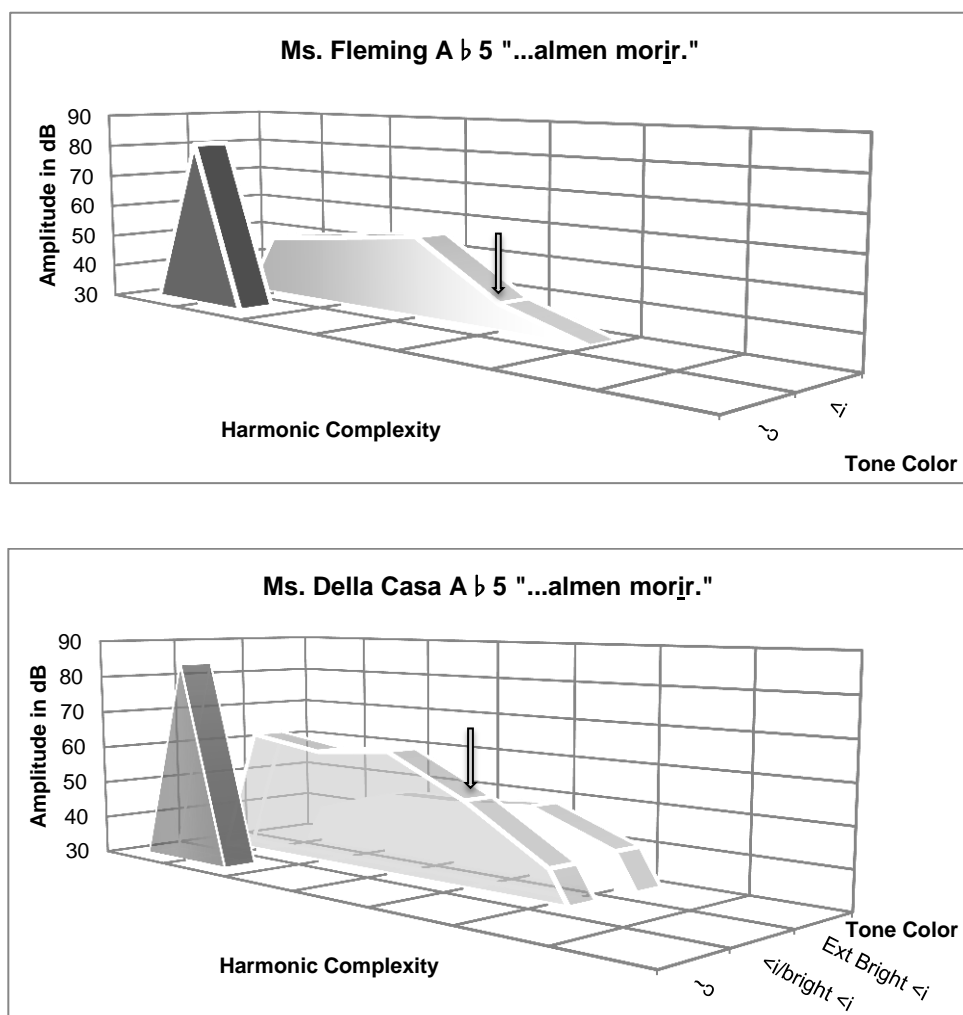


Figure 27: Parsed spectrum graphs of pitch A 5 from an excerpt of m. 36 of Renée Fleming's (top) and Lisa Della Casa's (bottom) recordings of Mozart's "Porgi amor" from *Le Nozze di Figaro*. Arrow indicates the area above which harmonics fall within the critical band. **Source:** Author's analysis of excerpt from (top) Renée Fleming, *Renée Fleming Signatures: Great Opera Scenes*, London Symphony Orchestra, conducted by George Solti, London Records, 1997, and (bottom) W.A. Mozart, *Mozart: Le Nozze di Figaro*, Vienna Philharmonic, conducted by Erich Kleiber, Decca, 1955. Analysis by author. Amplitudes in dB are relative, not calibrated.

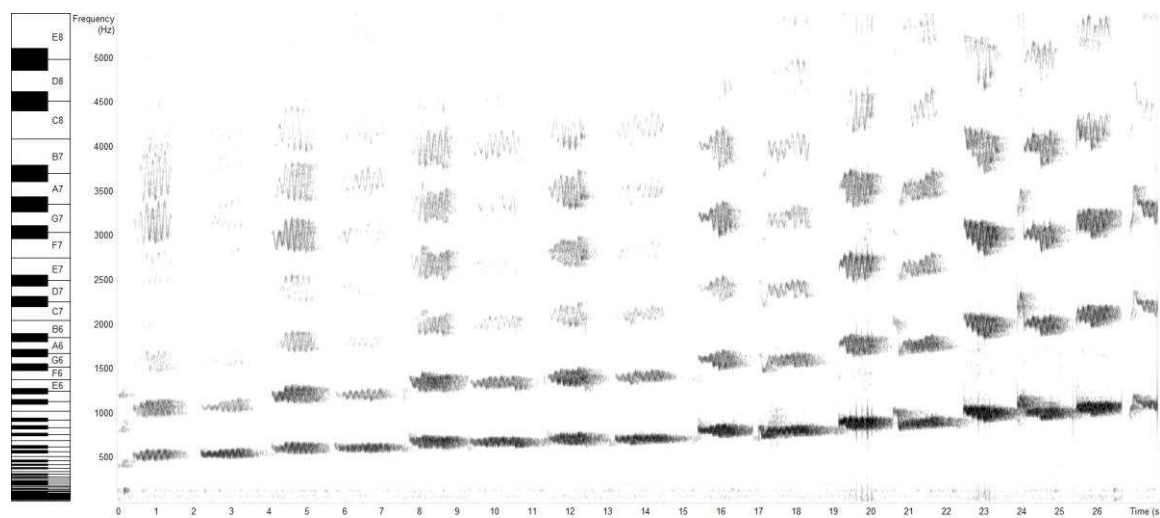


Figure 28: Spectrogram of a soprano singing a major scale C5-C6, alternating between a resonant vocal tract shape (of her choosing) and vocal tract shape of a speech level [u]. **Source:** Author's analysis.

c

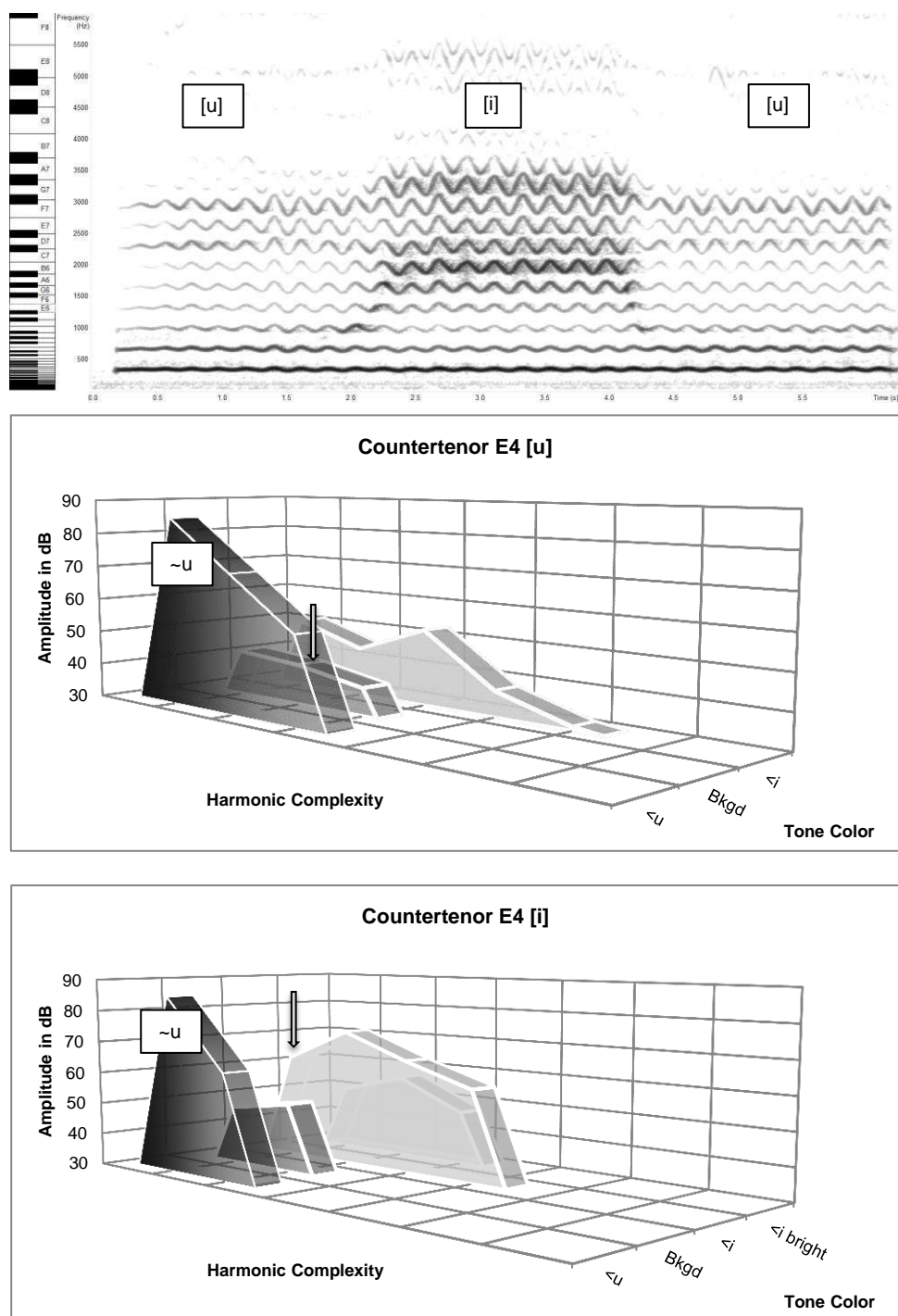


Figure 29: Spectrogram of a countertenor singing [u], [i], [u] from left to right (top), parsed spectrum graphs of the [u] (middle) and [i] (bottom). Arrow indicates the area above which harmonics fall within the critical band. The overlaid text box indicates the obvious true fundamental is perceptually separable from its local spectral segment.
Source: Analysis by author. Amplitudes in dB are relative, not calibrated.



Figure 30: m. 46-49 of Richard Strauss, “Beim Schlafengehen,” from *Vier Letzte Lieder*. **Source:** Author’s edition from *Vier Letzte Lieder* by Richard Strauss, Hermann Hesse, and Joseph Von Eichendorff © Copyright 1950 by Boosey & Co. Ltd. International Copyright Secured. All Rights Reserved. Reprinted by permission.

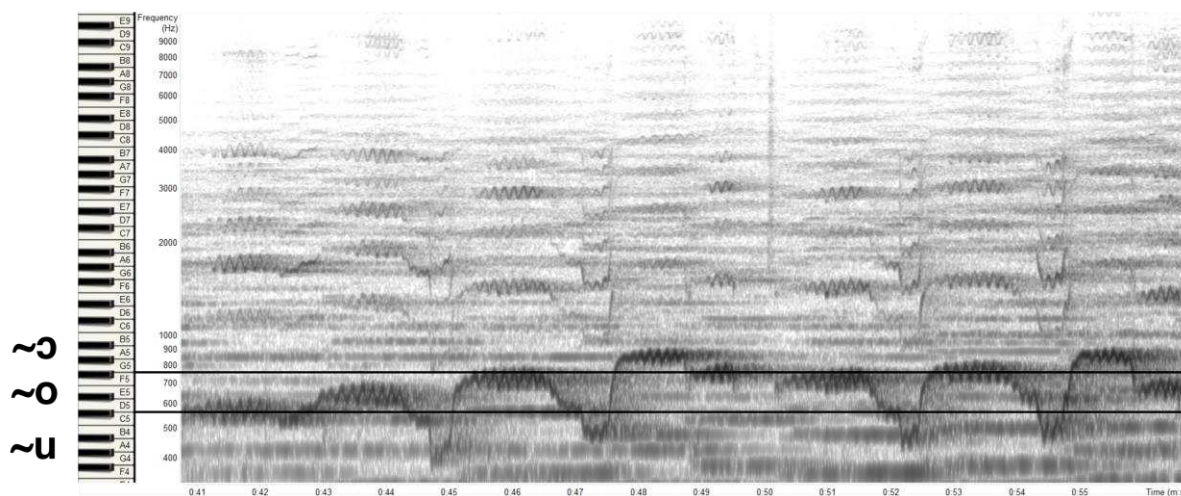
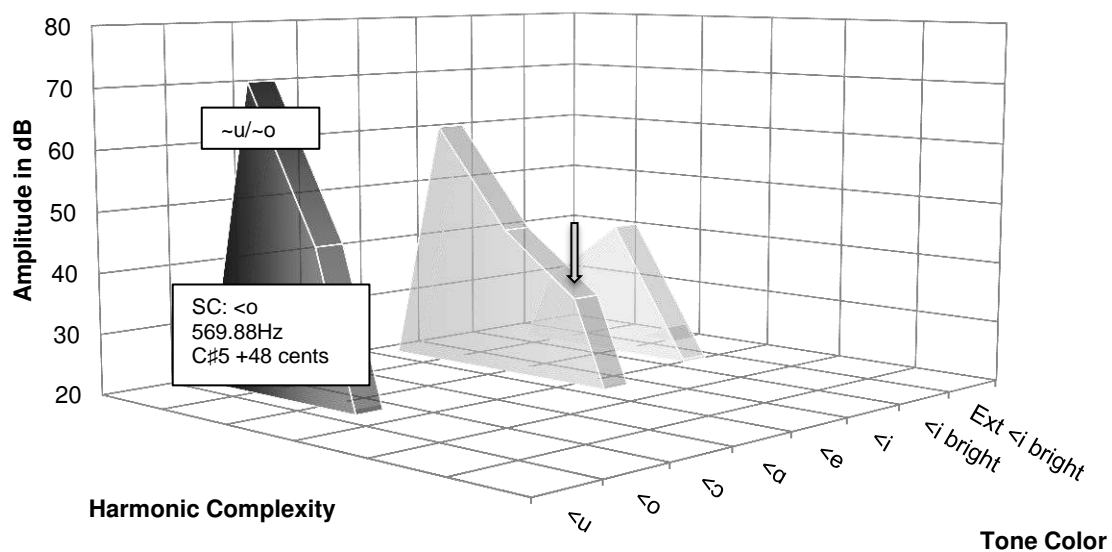
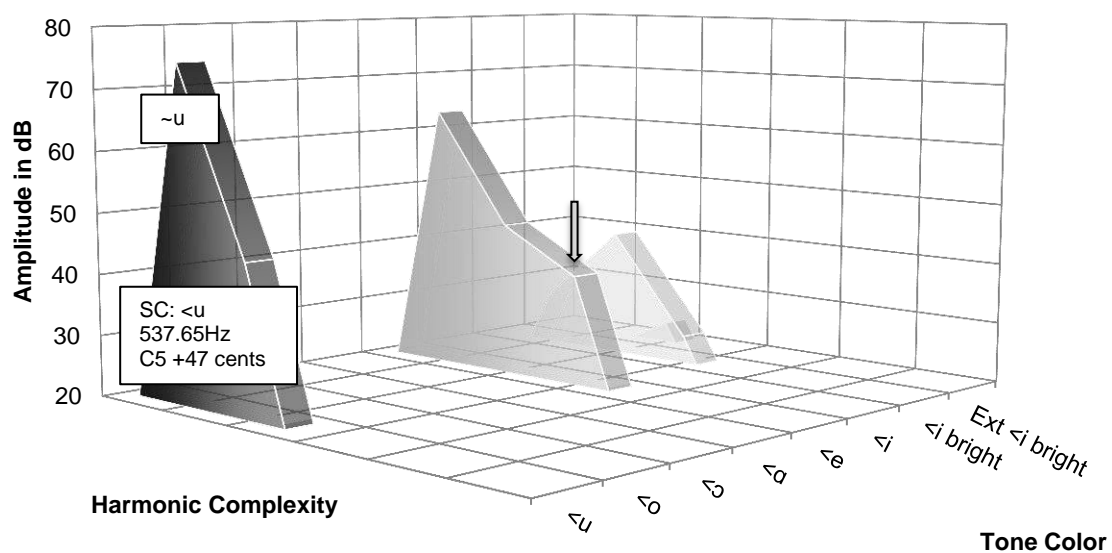


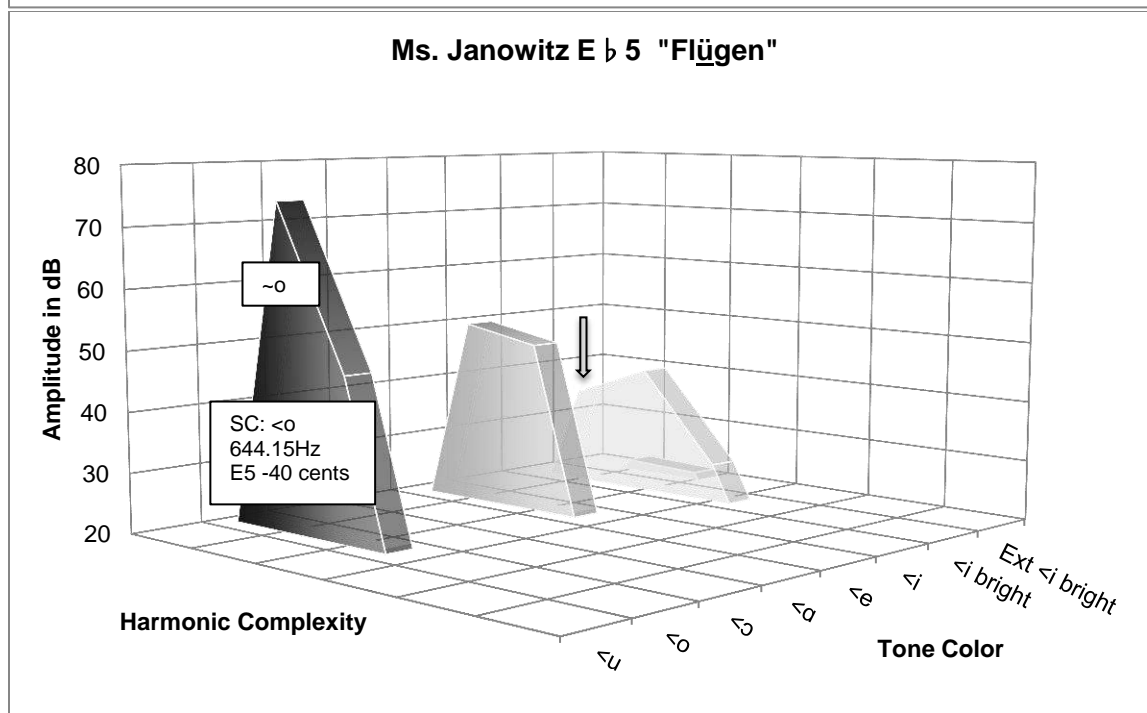
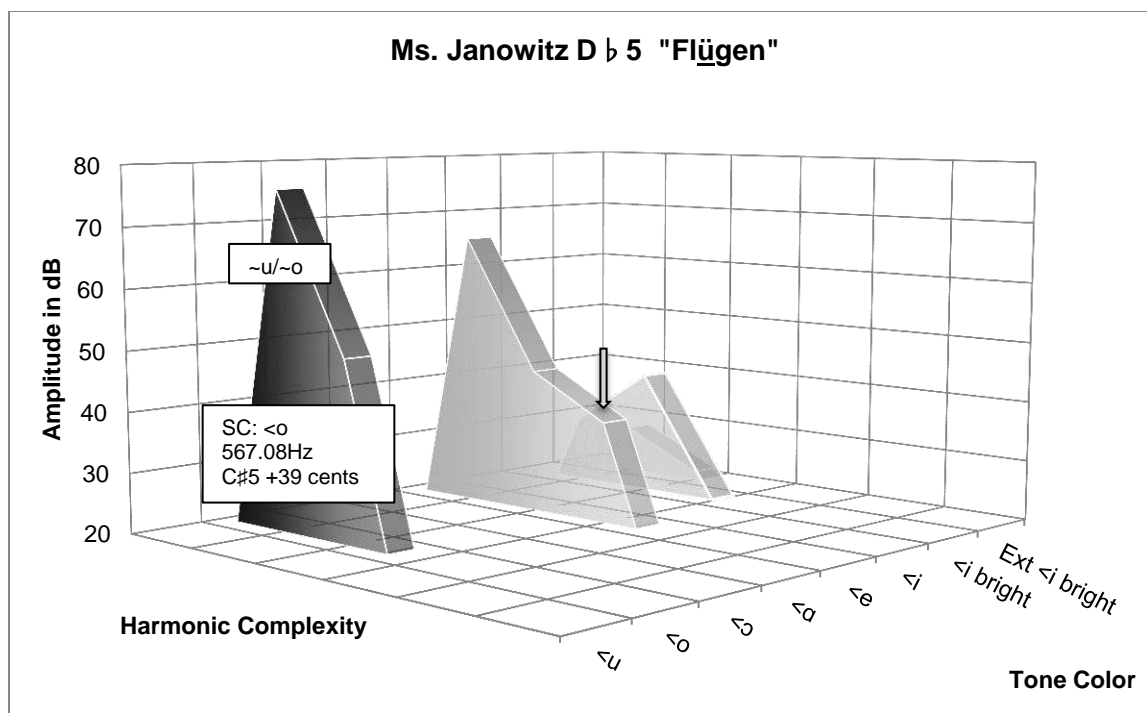
Figure 31: Spectrogram of m. 46-49 of Ms. Gundula Janowitz's performance of "Beim Schlafengehen" from Richard Strauss' *Vier Letzte Lieder*. Note that the vertical axis displays frequency logarithmically to aid aligning the fundamental (at the bottom of the spectrogram) with the pitches on the piano keyboard (left). Areas of absolute spectral tone color covered by the fundamental are indicated below (~u), between (~o), and above (~o) the horizontal black lines. **Source:** Gundula Janowitz, *Vier Letzte Lieder*, Berliner Philharmoniker, conducted by Herbert von Karajan, Deutsche Grammophon, 1971.

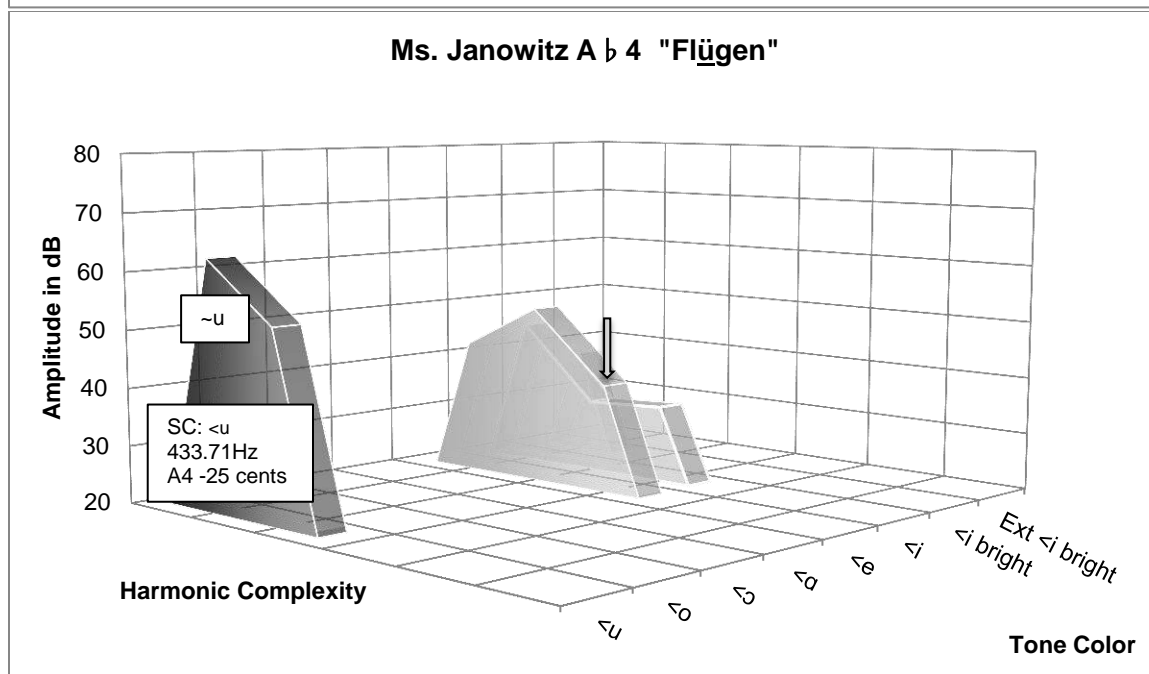
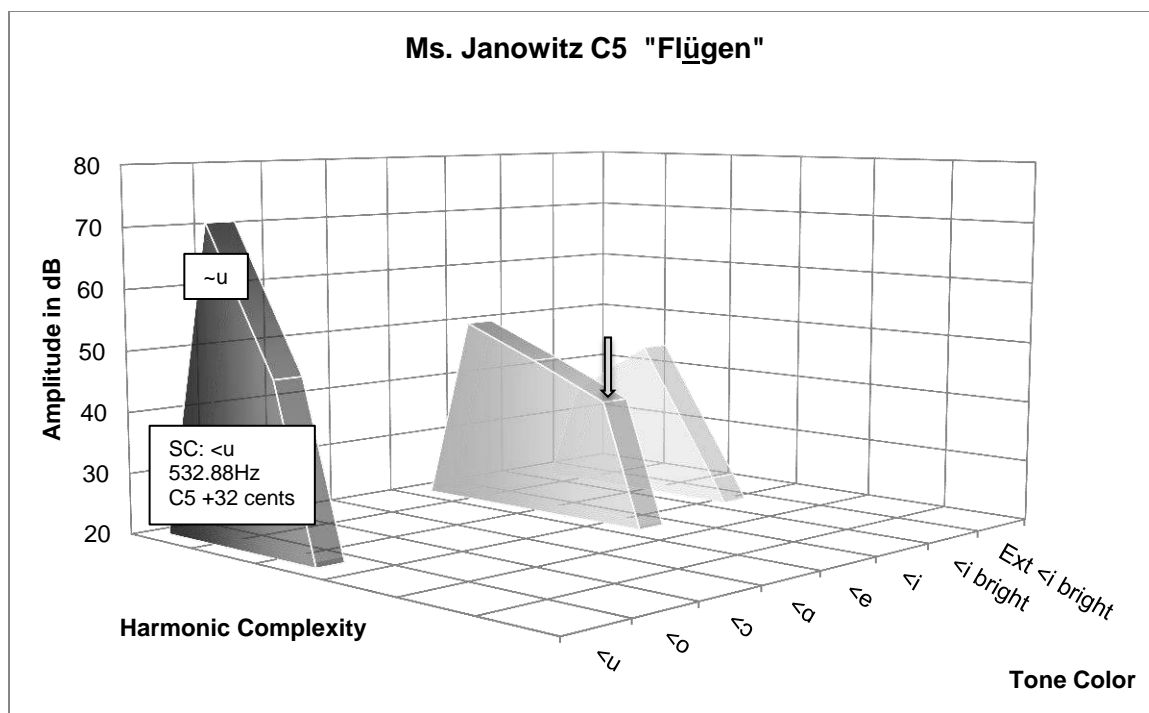
Ms. Janowitz D ♭ 5 "Flügen"

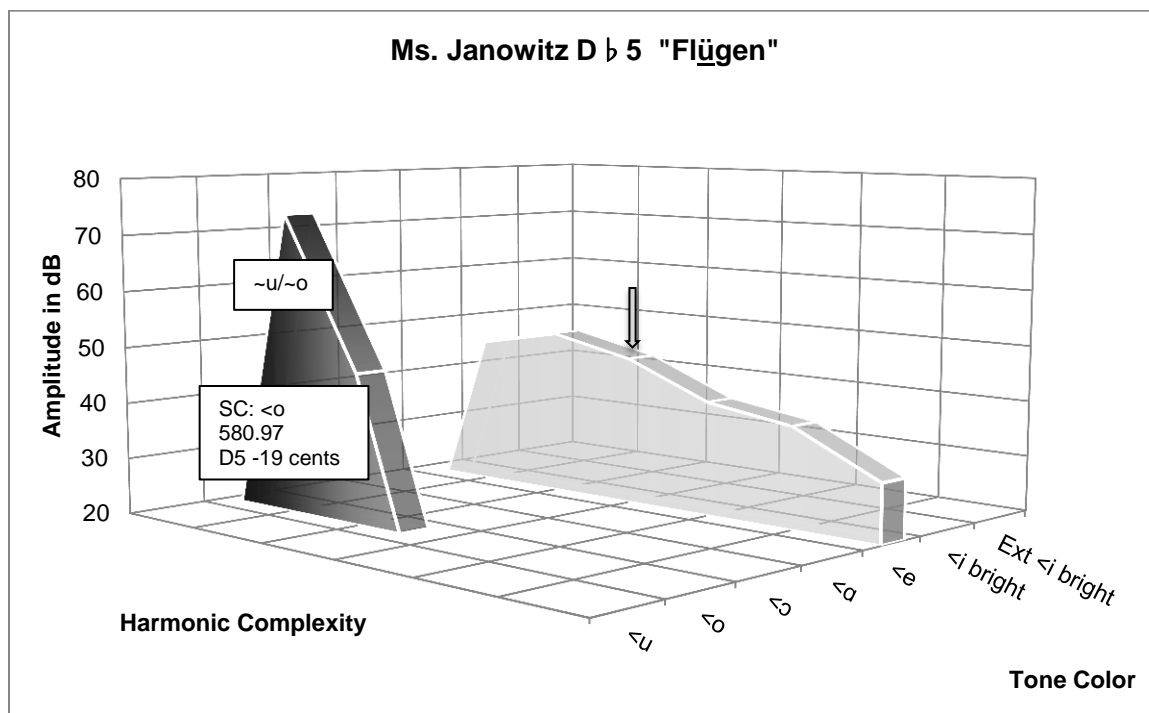
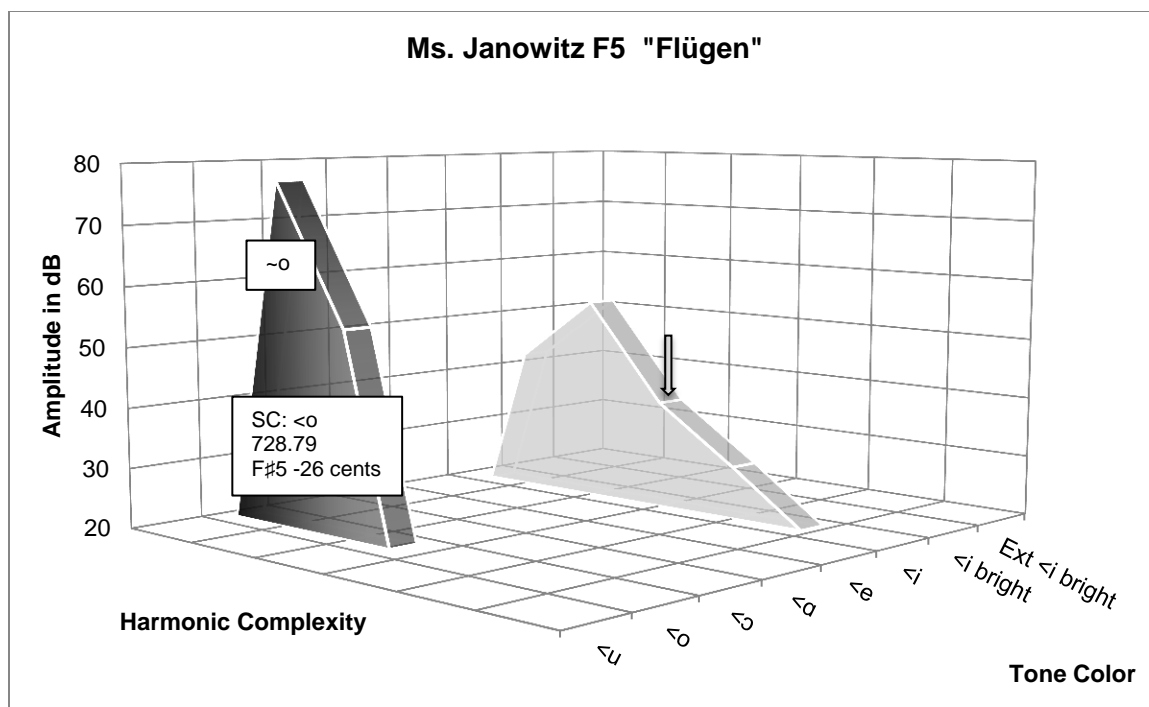


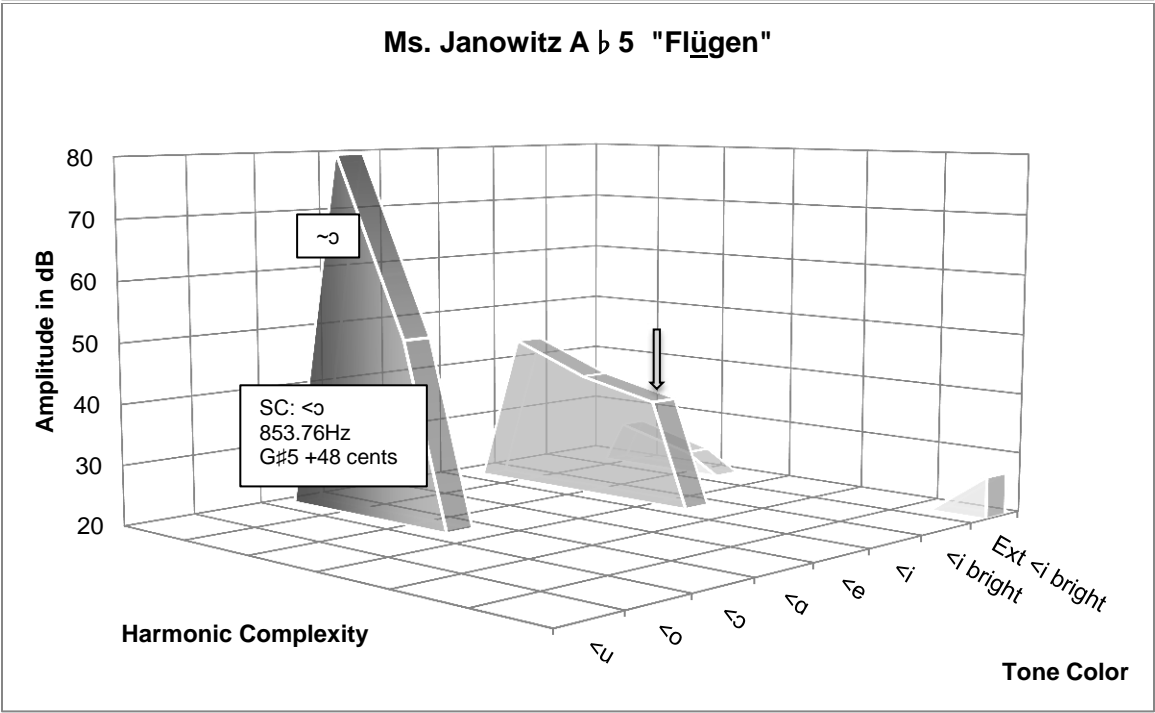
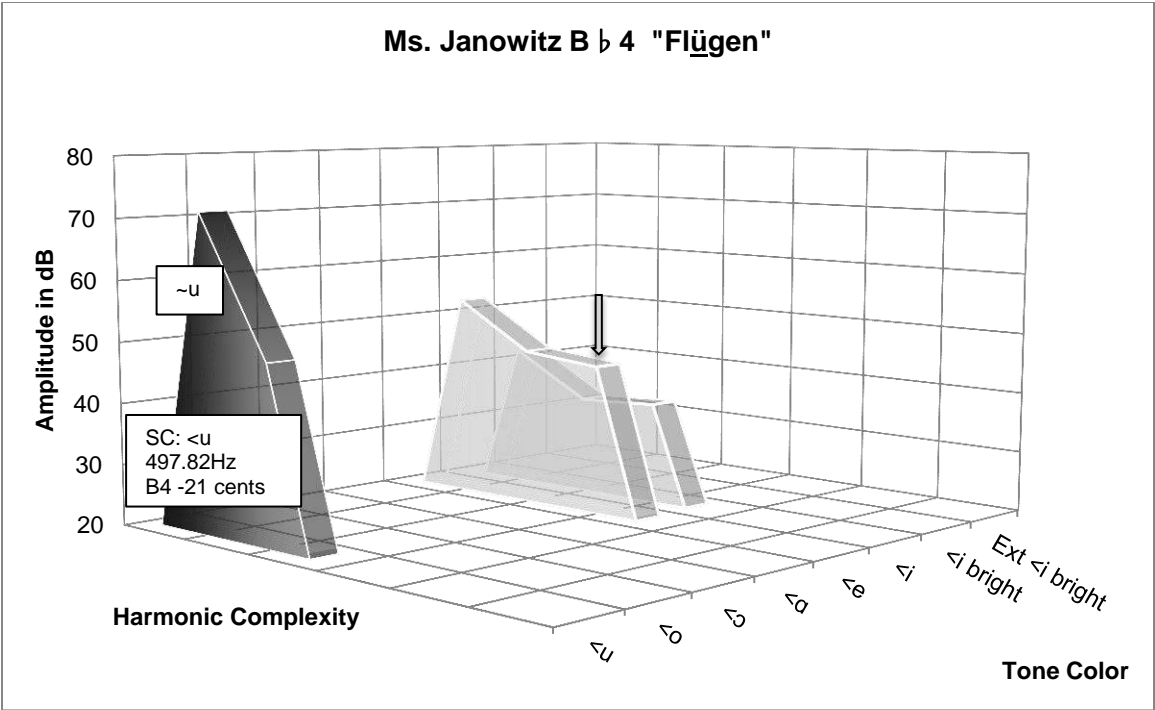
Ms. Janowitz C5 "Flügen"











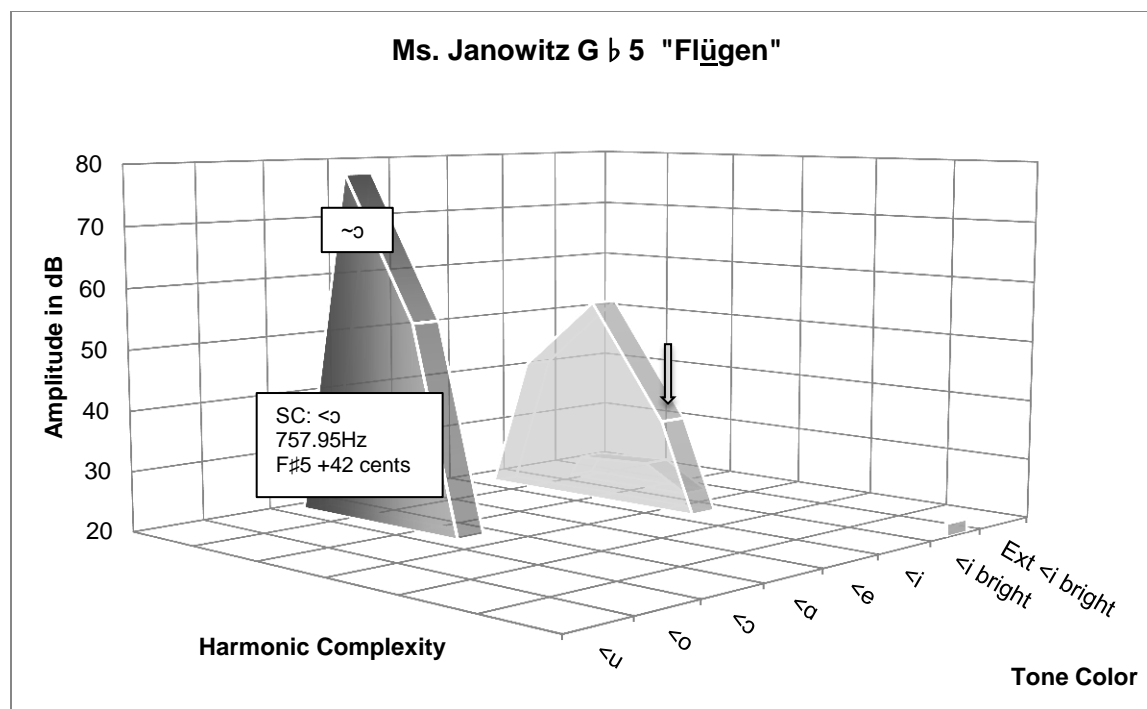


Figure 32: Parsed spectrum graphs for each note of the melisma on the word “Flügen” from m. 46-47 of Richard Strauss, “Beim Schlafengehen,” *Vier Letzte Lieder*. Values determined by the long term average spectrum (LTAS) of each note. Upper boxed text is the absolute spectral tone color of the first harmonic. Lower boxed text is the spectral centroid value (SC) for the first two harmonics **Source:** Author’s analysis using Praat (to determine spectral centroid values) based on Gundula Janowitz, *Vier Letzte Lieder*, Berliner Philharmoniker, conducted by Herbert von Karajan, Deutsche Grammophon, 1971. Amplitudes in dB are relative, not calibrated.

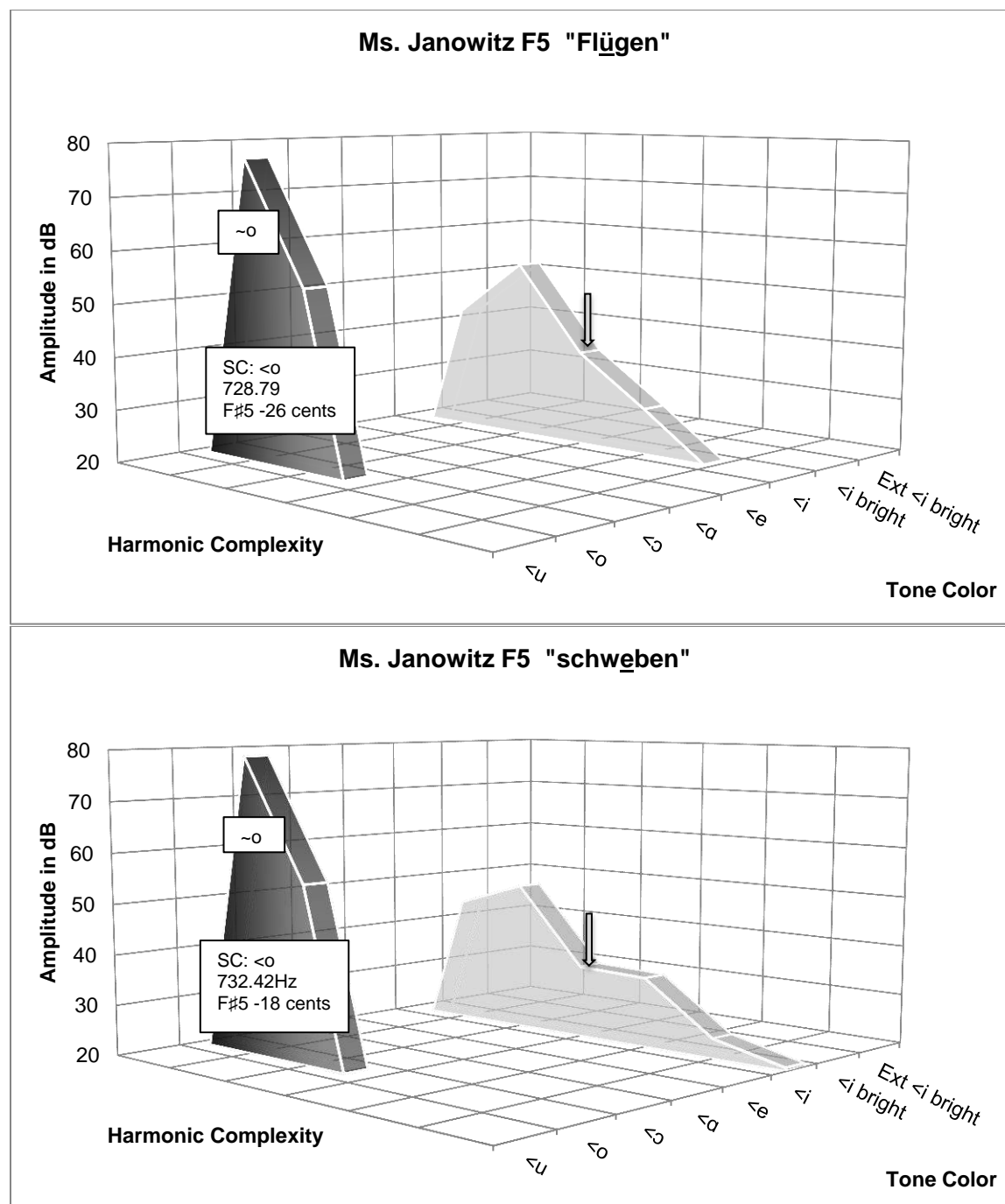


Figure 33: Parsed spectrum graphs for the F5 on the word “Flügen” ([y]) from m. 47 (top) and F5 on the word “schweben” ([e]) of Richard Strauss, “Beim Schlafengehen,” *Vier Letzte Lieder*. Values determined by use of the long term average spectrum (LTAS) of each note. Arrows indicate point above which harmonics fall into the critical band. **Source** Author’s analysis using Praat (to determine spectral centroid values) based on Gundula Janowitz, *Vier Letzte Lieder*, Berliner Philharmoniker, conducted by Herbert von Karajan, Deutsche Grammophon, 1971. Amplitudes in dB are relative, not calibrated.

Bibliography

- Aryafar, Kamelia and Ali Shokoufandeh. "Music Genre Classification Using Explicit Semantic Analysis." *MIRUM'11* (Nov. 30, 2011): 33-37.
- Ashcraft, Mark H. *Cognition*. Upper Saddle River, NJ: Pearson Prentice Hall, 2006.
- Balzano, Gerald J. "What Are Musical Pitch and Timbre?" *Music Perception: An Interdisciplinary Journal* 3/3 (1986): 297-314.
- Barry, William J. and Jürgen Trouvain, "The International Phonetic Alphabet: Do we need a symbol for a central open vowel?" *Journal of the International Phonetic Association* 38/3 (2008): 349-357, here 350.
- Becker, Christopher J. and Michael D. Hall. "Effect of Fundamental Frequency Removal and Low-Pass Filtering on Pitch Comparisons." *Psychomusicology: Music, Mind, and Brain* 24/3 (2014): 195-206.
- Beckers, Gabriël J.L., Brian S. Nelson, and Roderick A. Suthers. "Vocal-Tract Filtering by Lingual Articulation in a Parrot." *Current Biology* 14 (September 7, 2004): 1592-97.
- Bele, Irene Velsvik. "Dimensionality in Voice Quality." *Journal of Voice* 21/3 (2007): 257-272.
- Bizley, Jennifer, Kerry M. M. Walker, Andrew J. King, and Jan W. H. Schnupp. "Spectral timbre perception in ferrets: Discrimination of artificial vowels under different listening conditions." *Journal of the Acoustical Society of America* 133/1 (January 2013): 365-76.
- Borden, George A., Katherine S. Harris, and Lawrence J. Raphael. *Speech Science Primer: Physiology, Acoustics and Perception of Speech*. Philadelphia: Lippincott Williams & Wilkins, 1994.
- Bozeman, Kenneth W. *Practical Vocal Acoustics: Pedagogical Applications for Teachers and Singers*. Hillsdale, New York: Pendragon, 2013.
- Burgoyne, John Ashley, and Stephen McAdams. "A Meta-analysis of Timbre Perception Using Nonlinear Extensions to CLASCAL." *Computer Music Modeling and Retrieval: Sense of Sounds* 4969 (2008): 181-202.
- Coen, Michael H., Houri K. Vorperian, and Raymond D. Kent. "High fidelity analysis of vowel acoustic space." *The Journal of the Acoustical Society of America* 137/4 (2015): 2305.
- Cogan, Robert and Pozzi Escot. *Sonic Design: The Nature of Sound and Music*. Englewood Cliffs, New Jersey: Prentice-Hall, 1976.
- Cogan, Robert. *Music Seen, Music Heard: a picture book of musical design*. Cambridge: Publication Contact International, 1998.

- _____. *New Images of Musical Sound*. Cambridge: Harvard University Press, 1984.
- _____. *The sounds of song: a picture book of music for voice*. Cambridge: Publication Contact International, 1999.
- _____. "Toward a Theory of Timbre: Verbal Timbre and Musical Line in Purcell, Sessions, and Stravinsky." *Perspectives of New Music* 8/1 (Autumn - Winter, 1969): 75-81.
- Coutinho, Eduardo and Angelo Cangelosi. "The Use of Spatio-temporal Connectionist Models in Psychological Studies of Musical Emotions." *Music Perception* 27/1 (2009): 1-15.
- Denes, Peter B., and Elliot N. Pinson. *The Speech Chain: The Physics and Biology of Spoken Language*. New York: W.H. Freeman, 1993.
- Doscher, Barbara M. *The Functional Unity of the Singing Voice*. London: Scarecrow, 1994.
- Dyson, Freeman. *The Scientist as Rebel*. New York: New York Review of Books, 1995.
- Engel, G. "Über den Begriff der Klangfarbe." *Phil. Vorträge, Berlin, Neue Folge* II/12 (1886): 311-355.
- Fales, Cornelia. "The Paradox of Timbre." *Ethnomusicology* 46/1 (2002): 56-95.
- Fahey, Richard P. "The missing fundamental in vowel height perception." *Perception and Psychophysics* 58/5 (1996): 725-33.
- Fitch, Hollis L., Joseph J. Kupin, Irving J. Kessler, and James DeLucia. "Relating articulation and acoustics through a sinusoidal description of vocal tract shape." *Speech Communication* 39 (2003): 243-68.
- Giordano, Bruce, L., and Stephen McAdams. "Sound Source Mechanics and Musical Timbre Perception: Evidence From previous Studies." *Music Perception: An Interdisciplinary Journal* 28/2 (December 2010): 155-68.
- Grey, John M. "Multidimensional perceptual scaling of musical timbres." *Journal of the Acoustical Society of America* 61/5 (May 1977): 1270-77.
- Hanninen, Dora A. *A Theory of Music Analysis: On Segmentation and Associative Organization*. Rochester: University of Rochester Press, 2012.
- Harris, John. "Segmental Complexity and Phonological Government." *Phonology* 7/2 Phonological Government (1990): 255-300.
- Helmholtz, Hermann L. F. *On the Sensations of Tone as a Physiological Basis for the Theory of Music, Fourth edition (1877)*. Translated by Alexander J. Ellis. New York: Longmans, Green, and Co., 1912.

- Houtsma, A. J. M. and J. Smurzynski. "Pitch Identification and Discrimination for Complex Tones with Many Harmonics." *The Journal of the Acoustical Society of America* 87/1 (1990): 304-310.
- Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. "The Internal Structure of Phonological Elements: A Theory of Charm and Government." *Phonology Yearbook* 2 (1985): 305-28.
- Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. "Constituent Structure and Government in Phonology." *Phonology (Phonological Government)* 7/2 (1990): 193-231.
- Kaze, Jeffrey W. "Metaphony and Two Models for the Description of Vowel Systems." *Phonology* 8/1 (1991): 163-70.
- Kewley-Port, Diane and Yijian Zheng. "Auditory models of formant frequency discrimination for isolated vowels." *Journal of the Acoustical Society of America* 103/3 (1998): 1654-66.
- Kewley-Port, Diane, Yijian Zheng, and Amy T. Neel. "Fundamental frequency effects on thresholds for vowel formant discrimination." *Journal of the Acoustical Society of America* 100/4, Pt. 1 (October, 1996): 2462-70.
- Labuschagne, Ilse B. and Jonan J. Hanekom. "Preparation of stimuli for timbre perception studies." *Journal of the Acoustical Society of America* 134/3 (2013): 2256-67.
- Ladd, D. Robert, Rory Turnbull, Charlotte Browne, Catherine Caldwell-Harris, Lesya Ganushchak, Kate Swoboda, Verity Woodfield, and Dan Dediu. "Patterns of individual differences in the perception of missing-fundamental tones." *Journal of Experimental Psychology: Human Perception and Performance* 39/5 (Oct 2013): 1386-97.
- Lankeren, Hanny Van. "Voice characteristics of amateur female tenors are comparable with those of male tenors." *Logopedics Phoniatrics Vocology* 25 (2010): 51-58.
- Lass, Norman J., and Charles McGregor. Woodford. *Hearing Science Fundamentals*. St. Louis: Mosby/Elsevier, 2007.
- Mach, Ernst. "Zur Analyses des Tonempfindungen." *Sitzungsbericht Kaiserlich Akademie Wissenschaft*, 1885 92/2: 1283-89.
- McCoy, Scott. *Your Voice: An Inside View*. Delaware: Inside View Press, 2012.
- Miller, Donald. *Resonance in Singing*. Princeton: Inside View Press, 2008.
- Miller, Richard. *Solutions for Singers: Tools for Performers and Teachers*. Oxford: Oxford University Press, 2004.
- Miller, Richard. *Training Soprano Voices*. New York: Oxford University Press, 2000.

- Morrison, Geoffrey Stewart and Terrance M. Nearey. "Testing theories of vowel inherent spectral change." *Journal of the Acoustical Society of America* 122/1 (July, 2007): EL15-EL22.
- Nair, Garyth. *The Craft of Singing*. San Diego: Plural, 2007.
- Nearey, Terrance M. "Static, dynamic, and relational properties in vowel perception." *Journal of the Acoustical Society of America* 85/5 (May 1989): 2088-2113.
- Norman-Haignere, Nancy Kanwisher, and Josh H. McDermott. "Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex." *The Journal of Neuroscience* 33/50 (December 11, 2013): 19,451-19,469.
- Omori, Koichi, Ashutosh Kacker, Linda M. Carroll, William D. Riley, and Stanley M. Blaugrund. "Singing power ratio: Quantitative evaluation of singing voice quality." *Journal of Voice* 10/3 (1996): 228-35.
- Patterson, Roy D., Thomas C. Walters, Jessica K.M. Monaghan, and Etienne Gaudrain. "Reviewing the Definition of Timbre as it Pertains to the Perception of Speech and Musical Sounds." In *Neurophysiological Bases of Auditory Perception*, edited by Enrique A. Lopez-Poveda, Alan R. Palmer, and Ray Meddis, 223-33. New York: Springer, 2010.
- Paliwal, Kuldeep K. "Spectral Subband Centroid Features for Speech Recognition." *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 2* (1998): 617-20.
- Patil, Kailash, Daniel Pressnitzer, Shihab Shamma, and Mounya Elhilali. "Music in Our Ears: The Biological Bases of Musical Timbre Perception." *PLOS Computational Biology* 8/11 (November, 2012): 1-16.
- Peterson, Gordon E. and Harold L. Barney. "Control Methods Used in a Study of the Vowels." *Journal of the Acoustical Society of America* 24/175 (1952): 175-184.
- De Poli, Giovanni and Paolo Prandoni. "Sonological models for timbre characterization." *Journal of New Music Research* 26/2: 170-97.
- Plomp, Reinier. *Experiments on Tone Perception*. Soesterberg: Institute for Perception RVO-TNO, 1966.
- _____. *The Intelligent Ear: On the Nature of Sound Perception*. London: Lawrence Erlbaum, 2002.
- Pressnitzer, Daniel and Stephen McAdams. "Acoustics, psychoacoustics and spectral music." *Contemporary Music Review* 19/2 (2000): 33-59.
- Ragozzine, Frank. "Using Audacity for Demonstrations of Psychoacoustical Principles." *Teaching of Psychology* 34/4 (2012): 252-61.

- Reid, Cornelius. "Voice Science—An Evaluation." *Australian Voice* 11 (2005): page range unknown, here downloaded as PDF from the articles section of www.corneliusreid.com, 12-13.
- Reiterer, Susanne, Michael Erb, Wolfgang Grodd, and Dirk Wildgruber. "Cerebral Processing of Timbre and Loudness: fMRI Evidence for a Contribution of Broca's Area to Basic Auditory Discrimination." *Brain Imaging and Behavior* 2 (2008): 1-10.
- Renken, R, JEC Wiersinga-Post, S. Tomaskovic, and H. Duifhuis. "Dominance of missing fundamental versus spectrally cued pitch: Individual differences for complex tones with unresolved harmonics." *Journal of the Acoustical Society of America* 115/5 (May 2004): 2257-63.
- Schubert, Emery and Joe Wolfe. "Does Timbral Brightness Scale with Frequency and Spectral Centroid?" *ACTA Acustica United with Acustica* 92 (2006): 820-25.
- Smalley, Denis. "Defining timbre — Refining timbre." *Contemporary Music Review* 10/2 (1994): 35-48.
- Steege, Benjamin. *Helmholtz and the Modern Listener*. Cambridge: Cambridge University Press, 2012.
- Story, Brad H., Ingo Titze, and Eric A. Hoffman. "Vocal tract area functions from magnetic resonance imaging." *Journal of the Acoustical Society of America* 100/1 (July, 1996): 537-54.
- Stumpf, C. *Tonpsychologie, Vol. 2*. Leipzig: Verlag S. Hirzel, 1890.
- Strange, Winifred, James J. Jenkins, and Thomas L. Johnson. "Dynamic specification of coarticulated vowels." *Journal of the Acoustical Society of America* 74: 695-705.
- Suied, Clara, Trevor R. Agus, Simon J. Thorpe, Nima Mesgarani, and Daniel Pressnitzer. "Auditory gist: Recognition of very short sounds from timbre cues." *Journal of the Acoustical Society of America* 135/3 (March 2014): 1380-91.
- Sundberg, Johan. "Perceptual Aspects of Singing." *Journal of Voice* 8/2: 106-122.
- _____. *The Science of the Singing Voice*. Dekalb: Northern Illinois University Press, 1987.
- Titze, Ingo and Katherine Verdolini Abbott. *Vocology: The Science and Practice of Voice Habilitation*. Salt Lake City: National Center for Voice and Speech, 2012.
- Titze, Ingo. *Principles of Voice Production, 2nd Edition*. Iowa City: National Center for Voice and Speech, 2000.
- _____. "Another Incremental Step In Reviving And Revising Coffin's Favorable Vowel Chart." *Journal of Singing* 65/3 (Jan/Feb 2009): 329-31.

Vaggione, Horacio. "Timbre as syntax: A spectral modeling approach." *Contemporary Music Review* 10:2 (1994): 73-83.

Vassilakis, Pantelis N., and Roger A. Kendall. "Psychoacoustic and cognitive aspects of auditory roughness: definitions, models, and applications." *Proc. SPIE 7527, Human Vision and Electronic Imaging XV*, 75270O (February 17, 2010): doi:10.1117/12.845457.

Vennard, William. *Singing: The Mechanism and the Technic*. New York: Carl Fischer, 1967.

Winckel, Fritz. *Music, Sound and Sensation: A Modern Exposition*. Translated by Thomas Binkley. New York: Dover, 1967.

Zacharakis, Asterios, Konstantinos Pasiadis, and Joshua D. Reiss. "An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates." *Music Perception: An Interdisciplinary Journal* 31/4 (April 2014): 339-58.