UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INNER TUBING, OUTER TUBING: INTEGRATING THE ESTILL VOICE MODEL AND TITZE'S NONLINEAR SOURCE-FILTER COUPLING INTO SAXOPHONE VOICING INSTRUCTION

A DOCUMENT

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF MUSICAL ARTS

By

HEE CHENG SEAH
Norman, Oklahoma
2026

INNER TUBING, OUTER TUBING: INTEGRATING THE ESTILL VOICE MODEL AND TITZE'S NONLINEAR SOURCE-FILTER COUPLING INTO SAXOPHONE VOICING INSTRUCTION

A DOCUMENT APPROVED FOR
THE SCHOOL OF MUSIC

BY THE COMMITTEE CONSISTING OF

Dr. Jonathan Nichol, chair

Dr. Valerie Watts, co-chair

Dr. Leslie John Flanagan

Dr. Randall Hewes

ABSTRACT

This document discusses the limitations of existing saxophone voicing pedagogy and encourages an altered approach incorporating principles and research from the Estill Voice Model and Ingo Titze's concept of nonlinear source-filter model. Through explanation and application of EVM and Titze's concept, saxophonists will develop an anatomically-informed understanding of voicing, leading to more accurate self-diagnosis of oral cavity shape when encountering difficulties as well as improved control of tone, intonation, resonance, and altissimo register.

This document uses an implicit-explicit framework by integrating EVM principles with Donald Sinta's saxophone voicing exercises. First, this document outlines each of the 13 Compulsory Figures used in EVM. Next, it discusses which of the Figures are relevant to saxophone voicing, and how they influence aspects of sound production. Finally, this document provides voicing exercises incorporating principles from EVM. This approach integrates vocal tract mechanics, tone color control, and altissimo stability into a new method of saxophone voicing instruction.

ACKNOWLEDGEMENTS

To my *kaki-langs*, Ho Jin Xian, Lin Gang, Melissa Lim Yun Hao, Sng Yiang Shan, Yao Xing Ying, and my sister, Seah Hee An, for the endless karaoke sessions.

To my mentors, Mr. Boon Hui Quek, Dr. Jeffrey Loeffert, Dr. Jonathan Nichol, Dr. Leslie Kah Ho Wong, and Dr. Zechariah Toh Chai Goh, for instilling great love of music and pedagogy, and enriching my life with the best music educational experiences.

To my parents, Goh Lee Meng and Seah Jin Yong, for their beautiful singing voices and whistle tones.

To my husband, Dr. Parker Fritz, for your endless love and support, and the countless intellectual conversations we share.

Lastly, to any serious, developing saxophonist who is feeling frustrated while learning voicing, I hope this document gives you the information you need to confidently get started.

TABLE OF CONTENTS

ABSTRACT ... iv

ACKNOWLEDGEMENTS ... v

CHAPTER 1. DIFFICULTIES IN SAXOPHONE VOICING PEDAGOGY ... 1

What is Voicing? ... 1

Why Do We Need to Improve Saxophone Voicing Pedagogy? ... 3

Why Apply Vocal Pedagogy to Saxophone Voicing Pedagogy? ... 7

CHAPTER 2. SOUND PRODUCTION: THE BODY AND THE SAXOPHONE ... 11

Why Move Towards a Nonlinear Source-Filter Model? ... 11

How Can the Nonlinear Source-filter Model Benefit Saxophone Voicing? ... 12

Why Should Saxophonists Study the Estill Voice Method? ... 18

A New Pedagogical Approach ... 21

CHAPTER 3: EVM APPLIED TO SAXOPHONE VOICING EXERCISES ... 23

Estill Voice Model Principles ... 23

The 13 Compulsory Figures ... 24

Relevant Compulsory Figures in Saxophone Voicing ... 28

EVM Interpretation of Sinta's Voicing Exercises ... 31

Sinta's Modes ... 32

The First-Mode: Introduction to Saxophone Voicing ... 33

Primary Figures in the First-Mode ... 34

Secondary or Conditional Figures in the First-Mode ... 36

Low-Applicability Figures in the First-Mode ... 37

The Second-Mode: Isolating the Next Partial ... 38

Primary Figures in the Second-Mode ... 39

Secondary Figures in the Second-Mode ... 40

Low-Applicability Figures in the Second-Mode ... 41

The Third and Fourth-Modes: Discovery, Exaggeration, and Experimental Exploration ... 42

Primary Figures in the Third and Fourth-Modes ... 43

Secondary Figures in the Third and Fourth-Modes ... 44

Low-Applicability Figures in the Third and Fourth-Modes 45
Combining Modes: Aural Imagery and Tunes 46
Primary Figures in Combining Modes 48
Secondary Figures in Combining Modes 49
Low-Applicability Figures in Combining Modes 50
Altissimo Technique as Integration of Voicing Mechanism 50
Primary Figures in Altissimo 51
Secondary Figures in Altissimo 51
Low-Applicability Figures in Altissimo 52
Pedagogical Implications 53
CHAPTER 4. CONCLUSION 54
APPENDIX: FIGURES IN SINTA’S MODES 58
BIBLIOGRAPHY 59

CHAPTER 1. DIFFICULTIES IN SAXOPHONE VOICING PEDAGOGY

What is Voicing?

Broadly, voicing refers to the sounds produced when people speak or sing.[1] In phonetics, it specifically describes sound production that occurs when the vocal folds vibrate in a particular way.[2] In saxophone pedagogy, voicing refers to the awareness and control of the muscles and soft, flexible tissue in the oral cavity and vocal tract while playing the instrument.[3] Common examples and analogies for saxophone voicing include singing, humming, and whistling. All of these definitions point to a single idea: vocal sound is produced through phonation. In order to improve saxophone voicing pedagogy, students must internalize how they use their oral cavity and consider the resulting effects on sound control.

Many developing saxophonists find voicing difficult because it is an internal skill that relies on subtle tongue, throat, and air adjustments they cannot yet perceive or control. As a result, they often struggle to execute voicing exercises, such as overtone matching or altissimo notes, successfully in their first attempts. Although many developing saxophonists achieve a representative sound early in the learning process, particularly in the standard range, this does not apply to fine sound control, detailed tuning, or altissimo flexibility.[4] Playing the saxophone demands continual adjustments of the mouth, tongue, and throat that directly affect range, intonation and dynamics; some oral shapes are learned quickly, but others remain elusive and

[1] "Voice," *Cambridge Dictionary Online*.
[2] *Ibid*.
[3] Sinta, *Voicing*, 2.
[4] "What Wind Instrument Should I Play?," *Northwest School of Music*.

require persistence, unusual muscular exaggerations and repeated practice to build reliable muscle memory.[5]

Saxophone voicing method books consistently present overtone exercises as fundamental for developing tone quality, voicing control, and extended techniques, including the altissimo register.[6] Building on the legacy of German-American virtuoso Sigurd Raschèr, whose work effectively made altissimo playing essential for contemporary saxophonists, these materials treat the teaching of altissimo as a pedagogical necessity.[7] Because of this, overtone practice is widely encouraged as it cultivates a more resonant tone, voicing flexibility, and reliability into the altissimo register.[8] This positions voicing and vocal-tract influence alignment with the saxophone as an important component of modern saxophone technique.

Learning how to voice while playing the saxophone remains one of the most frustrating challenges for developing saxophonists because oral cavity manipulation is often unfamiliar because it is inside the body and cannot be seen through normal means. [9] To address this, students must realize how changes in the oral cavity and vocal tract affect aspects of playing the saxophone that require specific voicing configurations. My argument is saxophonists must understand how their inner tubing (vocal tract) interacts with their outer tubing (saxophone) in order to refine their voicing skills and maximize efficiency and accuracy in practice and performance.

In saxophone playing, the "inner tube" (the player's vocal tract) and the "outer tube" (the saxophone bore) function as a coupled system through oral cavity adjustment for the instrument

---

[5] Sinta, *Voicing*, 2.
[6] Loboda, *An Analysis of Overtone Production Techniques*.
[7] Raschèr, *Top Tones for The Saxophone*, Introduction.
[8] Loboda, *An Analysis of Overtone Production Techniques*.
[9] Sinta, *Voicing*, 3.

to respond as intended. This is done so when the player shapes the tongue position, throat, and oral cavity, adjusting the inner air column. Subtle movements of the larynx, such as raising, lowering, or slightly narrowing the glottal area, further adjust the effectiveness of the sound.

Why Do We Need to Improve Saxophone Voicing Pedagogy?

Existing method books regarding saxophone voicing have been helpful with the types of exercises to achieve voicing flexibility. Examples of exercises being emphasized in saxophone voicing pedagogy includes matching overtone exercises (where you finger and voice the actual note while alternating it into the fundamental fingering), pitch-bending exercises, and the common “F” trick (doing a descending pitch bend starting from F6 as this note is the bridge from tradition range into altissimo). Method books such as *Voicing* by Donald Sinta, *The Art of Saxophone Playing* by Larry Teal, and *Top Tones for the Saxophone* by Sigurd Raschér have been the cornerstones of pedagogical books utilized in high school and college instruction. The approaches created by Sinta, Raschèr, and Teal provide practical applied exercises, but they do not discuss or describe how the oral cavity shape is manipulated using anatomical terms, resulting in students having to guess how to manipulate their voicing or when they must self-diagnose issues.

Sinta’s *Voicing* is the most widely used saxophone voicing pedagogy book because it is structured like a conversation with the learner, closely mirroring an applied lesson setting. In the preliminary exercises, Sinta uses vocal exercises and analogies such as articulating aloud long vowels contained in the English alphabet, whistling a familiar tune, and singing or humming, and the “F” trick, offering evidence that vocal analogy is an important and effective method in understanding saxophone voicing.[10]

Moving forward, he introduces glottal articulation using syllables such as “kah” or “kuh” to develop the first “mode” of voicing (first overtone series from the fundamental). The goal is to jump the octave without depending on the octave key, rather using these syllables to drive the note up an octave to simulate voicing flexibility to create oral awareness.[11] When using such syllables, the most active area lies within the larynx. Sinta explains: “Avoid trying to slur upward but rather thrust the tongue into the new position for the upper note using the “kuh” or “kah” sound… remember[ing] what this oral cavity setting feels like.”[12] An explicit description with greater precision regarding the specific part of the tongue involved would more effectively illustrate how an anatomically-informed approach can prevent ambiguity about what he intended by “the tongue.”

The remainder of the book continues to strengthen the reader’s ability to succeed into the remaining higher partials of the overtone series, with the emphasis of tongue position and oral cavity, while occasionally giving quick tips through fingering venting to curb with the acoustical design of the saxophone that can impede producing certain voicing pitch.[13] While this might be useful, it still does not account for moments where you need to be relying on voicing more than fingering to fix the issue at hand. In general, although Sinta’s progression is useful for many students, he does not offer an anatomical explanation or insight into how the exercises develop.

The resources by Teal and Raschèr are quite dated, and saxophone pedagogy has advanced significantly since they were written. These texts remain widely used because they

[10] Sinta, *Voicing*, 3.
[11] *Ibid*., 11.
[12] *Ibid*.
[13] *Ibid*., 22.

established a shared technical and pedagogical baseline that still functions well. The combination of historical influence, pedagogical tradition, and solid fundamentals keeps them relevant. However, relying exclusively on these older resources is limiting and fails to account for the technical and musical demands of more contemporary repertoire.

In *The Art of Saxophone Playing*, Teal employs vocalization analogies to describe the phenomena of saxophone voicing, asserting: "saxophone tone is so similar in quality to the human voice that it is important to examine the parallel factors in tone production, both physical and mental."[14] Although Teal explicitly parallels saxophone tone to the human voice, he does not approach his explanation using concrete, anatomical terms. Nevertheless, he raises several points in his method book that ties into the vocal concepts I am using, but to be discussed in the later chapters. This example is therefore a particularly noteworthy source among saxophone voicing method books, given its explicit emphasis on the parallels between vocal and saxophone tone production.

Raschèr's *Top Tones for the Saxophone* uses singing as the first step to approaching overtone exercises, leading the reader through a process isolating the voicing position by alternatively playing and singing. He connects singing to saxophone playing not for its physical similarity, but for the way it builds aural awareness. This heightened listening helps players become more aware of their oral cavity and trains their ears for overtone matching exercises. Although this approach is effective in developing aural skills, it does not explain in anatomical terms how the oral cavity is manipulated throughout the process; it ultimately depends on hearing the sound clearly in the mind first, so it can then be translated into precise oral cavity manipulation.

---

[14] Teal, *The Art of Saxophone Playing*, 48.

A lack of anatomical understanding in saxophone voicing can significantly limit a player's technical and musical development. When students are unaware of how the embouchure, oral cavity, tongue position, and air support affect voicing, they often rely on trial-and-error or external physical movements that do not address the underlying physical processes of sound production. As a result, the tone may become unfocused, and intonation suffers since the player lacks the internal awareness needed to make precise adjustments. This gap in understanding frequently leads to bad habits, such as biting, embouchure distortion, too much or too little mouthpiece in the mouth, or being unable to play on harder reeds.[15] One of the most common examples is biting or applying excessive jaw pressure to produce or stabilize altissimo notes. Rather than coordinating voicing and air support, the player forces the sound rather than controlling it, which introduces tension and reduces flexibility.[16]

What begins as a localized misunderstanding, such as improper voicing or embouchure formation, can spread into articulation, dynamic control, and overall tonal consistency. Because the root cause is not clearly understood, students may reinforce ineffective habits through repetition, making them more difficult to correct later. In this way, a lack of anatomical awareness does not remain isolated; it contributes to a broader pattern of technical instability.

In contrast, an anatomically-focused understanding provides a strong foundation for efficient and sustainable practice. When students understand how their physical mechanisms function, they can approach practice with greater intention and focus. Instead of relying on trial-and-error, they can make targeted adjustments to airflow, tongue position, and embouchure, leading to more focused tone production and stable intonation. This efficiency not only improves technical accuracy but also reduces unnecessary physical strain.

---

15 Teal, *The Art of Saxophone Playing*, 98.
16 *Ibid*.

Anatomical awareness also enhances communication between teacher and student. With a shared, anatomically-informed vocabulary, instructors can give more precise feedback and students can better interpret and apply that guidance. This approach supports faster problem solving and encourages independence as students become more capable of identifying and addressing their own technical deficiencies. This awareness fosters a deeper connection to the body as part of the instrument.

Why Apply Vocal Pedagogy to Saxophone Voicing Pedagogy?

Although saxophonists often use vocal terms when describing sound quality or other musical ideas, the pedagogy does not fully embrace anatomical discussion when manipulating the oral cavity. Phonation occurs when air from the lungs moves through the larynx, setting the vocal folds into vibration. The vocal tract then shapes that vibration into what we recognize as the human voice. With this in mind, we can draw a direct parallel to saxophone playing: sound is produced when air passes through the reed, causing it to vibrate and generate tone. Teal notes in *The Art of Saxophone Playing* that the reed and the vocal folds in singing serve essentially the same function.[17]

In *Principles of Voice Production*, Ingo Titze likewise compares the voice to reed instruments, explaining that, unlike many reed or brass instruments whose oscillators are strongly governed by the resonating air column, the voice operates with variable degrees of coupling between the sound source and the vocal tract.[18] He further observes that simple reed

[17] Teal, *The Art of Saxophone Playing*, 48.
[18] Titze, *Principles of Voice Production*, 105.

models reproduce key oscillatory behaviors seen in woodwind instruments, where reed vibration always depends on the attached instrument body, the acoustic counterpart of the vocal tract.[19]

Titze explains that the voice differs from most reed and brass instruments because the strength of the interaction between its sound source and its resonating system changes with the task, such as speaking, singing high notes, or forming vowels.[20] In reed and brass instruments, the primary sound source (the reed or the lips) is largely governed by the acoustic properties of the air column in the instrument.[21] For saxophone voicing, the main sound source is the reed-lip system, which must be balanced with appropriate lip pressure to vibrate efficiently, and the resonator is the saxophone's air column. These two elements operate as a relatively fixed system once the player selects a particular setup. In the human voice, the source is the vibrating vocal folds and the resonator is the vocal tract, but the degree of interaction between them can vary much more than on saxophone because speakers and singers can continually reshape the vocal tract.

Now that we have a basic comparison between vocal and saxophone voicing, we can examine why developing saxophonists often struggle with approaching altissimo study and overtone-matching exercises. Established method books use singing, whistling, and humming as analogies to begin voicing study. Although a deep understanding of phonation and the anatomical role of the oral cavity are not often addressed in saxophone pedagogy, vocal pedagogy often does a better job at incorporating anatomical terms.

An established vocal pedagogical method that does this well is the Estill Voice Model (EVM). The principles of EVM teaches awareness of the vocal anatomy and its role in producing

[19] Titze, *Principles of Voice Production*, 94.
[20] *Ibid*., 105.
[21] *Ibid*.

the exact sound the vocalist desires. The application of its principles into saxophone voicing pedagogy will lead theoretical concepts into applied actions. An article from the American Choral Directors Association highlights a key concern about how vocal anatomy is taught in choir: "In chorus, it would probably help a lot if teachers talked about the physiology of the voice. I think choral teachers underestimate how much students are able to understand."[22] Lynn Swanson, a leading figure of the Choral Foundation in Atlanta, affirms this point, stating, "teach students the correct terms for the [oral cavity] and then use those terms in your daily teaching as well."[23]

I grew up and completed my undergraduate degree in Singapore. There, the standard of applied instrumental teaching in higher education only began to flourish in the 2000s and initially lagged behind more established programs in countries such as the United States. At that time, instrumental pedagogy in Singapore's higher education system was limited and needed significant development. In my own experience, altissimo was not part of the curriculum when I entered the program; it was introduced during my junior year and became a formal requirement only in the final year of study. By contrast, in the United States, advanced students often begin studying altissimo as early as high school. My saxophone teaching philosophy is that you cannot teach a wind instrumentalist how to play their instrument without teaching the basic functions of the oral cavity, just as an effective vocal instructor would. To be a more well-rounded instrumental teacher, I firmly believe that instructors must incorporate teaching the function of the oral cavity in their lesson, especially with advanced instrumental students. Instructors can

---

[22] Berroth and Swanson, "*Voice Anatomy for the Choir Classroom - Demonstrations and Activities*."
[23] *Ibid*.

better prepare students with tools to recognize, diagnose, and self-correct, deepening their foundation for effective learning and ultimately teaching them how to teach themselves.

CHAPTER 2. SOUND PRODUCTION: THE BODY AND THE SAXOPHONE

Why Move Towards a Nonlinear Source-Filter Model?

Linear source-filter coupling is considered the "traditional" voice concept because early voice and speech models assumed that the sound source (the vocal folds) and the filter (the vocal tract) operated independently and interacted in a simple, linear way. This linear, independent source-filter model was mathematically tractable, fit early acoustic data reasonably well, and became widely adopted in classic work on speech and singing. It formed the foundation of decades of pedagogy, analysis, and synthesis before newer research highlighted the importance of nonlinear source-filter interactions.[24]

The linear source-filter model of voice production, a foundational concept in speech science and acoustics for many decades, is fundamentally incomplete. The vocal tract is not a passive filter operating on an independent glottal source. Instead, *Bernoulli*'*s principle* helps explain how airflow velocity and pressure differences can contribute to vocal-fold movement. Within the traditional voice model, Bernoulli's principle describes how increased airflow velocity through a narrowed glottis lowers pressure and contributes to vocal-fold closure.[25] It explains linear source-filter coupling because airflow through the glottis is treated as a mostly one-way aerodynamic process: when the glottal opening narrows, flow speeds up and pressure drops, which influences vocal-fold motion without requiring strong feedback from the vocal tract.[26]

---

[24] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."
[25] "Bernoulli Effect," *The Voice Science Lexicon*.
[26] *Ibid*.

However, the vocal tract is not a passive filter operating on an independent glottal source. *Inertive reactance* (the vocal tract's resistance to changes in airflow, caused by air inertia, which helps shape pressure, flow, and sound during vocal fold movement) generated by the epilarynx tube and modulated by subglottal (below the glottis) and supraglottal (above the glottis) tract geometry feeds back energy to the vocal folds during the open glottal phase. This feedback shapes the glottal flow waveform, alters harmonic intensities and timbre, governs phonation threshold and register stability, and is under voluntary anatomical control. When the vocal folds and vocal tract interact in this way, the voice becomes easier to produce and carries better. In contrast, if the folds just vibrate on their own without this interaction, the sound tends to be less efficient, less powerful, and more limited, even if the pitch and words are correct.[27]

How Can the Nonlinear Source-filter Model Benefit Saxophone Voicing?

The reed-bore system on reed instruments works very much like the vocal fold and vocal tract system in the human voice. In this analogy, the space in the mouth and throat in front of the reed (the oral cavity and pharynx) acts like the subglottal system in singers, while the saxophone bore functions like the vocal tract, shaping and resonating the sound that the vibrating reed (like the vocal folds) produces.[28] Likewise, the tongue-arch position (high versus low) when playing classical saxophone is similar in function to epilarynx diameter (narrow versus wide) when singing. The ideal classical saxophone tongue position features a high arch because this position increases inertive reactance of the oral cavity entry section, raises the effective coupling

[27] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."
[28] Titze, *Principles of Voice Production*, 94.

parameter at the reed, and changes how evenly the reed opens and closes in a way that boosts brighter, more projected tone at reduced airflow.

Similarly, back pressure is a familiar sensation for saxophonists, especially in the upper register, when using front fingerings, or when playing soprano saxophone. It is often interpreted as a resistance problem that needs to be overcome with more air or specific tongue positioning. However, nonlinear source-filter theory offers a more accurate explanation: this resistance is not simply an obstacle, but the result of an interaction between the sound source (the reed) and the acoustic filter (the player's vocal tract).

In traditional pedagogy, instructions such as "blow faster air" or "raise the back of the tongue" are often given as universal solutions. While these can sometimes produce the desired result, they are ultimately based on trial-and-error rather than a clear understanding of the underlying mechanism. Nonlinear source-filter coupling shows that the reed and the vocal tract are not independent systems; they influence each other in real time. The shape of the vocal tract can either stabilize or destabilize the reed's vibration, particularly in sensitive registers. For example, in the altissimo register, the instrument's natural resonances are weaker and less supportive. In this situation, the player's vocal tract must take on a more active role. If the vocal tract is not shaped appropriately, the player experiences increased back pressure and instability as this is not simply a matter of insufficient air, but rather a mismatch between the source and the filter. This is where insights from vocal pedagogy become especially valuable. Singers regularly navigate high-frequency production, and they do so not by forcing more air, but by efficiently tuning the vocal tract to align with the desired pitch and resonance. Applying this concept to saxophone playing suggests that managing back pressure is less about force and more about coordination.

In this framework, technical challenges are no longer addressed through isolated commands, but through a vocally-integrated vocabulary. The goal is to adjust the vocal tract in a way that complements the reed's natural behavior, reducing unnecessary resistance and increasing efficiency. Rather than relying on "blind faith" instructions, saxophonists can develop a more reliable and repeatable technique grounded in acoustic interaction.

For decades, the dominant framework for describing voice production was the linear source-filter theory. It described a two-stage system: first the source, meaning the vibrating vocal folds, generates a quasi-periodic airflow waveform called the glottal flow pulse; then this signal passes through the filter, the vocal tract cavities of the throat and mouth, where resonances shape it into speech sounds and musical tones.[29] In this traditional view the source and filter act independently: the vocal folds create the sound and the vocal tract only shapes it, without sending any feedback to the source.[30]

However, as researchers continued to study the linear source-filter theory, they found patterns that the linear model could not explain. It did not fully account for the remarkable acoustic power and efficiency of highly trained singers, particularly in the upper register.[31] It also failed to reliably predict abrupt, irregular behaviors such as pitch jumps and voice "cracks," particularly when harmonics neared or crossed vocal-tract formants.[32] In addition, its predictions matched adult male voices better than female and child voices, whose higher fundamental

[29] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."
[30] Fant, *Acoustic Theory of Speech Production*.; Tokuda et al., "New Evidence That Nonlinear Source-Filter Coupling Affects Harmonic Intensities and Timbre."
[31] Fant, *Acoustic Theory of Speech Production*.; Rothenberg, "Acoustic Interaction Between the Glottal Source and the Vocal Tract."
[32] Titze, "Nonlinear Source-Filter Interaction in Singing"; Titze, Worley, and Story, "Source-Vocal Tract Interaction in Female Operatic Singing," 561-72.

frequencies sit closer to formants and show more frequent source instabilities.[33] Together, these issues indicated that an essential part of the system was missing from the linear framework.

Titze and other researchers responded to these limitations by advancing a new paradigm: *nonlinear source-filter coupling*, where the vocal source and vocal tract interact continuously in both directions rather than functioning separately.[34] In this updated model, the vocal folds (source) and vocal tract (filter) depend on each other and constantly exchange influence through acoustic pressures above and below the glottis.[35] This ongoing feedback reshapes how sound is produced, stabilized, and amplified in the human body.[36]

Two acoustical concepts are central to this process: glottal impedance and vocal tract impedance. Glottal impedance refers to the effective resistance and reactance encountered by airflow from the lungs as it passes through the glottis and is determined by glottal area, tissue configuration, and subglottal conditions.[37] Vocal tract impedance is how strongly the vocal tract resists or supports sound waves at different frequencies, depending on its shape, width, and whether it is open or closed.[38] Nonlinear coupling is especially strong when the magnitude of the vocal tract input impedance is comparable to the glottal source impedance, making the glottal flow highly dependent on supraglottal and subglottal acoustical pressures.[39]

---

[33] Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises"; Titze, "The Human Voice and Its Nonlinear Physics."
[34] Tokuda et al., "New Evidence That Nonlinear Source-Filter Coupling Affects Harmonic Intensities and Timbre."
[35] Titze, "The Human Voice and Its Nonlinear Physics."; Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."; Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."
[36] Rothenberg, "Acoustic Interaction Between the Glottal Source and the Vocal Tract."; Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."
[37] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."; Tokuda et al., "New Evidence That Nonlinear Source-Filter Coupling Affects Harmonic Intensities and Timbre."
[38] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."
[39] Kob and Neuschaefer-Rube, "Vocal Tract and Subglottal Impedance in High Performance Singing."; Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."

A powerful feedback loop is established when a singer adjusts the shapes of their epilaryngeal region (the short tube of the supraglottis immediately above the vocal folds) and vocal tract so that the positive inertive reactance of the tract aligns with the glottal condition.[40] Under these circumstances, acoustic pressure variations from the vocal tract (filter) reinforce pressures on the vocal folds (source), enhancing both positive inertive reactance of the vocal folds and tonal richness.[41] This alignment effectively allows the filter's acoustic "pushback" to support the oscillation, producing a more powerful and efficient vocal output without necessarily increasing subglottal pressure.[42]

This interactive, nonlinear model resolves several longstanding misunderstood concepts in vocal production. First, the rich and complex timbres of a trained singer no longer appear to stem solely from the mechanical vibrations of the vocal folds. Instead, the vocal tract reflects energy back toward the glottis, reshapes the pattern of airflow through the folds, and adds overtones that were not initially present.[43] Second, voice cracks and sudden jumps in pitch, although seemingly unpredictable, can now be explained as shifts that occur when the formant or its harmonics move through sensitive parts of the vocal tract's resonance chamber, changing the pressure on the folds and destabilizing its vibration.[44]

Nonlinear source-filter coupling provides an explanation for the extraordinary power and efficiency possessed by well-trained singers. Modeling and experimental research suggests that

---

[40] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."; Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."
[41] Titze, "Nonlinear Source-Filter Interaction in Singing."; Titze, Worley, and Story, "Source-Vocal Tract Interaction in Female Operatic Singing and Implications for Vocal Efficiency."
[42] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."; Titze, "The Human Voice and Its Nonlinear Physics."
[43] Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."; Titze, "Nonlinear Source-Filter Interaction in Singing."
[44] Rothenberg, "Acoustic Interaction Between the Glottal Source and the Vocal Tract."; Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."

when this feedback system is efficient, the acoustic output power can increase substantially while the system simultaneously conserves glottal flow and improves efficiency by converting airflow into stable vocal control.[45] Crucially, singers can gain these benefits without working harder, suggesting that expert singers are taking advantage of helpful acoustic conditions rather than simply pushing more air or increasing pressure.[46]

In Titze's framework, phonation can first be described in general terms as a coupled system: a self-oscillating vocal fold mechanism (the oscillator) interacts with an acoustic tube that has its own resonances (the resonator).[47] When described in speech-science language, this becomes the familiar "source-filter" model, with the vocal folds and glottal airflow as the source and the vocal tract as the filter. In its simplest form, this model assumes that the filter only shapes the outgoing sound and does not significantly affect how the source vibrates, so energy appears to flow in one direction from source to filter. Over time, however, researchers found that this linear, one-way model could not fully explain many real-world vocal behaviors, including voice breaks, sudden register shifts, subharmonics, and pitch instabilities near formant frequencies (a strong resonance peak in a sound that shapes its tone color).[48] These limitations made it clear that the "one-way conversation" between source and filter needed to be reconsidered.[49] Through experiments and mathematical modeling, Titze showed that the vocal tract actively influences vocal fold vibration rather than merely filtering a fixed source signal, demonstrating that the

---

[45] Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."; Tokuda et al., "New Evidence That Nonlinear Source-Filter Coupling Affects Harmonic Intensities and Timbre."
[46] Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."; Titze, "Nonlinear Source-Filter Interaction in Singing"; Titze, "The Human Voice and Its Nonlinear Physics."
[47] Titze, *Principles of Voice Production*, 105.
[48] Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."
[49] *Ibid*.

folds and tract form a coupled oscillator-resonator system with nonlinear source-filter interaction.[50]

This theoretical advance also helps bridge scientific understanding with established vocal pedagogy. Researchers have shown this nonlinear source-filter coupling interaction in computer models and live human voice observations, explaining how established exercises work by changing how the vocal tract feedbacks to the folds.[51] Techniques intuitively used by vocalists and teachers, which were not fully understood because of a lack of physical explanation, can now be interpreted through the lens of acoustic loading (the effect of sound pressure pushing on vocal tract tissue), inertive reactance, and controlled source-filter interaction.[52] This convergence of physics and pedagogy reveals that every vocal gesture is the result of an intricate interplay between aerodynamics, acoustics, and tissue mechanics.[53]

Why Should Saxophonists Study the Estill Voice Method?

The Estill Voice Model (EVM) is an approach that can tie together vocal and saxophone pedagogies, acting as a bridge to guide the reader through the process of using vocal science in an instrumental pedagogy setting. EVM provides a general knowledge on voice anatomy that is jargon-free, making the instrumental learning process easier to understand.

---

[50] Titze, "Nonlinear Source-Filter Coupling in Phonation: Theory."
[51] Kob and Neuschaefer-Rube, "Vocal Tract and Subglottal Impedance in High Performance Singing."; Titze, "Nonlinear Source-Filter Interaction in Singing."
[52] Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."; Titze, "The Human Voice and Its Nonlinear Physics."
[53] Titze, "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises."; Titze, Worley, and Story, "Source-Vocal Tract Interaction in Female Operatic Singing and Implications for Vocal Efficiency."

Josephine Antoinette Estill, vocal pedagogue and creator of EVM, stopped and asked herself how to sing efficiently. Estill did not have a clear understanding of how she was singing or how she was making her sound beautiful.[54] What she came to understand was threefold: she did not comprehend the pedagogical texts any better than she had understood her voice teachers, there was little agreement among those sources in either terminology or concepts, and what she read did not align with her own internal experience of singing successfully.[55] From this emerges a central idea: vocal knowledge is vocal power. Without an understanding of how the voice functions, even successful singing remains unreliable and difficult to reproduce consistently. The disconnect may instead lie in a lack of clarity about what actually produces successful vocalization, rather than in the outcomes themselves.[56]

One of the foundational concepts that informed this exploration was the Speech Chain, a model describing spoken communication.[57] It begins at the linguistic level, where intent is formed; moves to the physiological level, where the brain sends motor signals to the vocal mechanism; progresses to the acoustic level, where sound waves are generated and transmitted; returns to the physiological level as the listener's auditory system converts sound back into neural signals; and concludes again at the linguistic level, where meaning is interpreted. This framework provided a structured way to think about voice production and perception.[58]

A significant challenge in researching voice quality lies in its immense variability. No two voices produce identical sounds, making it difficult to determine which qualities to study. Estill mentions a useful comparison that can be made to human faces: although no two smiles are

---

[54] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 2.
[55] *Ibid*.
[56] Estill, J. (1991) "VOICE CRAFT: An alternative method for training voices." Oral version presented to Australian and New Zealand Association of Teachers of Singing, in Sydney on 3 August 1990.
[57] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 9.
[58] *Ibid*.

exactly alike, we rely on the same underlying musculature to produce them, and we understand how those muscles function. Similarly, despite vocal diversity, shared physiological mechanisms underlie all voice production.[59]

The voice itself is a highly complex acoustic signal that varies across multiple dimensions as the physiological system generating it changes. In principle, any change in this system that affects a listener's perception can be described as a change in voice quality. These qualities arise primarily from the manipulation of structures such as the larynx and pharynx, which shape the resulting sound. Previously, a dominant framework in speech science was the Source-Filter Theory, which separates sound production into a source (vocal fold vibration) and a filter (the vocal tract shaping the sound).[60] Estill expanded on this model by introducing "Power" to represent the role of breath, creating the Power-Source-Filter (PSF) model.[61] This provided not only an acoustic explanation of voice production but also a physiological one, forming a core component of EVM. Within this framework, Estill used terminology in a precise scientific sense to describe how the human "sound machine" operates. Breath, vocal fold vibration, and the shaping actions of the throat and mouth generate both noise and periodic sound signals, which combine into a complex waveform that we perceive as voice.[62]

Despite these models, defining clear objectives in voice training remains difficult because evaluation is largely subjective. What we hear internally differs from what others perceive externally, anatomical structures are not fully visible, and many of the mechanisms involved are difficult to sense directly. Most individuals cannot easily identify or feel the specific body parts

[59] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 11.
[60] Fant, G. (1970). *Acoustic Theory of Sound Production*. The Hague: Mouton.
[61] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 14.
[62] *Ibid*., 15.

involved in actions like "projecting" the voice.[63] While EVM addresses the principles of voice production, Estill Voice Training (EVT) tackles these challenges by providing a clear vocabulary for vocal structures, along with multiple forms of feedback (what we feel, see, and hear) to help guide development. It also incorporates principles from Dynamical Systems Theory (a mathematical framework used to study things that change over time), describing habitual vocal patterns as "attractor states." These default settings are shaped not only by physiology but also by tasks and environment. Understanding these attractor states allows singers to practice more intentionally and move toward deliberate control.[64]

Estill also emphasized that singing is unnatural because it requires overriding the body's natural patterns of respiration. To sing effectively, individuals must replace automatic laryngeal behaviors that may hinder vocal production with consciously controlled actions that support healthy and efficient sound. For this reason, developing bodily awareness is essential. Singers must learn to attend to the sensory feedback their muscles provide, even though many of these actions have become automatic over time. Initially, it may be difficult to perceive or interpret these sensations, but cultivating this awareness is a critical step toward gaining control over the vocal mechanism.[65]

A New Pedagogical Approach

I propose developing a new, theoretically grounded framework for teaching saxophone voicing, built around saxophone exercises that integrate contemporary voice pedagogy.

---

[63] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 15.
[64] *Ibid.*, 37.
[65] *Ibid.*, 38.

Saxophone instruction will benefit from a systematic resource rooted in voice science. This resource will utilize established saxophone voicing exercises from Sinta's Voicing, interpreted through the principles of EVM. This voice-informed framework will benefit saxophone voicing pedagogy by bringing instrumental technique in line with the actual physiological and acoustic mechanisms of vocal production.

CHAPTER 3: EVM APPLIED TO SAXOPHONE VOICING EXERCISES

Estill Voice Model Principles

The human voice has multiple small parts that can be moved to create different voice qualities. As mentioned in the previous chapter, Estill expanded the dominant framework in speech science (Source-Filter Theory) into the Power-Source-Filter model. Power is the respiratory system, the part of the human body that produces and controls the flow of air intake and outtake; source is the True Vocal Folds coming together in the airflow to make a buzzing sound. Filter is the resonance space in the vocal tract that shapes the buzz into vowels and tone quality.[66] Recent work in Dynamical Systems Theory shows that the voice does not behave in a simple, linear way.[67] From the beginning, Estill's research also showed that voice production is non-linear. The vibration of the vocal folds (Source) can change how the breath flow (Power) behaves. Changing the shape of the Aryepiglottic Sphincter (Filter) can change the effective mass and behavior of the True Vocal Folds (Source). We can hear this non-linearity when the voice "breaks" from the chest to the head as a singer moves up the scale.[68] Because of the complexity of sound production, even if it is a simple lullaby, Estill embraced it and translated her research observations into practical exercises for training the voice. To simplify the process, she asserts that learning the voice can be broken down into three stages, drawing parallels from learning a new sport.[69]

---

[66] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 14.
[67] Kelso, J.A.S. (1995). *Dynamic patterns*: *The self-organization of brain and behavior*. Cambridge, MA: The MIT press.; Steinhauer, K., Grayhack, J., Smiley-Oyen, A., Shaiman, S. & McNeil, M. (2004). The Relationship among voice onset, voice quality, and fundamental frequency: a dynamical perspective. Journal of Voice, 18(4):432-442.
[68] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 26.
[69] *Ibid.*, 27.

First, there is *Craft*, which is the stage where you develop your technical abilities. Here you become more conscious of your voice and how to move and control its different parts, much like practicing drills that teach you to handle the ball and navigate the field. Next comes *Artistry*, which is how you use those skills in real speaking and singing. Just as each sport has its own set of rules, different styles of expression call for using your skills in a variety of ways to create different artistic effects. Finally, there is *Performance Magic*, the moment where you bring all of your practice together in front of an audience. To excel in any game, it is essential to understand the rules and master the necessary skills, so it is helpful to begin by concentrating on the craft. The Compulsory Figures are available as a guide to support your learning and mastering journey.[70]

The 13 Compulsory Figures

The 13 Compulsory Figures consist of True Vocal Folds: Onset/Offset, False Vocal Folds, True Vocal Folds: Body-Cover, Thyroid Cartilage, Cricoid Cartilage, Aryepiglottic Sphincter, Larynx, Tongue, Velum, Jaw, Lips, Head & Neck, and Torso. While not all elements need further explanation, such as the Larynx, Tongue, Velum (soft palate), Jaw, Lips, Head & Neck, and Torso, it is important to discuss the other Figures that are not commonly taught in applied lessons. Not all of the 13 Compulsory Figures are applicable to saxophone performance; the relevant figures and their applications to saxophone performance will be provided in the "EVM Interpretation of Sinta's Voicing Exercises" section of this chapter.

True Vocal Folds: Onset/Offset refers to how the vocal folds start and stop vibrating, and it is one of the main mechanisms for shaping tone, clarity, and expressiveness in speech and

[70]Spadaro and McInnis, *The Estill Adventure*, 3.; Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 56.

singing. The onset is the pattern of coordination between breath and fold closure at the beginning of a sound, whereas the offset is the pattern of coordination between breath and fold opening or relaxing at the end of a sound. The three primary onset/offset options are: *glottal*, where the folds are close first, then air is released, giving a sharp, clean, or sometimes "attacky" edge; *aspirate*, where the air flows first, then folds meet, giving a breathy sound or a sigh-like release; and *smooth*, where the breath and fold movement begin and end together, giving a clean, almost invisible start and finish.[71] Changing onset/offset alters voice quality and how the breath behaves. Conscious control of onset/offset allows you to reduce hard, repetitive "attacks" that can irritate the folds, dial in specific types of voice quality with more precision and make mid-phrase choices support legato or percussive articulation as desired. Essentially, True Vocal Folds: Onset/Offset is the "power switch" technique for the folds, how to flip vibration on and off, and mastering those switches gives fine control over color and effort.[72]

The False Vocal Folds are a "second pair" of folds that sit right above the True Vocal Folds and can squeeze toward the midline (constrict), sit in a neutral position (mid), or pull widely apart (retract). The larynx has three closure levels: top Aryepiglottic Sphincter, middle False Vocal Folds, and bottom True Vocal Folds. The False Vocal Folds themselves are mostly fatty pads, not primary vibrators; when they move, they mainly change how air can flow toward the True Vocal Folds. During constriction, the False Vocal Folds move inward, disrupting airflow to the True Vocal Folds and adding noise which may be described as pressed, scratchy, grunt-like, or distorted. In mid, the False Vocal Folds are halfway between in and out, giving a more "natural," speech-like quality with moderate resistance and slight noise. When False Vocal Folds are pulled wide (retract), it minimizes constriction so the tone becomes open, resonant, free, and

---

[71] Spadaro and McInnis, *The Estill Adventure*, 6. Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 51.
[72] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 56.

clear, and airflow is more efficient. In practice, learning to sense and adjust False Vocal Folds retraction versus constriction gives a powerful method to reduce strain, avoid unwanted distortion in throat, and intentionally add or remove “grit” and intensity as a stylistic choice rather than an unconscious habit.[73]

True Vocal Folds: Body-Cover refers to how much of the vocal fold’s depth is active and the flexibility of the surface layer. It is a primary way to choose register and tone color at any pitch. The True Vocal Folds: Body-Cover have a deeper “body” tissue (muscle and ligament) and a more flexible “cover” at the surface. By changing the body length, thickness, and tension, it changes how long the folds stay closed each cycle and how freely the cover vibrates.[74] True Vocal Folds: Body-Cover is not applicable in wind playing because the reed or air column, not your vibrating vocal folds, is the primary sound source. Hence, those fine-grained choices about how the folds themselves vibrate do not map directly onto the saxophone in the same way they do for sung phonation.

The Thyroid Cartilage is the large shield of the voice box that the front ends of both True and False Vocal Folds attach. Tilting it changes how those folds vibrate and how the laryngeal tune is shaped, directly affecting pitch and voice quality. It is the largest cartilage of the larynx, forming a protective front wall around the airway and serving as the anchor point for the front ends of the vocal folds. The main contrast between vertical (resting point) and tilt (forward rotation) is that vertical tends to feel and sound more “earthy, blunt, chesty,” while tilt is heard as clearer and sweeter. Tilting changes true-fold length, mass, and tension and slightly reshapes the supraglottic space (False Vocal Folds and aryepiglottic area), so it participates in both pitch rising

---

[73] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 72.
[74] *Ibid*., 90.

and in refining timbre, allowing a singer to lighten and polish quality at the same pitch by increasing tilt rather than only by changing breath or volume.[75]

The Cricoid Cartilage is the ring-shaped "base" of the larynx that everything else sits and pivots on, and its tilt controls how tight and long the vocal folds are, making it a mechanism for loudness and belt-type qualities. It is the lowest laryngeal cartilage, a complete ring right above the trachea and shaped like a signet ring with a narrow arch in front and a tall lamina (the broad, high back wall of the ring) in back. Thyroid and arytenoid cartilages attach to it via the cricothyroid and cricoarytenoid joints, so any shift in cricoid position changes the length, mass, and tension on the vocal folds between thyroid (front) and arytenoids (back). When the larynx tilts, it stretches the vocal folds and lessens subglottal pressure, changing fold length, tension, and vibratory behavior. Estill contrasts vertical with tilt, where tilt is associated perceptually with louder, thicker, "belted" qualities. Tilting the cricoid helps decrease subglottal pressure and the closed-phase proportion of each vibratory cycle, boosting intensity without simply blasting more air, so it becomes one important ingredient in producing powerful belt while managing strain and avoiding over-pushing the breath.[76]

The Aryepiglottic Sphincter is the small, adjustable tube at the very top of the larynx, formed by the epiglottis, aryepiglottic folds, and arytenoids. Narrowing or widening it controls how "twangy" or "open" the voice sounds without changing pitch. It is the topmost of the larynx's three closure levels, tightening "like a purse string" between the epiglottis and arytenoids to form a short tube (the epilarynx) above the true and False Vocal Folds. Because the epiglottis is flexible, this tube can change shape, cackling like a witch narrows it, while a big "lion's yawn" posture widens it. Estill distinguishes wide Aryepiglottic Sphincter (relaxed,

---

[75] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 100.
[76] *Ibid*., 104-113.

opened) versus narrow Aryepiglottic Sphincter (tightened). Wide Aryepiglottic Sphincter gives a darker, flatter, more open and relaxed sound; narrow Aryepiglottic Sphincter concentrates energy into bright "ring" or twang (singer's formant, Broadway ping) and is often felt as buzzing "in the mask" even though it is not the same thing as nasal resonance.[77]

Relevant Compulsory Figures in Saxophone Voicing

Jonathan Kierspe's treatise of *A Theoretical Application of the Estill Voice Model for the Saxophone* identified 11 Compulsory Figures that are actively applicable to saxophone playing in some capacity, with the Velum essentially non-applicable (must always stay high) and True Vocal Folds: Body-Cover applicable only in limited extended technique contexts and requiring careful management to prevent vocal fold damage.[78]

Figures with limited or no direct saxophone application include the True Vocal Folds: Body-Cover mechanism. Since the reed, rather than the vocal folds, is the primary vibratory source in saxophone playing, this concept is generally not applicable. It becomes relevant only in extended techniques such as singing through the instrument or growling, and even then, must be approached cautiously to avoid strain on the vocal folds. The Velum is the least applicable structure in standard saxophone performance. It should remain in a consistently elevated position to prevent air from escaping through the nasal passage, ensuring that all available air pressure is directed into the mouthpiece and reed.[79] This focused airstream supports stable reed vibration, which is essential for a reliable response and consistent tone, especially in soft dynamics and extreme registers. When the velum lowers and air leaks through the nose, the effective air

[77] Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 116-121.
[78] Kierspe, *A Theoretical Application of the Estill Voice Model for the Saxophone*, 58.
[79] Lucía Mata-Pose et al., "Velopharyngeal Incompetence in Musicians: A State-of-the-Art Review," *Journal of Personalized Medicine* 13, no. 10 (2023): 1477.

pressure at the reed drops, which can cause notes to thin out, crack, or fail to speak cleanly. A high velum also minimizes unwanted nasal resonance, helping the player produce a more centered, resonant, and controlled sound.[80] However, as noted by Kierspe, EVM exercises involving the Velum can still be valuable for developing awareness and addressing issues such as a leaking velum.

Figures that relate directly to saxophone playing highlight several key anatomical structures and their functions. The True Vocal Folds: Onset/Offset patterns translate directly to articulation strategies: a glottal onset corresponds to double-tonguing, an aspirate onset aligns with air attacks, and a smooth onset reflects normal articulation. The False Vocal Folds tend to adduct as register rises, and this adduction facilitates altissimo playing and growling. They are also partially adducted during glottal articulations.[81]

The Thyroid Cartilage tilts in response to register changes, influencing both pitch and tone quality, while the Cricoid Cartilage works in coordination with the thyroid during these register adjustments.[82]

The Aryepiglottic Sphincter widens as range increases, which is the opposite of its behavior in singing. It also widens during pitch bends and overtone production, representing a key distinction between saxophone and vocal function. The Larynx, one of the most researched structures in this context, shows a strong correlation between height and register: it tends to be higher in the low range of soprano and alto saxophones and lowers in the palm keys, directly relating to pitch.[83]

---

[80] Lucía Mata-Pose et al., "Velopharyngeal Incompetence in Musicians: A State-of-the-Art Review," *Journal of Personalized Medicine* 13, no. 10 (2023): 1477.
[81] Kierspe, *A Theoretical Application of the Estill Voice Model for the Saxophone*, 59-63.
[82] *Ibid*., 64-9.
[83] *Ibid*., 71-4.

The Tongue is one of the most discussed Figure in saxophone pedagogy. Its position ranges from low to high, with compression models being less applicable. Instead, vowel shapes such as “e”, “u,” and “i" effectively guide register-specific positioning. The Jaw plays an active role in vibrato production, pitch adjustment, and reed pressure, operating across four positions with a neutral mid-position as the default.[84]

The Lips contribute to embouchure seal, tone color, and intonation. The Head & Neck should be kept steady to maintain upper vocal tract stability, but be aware that wearing a neck strap can change how the sternocleidomastoid (the large front neck muscle that helps turn and support the head) functions. The Torso is central to air support and breath control; an anchored torso promotes greater projection and stability, though the appropriate level of effort varies across pedagogical approaches.[85]

Overall, Kierspe’s application of the EVM clarifies that while not all Figures transfer directly to saxophone performance, a majority offer meaningful insight into the coordination of the vocal tract when playing saxophone. Structures such as the Tongue, Larynx, Jaw, and False Vocal Folds play especially significant roles in shaping register, articulation, and tone, whereas others like the Velum and True Vocal Folds are either consistently fixed or situationally relevant. Taken together, these findings reinforce the value of voice-based anatomical awareness in saxophone pedagogy, providing a more systematic framework for understanding how internal adjustments contribute to efficient, flexible, and expressive performance.

---

[84] Kierspe, *A Theoretical Application of the Estill Voice Model for the Saxophone*,75-9.
[85] *Ibid*.,79-83.

EVM Interpretation of Sinta's Voicing Exercises

This reframing provides a pedagogical vocabulary for interpreting Donald Sinta's *Voicing* exercises by evaluating the eleven active figures when voicing on the saxophone. Voicing moves from first-mode control to second, third, and fourth-mode development, and then towards altissimo technique. Success is framed in terms of voicing awareness, accuracy, consistency, aural intent, and deliberate oral cavity manipulation. This method will focus on modes which can be reinterpreted through EVM. Chapters in Sinta's book which focus on fingerings are outside the scope of this document.

Saxophone voicing does not require pitched vocal-fold vibration in the same way singing does. The reed vibrating on the saxophone mouthpiece is analogous to the vocal cords when singing. Rather than claiming direct equivalence between sung phonation and saxophone tone production, this reframing emphasizes the similar mechanics of vocal phonation and saxophone voicing. For that reason, EVM is useful not as a literal explanation of every aspect of saxophone playing, but as a practical framework for understanding and controlling the oral cavity and vocal tract.

I chose Sinta's progression for enhancement with EVM principles because it already frames voicing as a step-by-step journey rather than a set of isolated skills. The learning curve of exercises is thoughtfully organized and encouraging, inviting students to accept a slower, more organic development of control over the oral cavity. This progression is presented in a conversational manner which makes the material accessible to beginners.

By contrast, I decided not to adapt Teal's approach because his resource predates the idea of voicing as a pedagogical topic in saxophone. Likewise, there is very little material specific to oral cavity manipulation in regards to altissimo. While Teal does briefly liken saxophone playing to singing, and he presents an overtone chart with a quick description, it only scratches the surface on the topic of voicing.[86] Additionally, I did not select Raschèr's approach for enhancement with EVM because the sections that do correlate singing to saxophone voicing do not go into deep detail, even though it offers some valuable material on overtones and voicing.[87] Much of Raschèr's book consists of notated overtone exercises with relatively sparse accompanying text. In contrast to Sinta, Raschèr's limited prose on the mechanics of the oral cavity restricts the ease of which EVM can enhance his approach.

Sinta's, Teal's and Raschèr's resources recognize that the internal configuration of the oral cavity significantly affects voicing. Sinta's approach, however, stands out as the most primed to being enhanced with EVM because of its beginner-friendly language, gradual pacing, aid in troubleshooting, and encouragement. These attributes are shared with principles of EVM, which when combined create a more detailed and helpful approach on a unified pedagogical method on saxophone voicing. Adapting Sinta's exercises and approach with EVM equips both teachers and students with a richer, more precise vocabulary for describing and guiding voicing.

Sinta's Modes

Sinta differentiates his exercises according to modes. Sinta defines a *mode* as an overtone above the fundamental. Modes in this context are identical to overtones. Only the first four modes are explicitly discussed in *Voicing*, but the modes continue on past that. Likewise, this document will only focus on the first four modes.

---

[86] Teal, *The Art of Saxophone Playing,* 98.
[87] Raschèr, *Top Tones for the Saxophone, 12.*

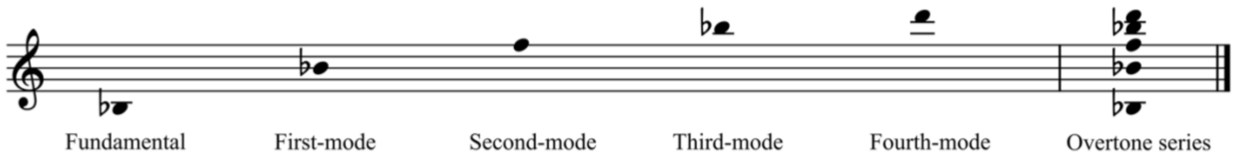


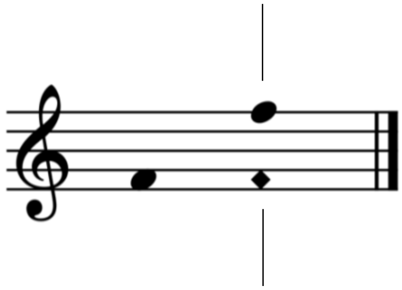
Figure 1: *Sinta*'*s Modes*

The First-Mode*: *Introduction to Saxophone Voicing

Sinta introduces first-mode exercises through overtone production using low note fingerings and using the syllables "kuh" and "kah" to gain awareness of the back of the throat. He recommends avoiding overreliance on the octave key and instead to use the tongue to isolate the next partial. The "kuh" and "kah" syllables designate a throat articulation strategy that prepares the oral cavity for reliable response and voicing in the upper register.[88] In EVM terms, this same coordination is realized through specific configurations of the Tongue, Larynx, and Jaw. This is also supported by Head & Neck and Torso anchoring, shaping the vocal tract. The False Vocal Folds are involved indirectly in that an optimal "kah/kuh" attack presupposes False Vocal Folds retraction rather than active constriction. The remaining EVM Figures (Velum, Lips, Aryepiglottic Sphincter, Thyroid Tilt, Cricoid Tilt, and Body-Cover) can vary and color the resulting sound, but they do not define the essential identity of the "kah/kuh" articulation. Pedagogically, this means that the student should not 'force' the note at the reed but accurately resonate at the desired upper partial. Through systematic overtone and voicing exercises, the saxophonist cultivates a consistent internal configuration, essentially a memorized tongue-throat posture, allowing the upper partial to emerge cleanly while the embouchure, air stream, and pitch center remain stable.

[88] Sinta, *Voicing*, 11.

Filled Notehead: Sounding pitch



Sinta: Kah/Kuh, back of throat, thrust tongue to new position
EVM:
 Primary: Tongue, Larynx, Jaw, Head & Neck, Torso
 Secondary: False Vocal Folds, True Vocal Folds, Lips
 Low Applicability: Thyroid and Cricoid Cartilage, Aryepiglottic Sphincter

Diamond Notehead: Fingered pitch

Figure 2: *Sinta's First-Mode*

Figure 2 show Sinta's vocabulary as he discusses how to produce the first-mode. The accompanying EVM Figures shows exactly what parts of the anatomy are active when producing the sounding pitch in hierarchical order.

Primary Figures in the First-Mode

Primary EVM Figures:
 Tongue: Sides of tongue dorsum beside upper molars
 Larynx: Wide open throat and effort pulling down
 Jaw: Mouth vertically wide open
 Head & *Neck*: Anchor; effort in the neck and roof of mouth
 Torso: Anchor; effort stretching the chest and back muscle

Figure 3: *EVM Principles in Sinta's First-Mode*

Figure 3 outlines the fundamental pitches of the bottom octave of the saxophone's range in diamond noteheads and shows the first-mode octave in solid noteheads. The primary EVM figures listed are involved in producing the first-mode.

The Tongue is the dominant figure in the first-mode. Sinta's instruction to use the tongue as the octave key and to thrust the tongue into a new position for the next partial refers to how the act of raising the back of the tongue (the dorsum) controls the shift from the fundamental to the first partial. EVM distinguishes three tongue positions: low, mid, and high. The tongue position affects overall sound quality. EVM's tongue positions provide a useful descriptive language for this process because first-mode voicing can be characterized as a tendency toward a higher, more focused tongue position, which facilitates access to upper partials while minimizing the need for embouchure adjustments. The high tongue position can be found by saying the vowel sound /i/ as in "beet."[89]

The Larynx also plays an important role in this process, functioning more as a stabilizing structure. Sinta repeatedly refers to the "feeling of an enlarged or open throat" and warns against tightening embouchure muscles when matching pitch and timbre. When reframing this through EVM, the Larynx prefers a stable or lowered laryngeal posture during early overtone learning, not because a low larynx is always required, but because a high larynx position can shrink the vocal tract and make it more difficult to isolate higher partials.

The Jaw becomes important when the first-mode extends downward to low B-flat. Sinta recommends dropping the jaw and slightly loosening the embouchure.[90] Dropping the jaw lengthens and opens the oral cavity, giving more space to isolate the fundamental.

[89] Spadaro and McInnis, *The Estill Adventure*, 20.
[90] Sinta, *Voicing*, 15.

The Head & Neck and Torso are supporting Figures rather than voicing figures, but they are still relevant because Sinta repeatedly asks the player to sustain notes, maintain control, and avoid biting. EVM distinguished between *anchored* and *relaxed* positions for both the Head & Neck and Torso. When these Figures are anchored the sound is clear and steady during sound production. Conversely, when these Figures are relaxed the sound is unsupported.[91] Reframed for saxophone, these Figures justify an emphasis on avoiding biting and reducing oral cavity constriction and body tension.

Secondary or Conditional Figures in the First-Mode

The False Vocal Folds are indirectly applicable to the first-mode. Although Sinta never names them, the first-mode's emphasis on an open and consistent timbre strongly implies the avoidance of supraglottic constriction. EVM's distinction between False Vocal Folds constriction and retraction offers a clear caution: if a player understands "attack" as tightening or closing the throat, the onset will probably sound forced or unstable instead of producing a clean change in voicing.

True Vocal Folds: Onset/Offset has limited but meaningful relevance because Sinta repeatedly recommends a glottal sound onset. On saxophone, this should not be read as literal phonation onset in the singing sense, but instead reframing it through EVM to help interpret the glottal sound onset as favoring a clean, coordinated start over either an aspirate release or a hard glottal catch.

[91] Spadaro and McInnis, *The Estill Adventure*, 28-31.; Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 191-209.

The Lips are not central in the first-mode. Sinta's instructions focus far more on internal oral cavity configuration than on external lip shaping. Since there are already established resources on forming a correct embouchure, this will not be a focus of this document. The purpose of this document is to explore the underdeveloped aspects of voicing.

Low-Applicability Figures in the First-Mode

Thyroid Cartilage is not centrally applicable in the first-mode because Sinta does not ask the player to generate specific "cry-like thinning" or sung phonatory quality changes associated in EVM with thyroid tilt. Any thyroid movement does not affect any production of the first-mode. Cricoid Cartilage is likewise not centrally applicable. EVM links cricoid tilt to loud, excited, thicker-fold phonatory events, whereas first-mode is concerned with overtone production, overtone matching, and vocal tract awareness.

The Aryepiglottic Sphincter plays only a small role in first-mode playing. In EVM it is linked with "ring" or "twang," so a bit of narrowing might help projection or focus, but the first-mode stresses timbre matching, openness, and avoiding unnecessary effort. For that reason, Aryepiglottic Sphincter activity should be treated as an occasional side effect at most, not a main teaching tool for first-mode pedagogy.

The Second-Mode*: *Isolating the Next Partial

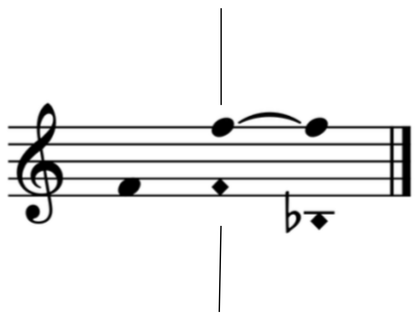


Figure 4: *Sinta*'*s Second-Mode*

Figure 4 builds upon *Figure 2* by transitioning to the second-mode by fingering a low B-flat and producing middle F. The primary EVM Figures remain relatively the same with small positioning adjustments.

Once comfortable in the first-mode, Sinta guides the player to move on to the second-mode by maintaining an upper pitch while changing to a new fundamental fingering, extending range chromatically, and venting the octave key. Here, the EVM-inspired reframing becomes more complex, because the student's task is no longer simply discovering a new tract shape but preserving a single resonance target while the fingering and acoustic impedance conditions change. The second-mode includes refinement, inhibition of old habits, and more differentiated internal coordination. Although Sinta encourages "venting," which is pressing a key to alter the airstream to help isolate certain partials, as a way to deal with the acoustical properties of the instrument, this pedagogical approach encourages internalizing the relevant Figures in order to isolate the desired partial rather than relying on the instrument.

Primary Figures in the Second-Mode

Primary EVM Figures:
Tongue: Sides of tongue dorsum beside upper molars or sides of tongue dorsum between upper and lower molars
Larynx: Somewhat flexible; monitored for unnecessary motion
Jaw: Mouth vertically wide open; relaxed

Figure 5: *EVM Primary Principles in Sinta's Second-Mode*

Figure 5 outlines the fundamental pitches of the bottom octave of the saxophone's range in diamond noteheads and shows the second-mode in solid noteheads. The primary EVM figures listed are involved in producing the second-mode.

The Tongue remains the central figure. Sinta explicitly states that the oral cavity must learn a "new physiological setting" for second-mode pitches and repeatedly asks the student to hold voiced notes long enough to develop awareness of different settings for different registers.[92] In EVM terms, this means that second-mode is best understood as a more exacting series of tongue-dorsum calibrations than first-mode, with finer distinctions between adjacent partials.

The Larynx is more important in the second-mode than in the first-mode because second-mode extension places greater pressure on tract consistency across register shifts. EVM's larynx positions offer an explanation why students occasionally overshoot or undershoot target overtones: unstable laryngeal height can alter tract length and timbral brightness enough to derail the intended partials. The pedagogical implication is not that the larynx must be fixed rigidly, but that it must be monitored for unnecessary motion.

[92] Sinta, *Voicing*, 20.

The Jaw also remains directly applicable. As second-mode pitches rise, many players compensate by biting or narrowing the oral cavity space. EVM's framework clarifies that a relatively free, adjustable jaw helps maintain tract flexibility and avoid tongue immobilization. Since Sinta repeatedly stresses consistency and gradual addition of notes, an EVM-inspired reframing argues that jaw freedom is part of the mechanism that allows not only comfort, but pitch accuracy.

Secondary Figures in the Second-Mode

Secondary EVM Figures:
False Vocal Folds: Relaxed in the center of the throat or effort in the neck and roof of mouth to open the throat space
Head & *Neck*: Anchor; effort in the neck and roof of mouth
Torso: Anchor; effort stretching the chest and back muscle
True Vocal Folds: Effort in the neck, roof of mouth, and torso

Figure 6: *EVM Secondary Principles in Sinta's Second-Mode*

The False Vocal Folds are again indirectly applicable, especially because the second-mode introduces more difficult tasks and therefore greater risk of physical discomfort. EVM's retract versus constrict contrast provides a pedagogical perspective: when second-mode production becomes difficult, the teacher can distinguish proper oral cavity shape from tension. This is particularly useful when students mistake "support" for biting.

The Head & Neck and Torso become more prominent because the second-mode culminates in exercises requiring repeated accuracy, consistency, and flexibility. EVM's anchoring figures explain why body awareness matters here: efficient postural support allows the oral and laryngeal figures to change without glottal collapse or neck tension. The roles of the Head & Neck and Torso are still secondary, but easier to justify in the second-mode than in the first-mode.

The True Vocal Folds: Onset/Offset remains only somewhat relevant. The second-mode instructs the student to practice both tongued and slurred and also to use the throat "kuh" attack. EVM onset language helps distinguish a coordinated start of notes from air leakage or percussive throat closure. However, because saxophone tone is reed-driven, this Figure should still be described as an analogy for coordination rather than literal explanation of sung onset types.

Low-Applicability Figures in the Second-Mode

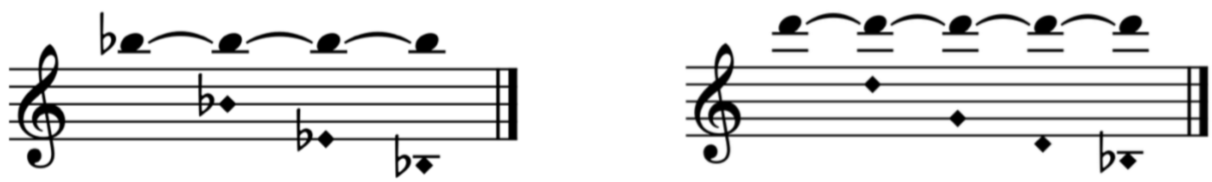
The second-mode gives no evidence that the Lip protrusion, spreading, or other lip-shape manipulations are a primary route to second-mode mastery. Some students may make minor lip adjustments, but these should be secondary to tongue and jaw organization.

The Aryepiglottic Sphincter remains at low applicability. Narrowing the Aryepiglottic Sphincter could potentially influence focus and accuracy, yet the second-mode's pedagogical purpose is to maintain upper notes, learning new oral cavity shapes, and using venting as needed.

The Thyroid Cartilage and Cricoid Cartilage remain largely non-applicable for the same reason as in the first-mode: the second-mode does not center phonatory quality changes associated with tilt figures, but rather with resonance tuning and fingering-to-tract coordination. Despite this, mentioning them is still useful because it clarifies that not every EVM figure can be used in an instrumental settings in a directly relevant way.

The Third and Fourth-Modes**:** ***Discovery, Exaggeration, and Experimental Exploration***

Sinta: Reinforcement of first and second-mode, increased awareness of oral cavity shape, introduction to third and fourth-mode
EVM:
- **Primary**: Tongue, Larynx, Jaw
- **Secondary**: False Vocal Folds, Head & Neck, Torso, Aryepiglottic Sphincter
- **Low Applicability**: Lip, Thyroid and Cricoid Cartilage, True Vocal Folds



Third-Mode **Fourth-Mode**

Figure 7: *Sinta's Third and Fourth-Modes*

Figure 7 approaches the modes in reverse. By sustaining the top B-flat and moving through the diamond notehead fingerings, the saxophonist will progress through the first three modes.

The third and fourth-modes are the most explicitly exploratory mode in this group. Sinta describes its goals as increased awareness, discovery of new oral cavity positions, and reinforcement of prior skills, then presents "experimental exercises" designed to force exaggerated oral cavity changes and break old physiological habits. These modes align especially well with EVM because both pedagogies value deliberate contrast, exploratory exaggeration, and the development of a larger coordinative vocabulary.

Primary Figures in the Third and Fourth-Modes

Primary EVM Figures:
Tongue: Sides of tongue dorsum beside upper molars or sides of tongue dorsum between upper and lower molars
Larynx: Balanced effort in the center of the throat or small throat space and effort lifting up
Jaw: Mouth vertically wide open or relaxed

Third-Mode **Fourth-Mode**

Figure 8: *EVM Primary Principles in Sinta's Third and Fourth-Modes*

Figure 8 isolates the transitions into the higher modes by relating it to the more comfortable, lower modes. The left "third-mode" example sustains each pitch while first using a second-mode fingering and then switching to a third-mode fingering, and the right "fourth-mode" example does the same but by switching from a third-mode fingering to a fourth-mode fingering.

The Tongue remains primary because the third and fourth-modes continue to build mode changes through oral-cavity manipulation and production of higher overtones. However, the EVM-inspired insight here is broader than "high tongue for high notes." The third and fourth-mode's emphasis on discovering unexpected pitches and analyzing new feelings suggests a pedagogy of figure exploration: students are not merely repeating fixed settings but expanding their accessible range of tongue configurations and learning to identify them kinesthetically.

The Jaw is highly relevant because exaggerated experimentation often succeeds only when the jaw is not locked into the "habit" position. Sinta frames these exercises as ways to add new muscle skills and free the body from prior reinforcement.[93] EVM specifies that jaw mobility can either permit or inhibit those new skills, making jaw monitoring a concrete pedagogical task rather than broad advice to "stay relaxed."

The Larynx also becomes more relevant in the third and fourth-modes because it invites the player to maintain low pitches while adding the octave key, and then to reverse or alter the expected pitch through oral tract changes.

Secondary Figures in the Third and Fourth-Modes

Secondary EVM Figures:
False Vocal Folds: Relaxed in the center of the throat or effort in the neck and roof of mouth to open the throat space
Aryepiglottic Sphincter: Wide in the center of throat or high, small space in throat/buzz in cheeks
Head & Neck: Anchor; effort in the neck and roof of mouth
Torso: Anchor; effort stretching the chest and back muscle
True Vocal Folds: Effort in the neck, roof of mouth, and torso

Third-Mode **Fourth-Mode**

Figure 9: *EVM Secondary Principles in Sinta's Third and Fourth-Modes*

The False Vocal Folds are more relevant here than in prior modes because experimental practice often leads to physical tension. Sinta invites students to explore wildly and even "vent frustrations," yet also asks them to observe what is produced and convert discovery into usable

[93] Sinta, *Voicing*, 21.

awareness.[94] EVM's False Vocal Folds framework is ideal for this distinction: experimentation is productive only when the throat remains sufficiently retracted to permit clear airflow rather than constricting due to tension.

The Aryepiglottic Sphincter has its strongest possible relevance in the third and fourth-modes, though still as a secondary figure. Because these modes involve experimentation with upper harmonics and timbre, Aryepiglottic Sphincter narrowing can be understood as one factor that helps produce very concentrated, bright, ringing tones and, at the extremes, even high, squeak-like sounds.

Head & Neck and Torso remain supportive but significant. Experimental tasks often fail when the player excessively moves their body compared to making subtle internal changes; EVM anchoring provides a rationale for stable support without rigidity. Since the third and fourth-modes are about learning unfamiliar oral cavity shapes, supportive anchoring can be framed as method to allow experimentation to remain organized.

Low-Applicability Figures in the Third and Fourth-Modes

True Vocal Folds: Onset/Offset continues to be only partly applicable. The third and fourth-modes are less about attack type than sustained manipulation and transitions. EVM onset language may still help when a student initiates an exercise with excess breathiness or throat tension, but it is not the third and fourth-mode's core explanatory figure.

[94] Sinta, *Voicing*, 20.

Lips remain of low relevance because Sinta's experimental focus is internal. Lip changes may color resonance slightly, yet they are not foregrounded as a pedagogical means of producing third and fourth-modes.

Thyroid and Cricoid Cartilages are largely non-relevant here. They may be engaged indirectly when students shout internally or change tract thickness through associated laryngeal adjustments, but third and fourth-modes voicing still does not intentionally address the phonatory tilt contrasts that those Estill figures target. Their best use in analysis is to delimit the boundary between useful analogy and overextension.

Combining Modes: *Aural Imagery and Tunes*

Sinta combines the different modes by reframing voicing through familiar songs, bugle calls, and scales, while explicitly insisting that the player hear each pitch before it is played and send the appropriate physiological message to mind and vocal tract. Sinta implies that voicing success depends not only on tract configuration but on pre-phonatory inner hearing that guides that configuration.[95]

---

[95] Sinta, *Voicing*, 33.

Figure 10a: *"Mary Had a Little Lamb" combining third and fourth-modes*

Figure 10b: *"Mary Had a Little Lamb" combining second, third, and fourth-modes*

Figures 10a and *10b* shows a progression of playing "Mary Had a Little Lamb" which gradually incorporates more voicing manipulation by mixing the modes. By using a familiar tune, the student will likely imagine the pitch accurately, and the challenge is correlating an oral cavity shape dependent on the pitch and mode.

Figure 10c: *"Mary Had a Little Lamb" showing all possible combinations of first through fourth modes*

Figure 10c shows all possible fingering options for producing each pitch of the "Mary Had a Little Lamb" melody within the first four modes. Students are encouraged to change which mode they are utilizing as the melody progresses, particularly when the melody repeats pitches.

Primary Figures in Combining Modes

The Tongue remains relevant as voiced melodies and scales require rapid and differentiated oral cavity changes across the range. EVM's tongue framework makes it possible to describe these changes as a sequence of maneuvers. In pedagogical terms, combining the modes encourages the student to associate heard intervals with reproducible tongue-dorsum gestures.

The Larynx is also relevant because melodic voicing on a single fundamental or across fundamentals requires subtle timbral and oral tract changes. A stable but responsive laryngeal position helps preserve the internal link between imagined pitch and produced pitch. EVM's low-mid-high larynx options support this: the student can notice whether unwanted brightness or darkness reflects a laryngeal shift.

The Jaw is relevant because scales and melodies require quick changes without losing clarity or flexibility. EVM's jaw options clarify that a "frozen jaw" is not equivalent to control; in fact, a small amount of jaw flexibility may be necessary so the tongue can still move inside the oral cavity. Since combining the modes explicitly urges the player to challenge previous limitations creatively, the jaw can be framed as a slightly flexible part of the oral cavity, not a rigid structure.

Secondary Figures in Combining Modes

True Vocal Folds: Onset/Offset is more relevant here because combining the modes includes scales and melodies that may be practiced with slight articulation or "throat kah/kuh." EVM onset categories can help classify whether a student's attacks are clean, breathy, or tense. In combining the modes, onset language is useful because it intersects with rhythmic and melodic accuracy.

False Vocal Folds remains indirectly relevant as a way to diagnose tension. Since combining the modes links success to imagining the pitch before playing it, excessive tension can be interpreted as evidence that the student is not internally audiating the pitch. EVM supports an important pedagogical focus: accurate inner hearing must be paired with a flexible oral cavity shape in order to accurately produce the imagined pitch.

The Head & Neck and Torso is also relevant because phrase-like voicing patterns are easily destabilized by poor posture or forcing. EVM anchoring helps explain how posture stability supports consistency across the range. This aligns with Sinta's daily goal of one new note or increased awareness, since consistent support makes incremental gains more reproducible.[96]

Low-Applicability Figures in Combining Modes

The Lips continue to have limited relevance. Familiar tunes may tempt students to distort the embouchure as a substitute for oral cavity shaping, but Sinta's intention clearly points the student back toward ear, mind, and vocal tract.

Aryepiglottic Sphincter can be mentioned as a possible contributor to tonal brightness or focus, but this does not provide enough evidence to make it central.

Thyroid Cartilage and Cricoid Cartilage remain outside the main explanatory figures as combining the modes is about audiation-guided tract shaping, not voice-quality figures based on deliberate tilt coordination. Their inclusion in this analysis is of low importance still valuable as it prevents the reader from confusing "inner hearing" with sung phonation strategies.

Altissimo Technique as Integration of Voicing Mechanism

Instruction on developing the altissimo register has shifted from an integrated, concept-based approach to a more technique-focused method, emphasizing front E and F fingerings, palm-key coordination, mirror practice, efficient hand motion, intonation monitoring, and repetitive pattern work. In EVM terms, altissimo still depends on voicing figures, but the

[96] Sinta, *Voicing*, 42.

pedagogical problem is now integrative rather than exploratory: the player must continue to manipulate the oral cavity under increased challenge.

Primary Figures in Altissimo

The Tongue remains primary because front fingering success in altissimo depends on maintaining appropriate voicing. Even when the text discusses hand position, altissimo presupposes that the student can produce F and add front E with voicing and fingering consistency.

Larynx is highly relevant because altissimo often fails due to the oral cavity shifting. EVM allows altissimo to be described as a failure of change in laryngeal and supraglottic position rather than as a mysterious technical inconsistency. The pedagogical value is substantial: students can be taught to monitor whether palm-to-front fingering transitions alter the oral cavity.

Head & Neck and Torso become particularly important in altissimo because the text explicitly addresses efficient physical movement, mirror practice, and avoidance of unnecessary wrist arching. They resonate with EVM anchoring principles: stable upper-body organization supports efficient local motion.

Secondary Figures in Altissimo

The Jaw remains relevant because fast oral cavity changes or finger technique often result in tension. A flexible but controlled jaw helps preserve the tract shape needed for front fingerings while the hands change rapidly. Though Sinta does not emphasize the jaw explicitly in altissimo, it is still useful to recognize when finger technique disrupts the oral cavity shape.

False Vocal Folds are also useful in diagnosing inconsistencies. Altissimo repeatedly insists on intonation accuracy and careful listening. From an EVM perspective, poor intonation or spread tone during altissimo suggest oral cavity tension, so false-vocal-fold retraction remains useful in maintaining flexibility. This is because it keeps the space above the true vocal folds open, which lets the true folds vibrate more freely and reduces unwanted constriction. In practical terms, that often means less throat tension, easier onset, clearer tone, and more flexibility across styles and registers.[97]

Aryepiglottic Sphincter has more relevance here because altissimo production is often tonally bright and uncontrolled. EVM's narrow Aryepiglottic Sphincter option can offer a possible explanation to focused altissimo tone. Still, because the altissimo section discusses fingerings, hand movement, and intonation more explicitly than throat narrowing, Aryepiglottic Sphincter should remain a secondary tool.

Low-Applicability Figures in Altissimo

True Vocal Folds: Onset/Offset is of only minor relevance in altissimo. Although tongued and slurred practice may involve different methods of starting the sound, the altissimo's main issue is continuity through technical patterns rather than relying on starting notes in the same way. As in earlier modes, this Figure is best used only when tone quality at the start of the note clearly interferes with stability.

---

[97] Spadaro and McInnis, *The Estill Adventure*, 8.; Steinhauer, Klimek, and Estill, *The Estill Voice Model*, 62-7.

The Lips remain non-central because the stated technical priorities lie elsewhere. Overemphasis on lip reshaping could even work against the efficient, repeatable movement pattern outlined in altissimo.

Thyroid Cartilage and Cricoid Cartilage are again of low relevance. Altissimo accuracy is not described through vocal-fold quality contrasts, loud calling gestures, or cry-like tilt behaviors, but through combined fingering and voicing practice.

Pedagogical Implications

Sinta's language is pedagogically effective, but often broad; EVM adds a refined vocabulary that can help teachers diagnose whether a student is struggling because of tongue position, jaw position, laryngeal instability, False Vocal Folds constriction, or poor posture. This gives the teacher a more detailed method of communicating while still staying aligned with Sinta's progression of exercises.

This reframing also opens a strong methodological claim: Sinta's modes can be taught as a sequence. The first-mode establishes foundational tongue-led oral tract shaping, the second-mode differentiates mode-specific positions more precisely, the third and fourth-modes expand the player's vocabulary, the combining modes link voicing to audiation and producing melodies, and altissimo integrates the oral cavity with finger technique.

CHAPTER 4. CONCLUSION

Saxophone voicing remains one of the most difficult skills for developing players to understand because the relevant adjustments occur inside the body and are therefore difficult to see, describe, and consistently reproduce. Although standard pedagogical texts such as Donald Sinta's *Voicing*, Larry Teal's *The Art of Saxophone Playing*, and Sigurd Rascher's *Top-Tones for the Saxophone* remain influential, they often rely on effective but comparatively broad analogies rather than a fully anatomical or acoustical explanation of how voicing works. The central argument of this document has been that saxophone voicing pedagogy can be strengthened by treating the player's vocal tract and the instrument bore as an interacting system and by adopting a more explicit vocabulary for the internal co-ordinations required to control that interaction.

The need for this reframing by identifying a persistent gap in saxophone pedagogy: students are frequently asked to sing, hum, whistle, or imagine vowel shapes, but they are given limited guidance for understanding what the tongue, larynx, jaw, and surrounding structures are actually doing. As a result, many players approach overtone matching and altissimo development through trial and error, often compensating with jaw pressure, embouchure distortion, or excess force rather than with efficient voicing adjustments. In response to this problem, I argue that a more anatomically-informed pedagogy can improve technical accuracy, support more efficient practice, and help students develop the self-diagnostic skills necessary for independent learning.

The theoretical basis for that pedagogical shift by drawing on Ingo Titze's concept of nonlinear source-filter coupling and on EVM. Nonlinear source-filter theory challenged the older linear model by demonstrating that the source and resonator do not operate independently, but instead influence one another continuously through acoustic feedback. Applied to saxophone playing, this framework clarifies that the player's oral cavity and vocal tract are not passive spaces through which air merely passes; rather, they actively affect reed behavior, resonance, timbre, response, and stability, especially in the upper register and altissimo. EVM then supplied a practical pedagogical bridge by organizing vocal function into distinct, trainable Figures that can be isolated, sensed, named, and combined. In this context, EVM offered a vocabulary and training structure capable of making internal voicing behaviors more accessible to saxophonists.

Drawing the two frameworks into direct dialogue with Sinta's *Voicing* and argued that Sinta's method can be read as an implicit pedagogy of vocal-tract coordination. Reframed through EVM, the exercises *Voicing* emerge not as isolated tricks for producing overtones or altissimo notes, but as a progressive sequence of coordinative problems involving supraglottic shaping, support, aural intention, and technical integration. Across that sequence, the most consistently relevant Figures were the Tongue, Larynx, and Jaw, supported by Head & Neck and Torso anchoring, with False Vocal Folds functioning as an important diagnostic indicator of *retract* versus *constriction*. By contrast, Figures such as Thyroid Cartilage tilt, Cricoid Cartilage tilt, and True Vocal Folds: Body-Cover proved far less central, which reinforced the need for a selective rather than total transfer of voice-pedagogical concepts into the instrumental domain.

One of the most important outcomes of this study is the claim that EVM does not need to be integrated into saxophone pedagogy as a literal theory of sound production in order to be useful. Saxophone tone is produced by reed vibration rather than by vocal-fold oscillation, and for that reason, not every Figure in EVM maps directly onto the instrument. Nevertheless, the model remains highly valuable as a framework for describing and training the parts of the body that shape the oral cavity and vocal tract. This way, EVM offers teachers a more precise language for identifying whether a student's difficulty is primarily related to tongue position, jaw fixation, laryngeal instability, False Vocal Folds constriction, or inadequate bodily support, rather than leaving such problems under the vague label of poor voicing.

Inner tubing and outer tubing are more than a metaphor, they describe a useful pedagogical concept: the player's internal tract and the saxophone's bore must be understood as a single system whose coordination determines much of what the player experiences as resonance, resistance, timbre, stability, and upper-register access. From this perspective, back pressure, pitch instability, and inconsistent altissimo response are not only signs that the player must blow harder or bite less, but indications that the relationship between the reed, the tract, and the instrument has not yet been optimized. This reframing shifts saxophone instruction away from generalized instruction and toward more specific, reproducible, and anatomically grounded experimentation.

Integrating vocal and saxophone pedagogies is beneficial for students and teachers in several ways. First, a voice-informed approach provides teachers with a shared anatomical vocabulary that can improve communication in applied lessons. Instead of relying exclusively on imagery such as "more open," "more focused," or "sing the note," instructors can pair those cues with more specific descriptions involving tongue height, larynx position, jaw freedom, False Vocal Folds retraction, and postural anchoring. This integrated pedagogy also supports the development of student independence. When students can identify which of the internal adjustments are producing success or failure, they are better able to diagnose their own difficulties and to practice with greater intention and efficiency. Additionally, the framework may help reduce the formation of harmful compensatory habits by directing attention toward internal coordination rather than external force.

Ultimately, this document has argued that saxophone voicing should no longer be taught as a mysterious internal talent that some students simply discover and others do not. When examined through nonlinear source-filter coupling and interpreted through EVM, voicing becomes a coordinated and teachable process. Sinta's *Voicing* remains a valuable pedagogical resource, but its effectiveness can be expanded when its exercises are supported by a more explicit understanding of anatomy and acoustics. Integrating vocal pedagogy into saxophone pedagogy is not a departure from established traditions, but a continuation of its deepest goal: helping players understand how their bodies participate in the making of sound. A pedagogy that teaches saxophonists to hear, feel, name, and intentionally adjust their inner tubing offers a path toward more efficient practice, reliable altissimo technique, nuanced tone control, and greater artistic freedom.

APPENDIX: FIGURES IN SINTA'S MODES

First-Mode:

Primary Figures	Secondary Figures	Low-Applicability
Tongue, Larynx, Jaw, Head & Neck, Torso	False Vocal Folds, True Vocal Folds, Lips	Thyroid Cartilage, Cricoid Cartilage, Aryepiglottic Sphincter

Second-Mode:

Primary Figures	Secondary Figures	Low-Applicability
Tongue, Larynx, Jaw	False Vocal Folds, True Vocal Folds, Head & Neck, Torso	Lips, Thyroid Cartilage, Cricoid Cartilage, Aryepiglottic Sphincter

Third and Fourth-Modes:

Primary Figures	Secondary Figures	Low-Applicability
Tongue, Larynx, Jaw	False Vocal Folds, Aryepiglottic Sphincter, Head & Neck, Torso	Lips, Thyroid Cartilage, Cricoid Cartilage, True Vocal Folds

Combining Modes:

Primary Figures	Secondary Figures	Low-Applicability
Tongue, Larynx, Jaw	True Vocal Folds, False Vocal Folds, Head & Neck, Torso	Lips, Thyroid Cartilage, Cricoid Cartilage, Aryepiglottic Sphincter

Altissimo:

Primary Figures	Secondary Figures	Low-Applicability
Tongue, Larynx, Head & Neck, Torso	Jaw, False Vocal Folds, Aryepiglottic Sphincter	True Vocal Folds, Lips, Thyroid Cartilage, Cricoid Cartilage

BIBLIOGRAPHY

Benson, Elizabeth Ann. “The Estill Voice Model: Theory & Translation.” Voice and Speech Review 11, no. 1 (2017): 112-114.

“Bernoulli Effect,” The Voice Science Lexicon, accessed July 22, 2026, https://www.voicescience.org/lexicon/bernoulli-effect/

Berroth, Jennifer, and Swanson, Lynn. “Voice Anatomy for the Choir Classroom - Demonstrations and Activities,” ChoralNet, posted September 04, 2018, accessed March 18, 2025, https://choralnet.org/archives/576781

Estill, Jo. Voice Craft: An Alternative Method for Training Voices. Oral presentation presented to the Australian and New Zealand Association of Teachers of Singing, Sydney, August 3, 1990.

Estill Voice Training. “Estill Figures for Beginners Class.” YouTube video, April 14, 2022. https://www.youtube.com/watch?v=Na30brdz7ag.

“Estill Voice Training.” Estill Voice Training. Accessed February 05, 2024. https://estillvoice.com/about/.

Fant, Gunnar. Acoustic Theory of Speech Production. The Hague: Mouton, 1970.

Kelso, J. A. S. Dynamic Patterns: The Self-Organization of Brain and Behavior. Cambridge, MA: The MIT Press, 1995.

Kierspe, Jonathan. A Theoretical Application of the Estill Voice Model for the Saxophone. DMA diss., Florida State University, 2022.

Kob, Malte, and Carolin Neuschaefer-Rube. “Vocal Tract and Subglottal Impedance in High Performance Singing.” Journal of Voice (forthcoming).

Loboda, Emily Jane. An Analysis of Overtone Production Techniques in Saxophone Teaching Methods. DMA diss., University of North Carolina at Greensboro, 2018.

Mata-Pose, Lucía, Miguel Mayo-Yáñez, Carlos M. Chiesa-Estomba, Jérôme R. Lechien, Luigi A. Vaira, Antonino Maniaci, Alberto M. Saibene, Christian Calvo-Henríquez, and Irma Cabo-Varela. 2023. "Velopharyngeal Incompetence in Musicians: A State-of-the-Art Review" Journal of Personalized Medicine 13, no. 10: 1477. https://doi.org/10.3390/jpm13101477

Raschèr, Sigurd M. Top-Tones for the Saxophone: Four-Octave Range. Elkhart, IN: Carl Fischer, 1941.

"Research & Publications." Estill Voice International. Accessed September 27, 2023. https://estillvoice.com/research/.

Rothenberg, Martin. "Acoustic Interaction Between the Glottal Source and the Vocal Tract." In Vocal Fold Physiology, edited by K. N. Stevens and M. Hirano, 305-328. Tokyo: University of Tokyo Press, 1981.

Steinhauer, Kimberly. "The Estill Voice Model: A Paradigm for Voice Training and Treatment." Journal of the Acoustical Society of America 154, no. 4, suppl. (October 2023): A352-A353.

Steinhauer, Kimberly, Judith Preston Grayhack, Ann L. Smiley-Oyen, Susan Shaiman, and Malcolm R. McNeil. "The Relationship among Voice Onset, Voice Quality, and Fundamental Frequency: A Dynamical Perspective." Journal of Voice 18, no. 4 (2004): 432-442.

Steinhauer, Kimberly, Mary McDonald Klimek, Mara Behlau, and colleagues. "Connecting Auditory-Perceptual Prompts Used in Voice Therapy to Anatomy and Physiology via the Estill Voice Model and the Rehabilitation Treatment Specification System." American Journal of Speech-Language Pathology (2024).

Steinhauer, Kimberly, Mary McDonald Klimek, and Jo Estill. The Estill Voice Model: Theory and Translation. Pittsburgh: Estill Voice International, 2017.

Sinta, Donald. Voicing: An Approach to the Saxophone's Third Register, rev. ed. East Lansing, MI: Blaris Publications, 2008.

Spadaro, Esther, and Kimberly McInnis. The Estill Adventure: Figures for Beginners. Pittsburgh, PA: Estill Voice International, 2018.

Teal, Larry. The Art of Saxophone Playing. Princeton, NJ: Summy-Birchard, 1963.

"The Estill Adventure: Figures for Beginners Teaching Package." Estill Voice International. Accessed March 18, 2026. https://store.estillvoice.com/pages/the-estill-adventure-figures-for-beginners-teaching-package.

Titze, Ingo R. "Nonlinear Source-Filter Coupling in Phonation: Theory." Journal of the Acoustical Society of America 123, no. 5 (2008): 2733-49.

Titze, Ingo R. "Nonlinear Source-Filter Coupling in Phonation: Vocal Exercises." Journal of the Acoustical Society of America 123, no. 4 (2008): 1902-15.

Titze, Ingo R. "Nonlinear Source-Filter Interaction in Singing." Paper presented at the Pan-American/Iberian Meeting on Acoustics, Cancun, Mexico, 2004.

Titze, Ingo R. Principles of Voice Production. Englewood Cliffs, NJ: Prentice Hall, 1994.

Titze, Ingo R., Allison Worley, and Brad H. Story. "Source-Vocal Tract Interaction in Female Operatic Singing and Implications for Vocal Efficiency." Journal of Singing 67, no. 5 (2011): 561-72.

Titze, Ingo R. "The Human Voice and Its Nonlinear Physics." In Nonlinear Dynamics in Physiology and Medicine, edited by Anne Beuter et al., pages 1-35. New York: Springer, 2003.

Tokuda, Isao T., et al. "New Evidence That Nonlinear Source-Filter Coupling Affects Harmonic Intensities and Timbre." Journal of the Acoustical Society of America 140, no. 1 (2016): 369-79.

"Voice," Cambridge Dictionary Online, accessed March 18, 2026, https://dictionary.cambridge.org/dictionary/english/voice.

"What Wind Instrument Should I Play? A Look at 5 Popular Wind Instruments and Factors To Help You Know Which One To Choose," Northwest School of Music, accessed March 18, 2025, https://nwschoolofmusic.com/what-wind-instrument-should-i-play/.