Flute in the Throat:

A Vortex-Acoustic Lock-In Hypothesis for Overtone Singing

Ioannis Psallidakos

Abstract

Overtone singing, particularly the Tuvan *sygyt* style, achieves selective harmonic amplification with formant bandwidths on the order of 20–40 Hz and neighbor-harmonic attenuation exceeding 35 dB. The foundational work of researchers including Levin and Edgerton (1999), Adachi and Yamada (1999), Kob (2004), Saus (2009), and Bergevin et al. (2020) has established formant convergence as the primary acoustic mechanism. However, a bandwidth of 20–40 Hz at 1500 Hz implies an effective Q-factor of 40–75—values that may exceed what passive soft-tissue resonators typically achieve (Q on the order of 10–20, or up to 15–40 under optimized conditions). The hypothesis is most strongly motivated by the extreme end of this range. This discrepancy suggests that purely passive linear models, while descriptively accurate for frequency selection, may not fully account for the energetic profile (bandwidth and amplitude). We propose a **hybrid linear-nonlinear framework** that extends current understanding by integrating three components: (1) a **supra-lingual Helmholtz cavity** with **inter-molar lateral orifices** as a previously uncharacterized anatomical structure; (2) **vortex-acoustic coupling (lock-in)** providing regenerative, nonlinear energy transfer that offsets tissue damping; and (3) a **multi-parameter control framework** explaining performer variability and skill progression. Demetrio Stratos (1978) demonstrated 'flautofonia'—flute-like sounds via supraglottic resonance without vocal fold vibration. This hypothesis proposes a complementary mechanism: analogous flow-acoustic coupling preserving modal phonation, achieving spectral focusing through vortex lock-in rather than replacing the glottal source. This framework generates testable predictions amenable to empirical investigation.

Keywords: overtone singing, throat singing, sygyt, Helmholtz resonator, vortex-acoustic coupling, laminar vortex shedding, edge-tone, Q-factor, source-filter theory, formant bandwidth, biphonic singing, vocal tract acoustics

1. Introduction: The Q-Factor Paradox

Overtone singing presents one of the most remarkable phenomena in human vocalization. Through precise articulatory control, singers produce two simultaneously perceived pitches: a low-frequency drone and a high-frequency melodic tone corresponding to a selectively amplified harmonic. The *sygyt* style of Tuvan throat singing represents the most acoustically extreme variant, producing whistle-like melodic overtones of exceptional clarity in the 1–2 kHz range.

The scientific study of this phenomenon has benefited enormously from dedicated researchers. The pioneering work of Levin and Edgerton (1999) brought serious scientific attention to Tuvan throat singing. Adachi and Yamada (1999) provided foundational acoustic analysis. Kob (2004) established important relationships between vocal tract geometry and overtone production. Saus (2009) contributed valuable insights into formant control and the role of the sublingual cavity. Bloothooft et al. (1992) reported second-formant bandwidths that seldom exceed 40 Hz and are often much smaller, and Klingholz (1993) similarly reported very narrow bandwidths in expert overtone singers. Bergevin et al. (2020) provided comprehensive analysis demonstrating F2-F3 formant convergence with dynamic MRI visualization.

This substantial body of work explains *which* acoustic phenomenon occurs (formant convergence) and *where* in the vocal tract it originates. However, a quantitative puzzle remains: the **Q-factor paradox**.

1.1 The Physics of the Problem

The Q-factor (quality factor) of a resonator is defined as the ratio of resonant frequency to bandwidth:

$$Q = f_0 / \Delta f$$

For expert overtone singing with a target frequency around 1500 Hz and bandwidth on the order of 20–40 Hz:

$$Q = 1500 / 30 \approx 50 \quad (range: 40–75)$$

These are remarkably high values. Biological tissues—mucosa, muscle, cartilage—are soft and highly damped. Passive vocal tract resonances are typically modeled with Q-factors on the order of **10–20** (Titze, 1994; Story, 2002), though measured singing formants can achieve Q ≈ 15–40 under optimized conditions (Sundberg, 1987). The Q-factors achieved in expert overtone singing (≈ 40–75) may still exceed even these optimized passive limits, though direct in-vivo measurement during sygyt is needed to confirm this. The hypothesis is most strongly motivated by the extreme end of reported bandwidths (≤20 Hz, Q ≈ 75); if typical expert performance falls closer to Q ≈ 40, passive optimization alone might suffice.

1.2 The Central Question

While linear source-filter models correctly predict frequency selection (which harmonic gets amplified), they do not as readily predict the energetic profile (how narrow the bandwidth becomes, how extreme the neighbor attenuation). This suggests two possibilities:

- Either passive biological resonators can achieve Q ≈ 40–75 under conditions not yet fully understood, or
- An active, regenerative mechanism supplements passive filtering—analogous to how a flute maintains high Q through jet-edge feedback rather than passive tube resonance alone.

This paper explores the second possibility.

2. What Linear Models Explain—and Where They Reach Their Limits

2.1 The Success of Source-Filter Theory

The source-filter model of speech production, formalized by Fant (1960) and developed extensively by Stevens (1998), treats vocalization as a two-stage process: the glottal source produces a harmonic-rich spectrum, which is then shaped by the resonant properties (formants) of the vocal tract filter.

As Bergevin et al. (2020) demonstrated, the biphonic sound of throat singing arises from linear filtering—specifically, the convergence of F2 and F3 onto a single harmonic frequency. They showed that "nonlinearities are not a priori essential" for explaining the basic phenomenon.

This linear framework correctly predicts:

- The relationship between tongue position and overtone frequency
- The role of front-cavity and back-cavity geometry in formant control
- The general mechanism of harmonic selection through formant convergence
- The spectral location of the focused overtone

2.2 The Damping Problem

However, linear models assume passive filtering, where the vocal tract shapes but does not add energy to the acoustic signal. Under passive conditions, bandwidth is determined by damping—energy losses to tissue absorption, viscous friction, and radiation.

The challenge is quantitative:

- Formant convergence (F2 + F3 merging) **doubles effective amplitude but does not inherently narrow bandwidth**—damping remains governed by tissue losses ($Q \approx$ 10–20). To achieve $Q \approx 40$–75 requires additional mechanisms beyond peak sharpening
- Damping in soft-tissue systems is substantial due to viscoelastic losses
- Even optimized passive resonators in biological systems are typically modeled with Q on the order of 10–20

To achieve $Q \approx 40$–75, damping would need to be reduced substantially below typical values. Passive linear models do not offer a clear mechanism for this.

2.3 Observations Suggesting Additional Mechanisms

Extreme Harmonic Attenuation: Bergevin et al. (2020) report that neighboring harmonics are attenuated by 15–35 dB, with harmonics two steps removed attenuated by 35–65 dB. While formant convergence concentrates energy at the target frequency, the extreme suppression of neighbors—particularly beyond 35 dB—suggests mechanisms beyond simple resonant selectivity.

Dynamic Stability: Expert singers maintain stable overtone production despite variations in breath pressure and articulatory micro-movements. This robustness suggests feedback mechanisms that passive linear models do not incorporate.

Expert-Novice Gap: Years of training are required to achieve master-level production, yet the basic articulatory positions are not extraordinarily complex. What exactly are experts learning that novices lack?

Alternative Mechanisms to Consider: Several other mechanisms could contribute to narrow bandwidths: glottal pulse sharpening (Bloothooft et al., 1992), lip radiation impedance effects (Sundberg et al., 2021), and piriform fossa zeros (Takemoto et al., 2006). We do not exclude these contributions but propose that lateral-channel mechanisms may provide an additional, testable component. Predictions 8.2.1–8.2.4 are designed to distinguish lateral-specific effects from these alternatives.

2.4 A Conceptual Summary

Aspect	Linear Model Predicts	Expert Singers Achieve
Frequency selection	Correct (via F2-F3 convergence)	Matches prediction
Formant bandwidth	50–150 Hz (passive Q ≈ 10–20)	20–40 Hz (effective Q ≈ 40–75)
Neighbor attenuation	10–20 dB typical	35–65 dB observed
Energy source	Passive shaping only	Possible active component?

3. A Hybrid Linear-Nonlinear Framework

3.1 Central Thesis

We propose that expert overtone singing involves a **hybrid system** combining:

- **Linear resonance:** Vocal tract geometry sets formant frequencies through passive acoustic filtering (as established by existing research)
- **Nonlinear flow-acoustic coupling:** Airflow through narrow orifices creates vortex shedding that, under appropriate conditions, locks onto acoustic resonances—providing regenerative energy transfer that offsets tissue damping
- **Active motor control:** Real-time articulatory adjustments maintain the locked state and navigate a multi-parameter control space

This is analogous to how wind instruments achieve high Q-factors. A flute's resonance is not simply passive tube acoustics—the jet-edge interaction continuously **injects energy** into the acoustic mode, compensating for damping losses. We propose that the human vocal tract, under specific configurations, may operate similarly.

3.2 Why "Hybrid" Rather Than "Nonlinear"?

We emphasize "hybrid" because the bulk of acoustic behavior remains linear:

- The glottal source is not modified

- Formant frequencies are still determined by vocal tract geometry
- The basic spectral shaping follows source-filter principles

The nonlinear element is localized and specific: flow-acoustic coupling at narrow orifices that injects energy to offset damping. This extends rather than replaces linear acoustics.

3.3 Scope of the Proposed Nonlinearity

It should be noted that other nonlinear phenomena exist in voice production—including source-filter interaction at high sound pressure levels, subharmonic generation, and chaotic dynamics in certain phonation modes. Nonlinear source-filter coupling (Titze, 2008) can narrow spectral peaks via glottal flow skewing when harmonics approach formants. However, sygyt's extreme 35–65 dB neighbor suppression and $Q \approx 40–75$ exceed what source adjustments alone typically achieve (Story, 2002). We do not dismiss these general nonlinear effects but posit an **additional** flow-acoustic mechanism in lateral channels as a candidate explanation for the most extreme selectivity observed. Whether this mechanism operates independently of, or in conjunction with, source-filter coupling remains an open question for future investigation.

4. Component 1: The Supra-Lingual Helmholtz Resonator

4.1 Relationship to Previous Work

Saus (2009) proposed that the **sublingual** space (*below* the tongue) contributes to F3 control, noting that this space extends laterally around the tongue frenulum. Sundberg et al. (2021) demonstrated that the front cavity functions as a Helmholtz resonator with the lip opening as the neck. We propose a **complementary** structure: a **supra-lingual** cavity (*above* the tongue) with a **different** orifice location.

These are geometrically distinct:

- **Sublingual (Saus):** Below the tongue, between tongue underside and floor of mouth, extending laterally around the frenulum
- **Supra-lingual (proposed):** Above the tongue, between tongue dorsum and hard palate, with lateral exit pathways

Both may contribute to overtone production—potentially explaining the multiple focused states observed in some singers (Bergevin et al., 2020, Singer T2).

4.2 Anatomical Description

The Supra-Lingual Cavity (Volume):

- The tongue dorsum assumes a concave (bowl-shaped) configuration beneath the hard palate, forming acoustic cavity volume (V)
- Sealed anteriorly and laterally by U-shaped tongue-palate/dental contact, open posteriorly to pharynx via dynamic posterior constriction (tongue-uvula space)
- Cavity volume adjustable via tongue dorsum curvature

The U-Shaped Perimeter Seal:

- Lateral tongue edges contact the palate and upper dental arch along the sides
- Anterior boundary: Tongue body curvature (tip retracted/lowered per sygyt double-cavity configuration, Bergevin et al., 2020)
- This creates a U-shaped seal with the open end oriented posteriorly toward the pharynx

The Inter-Molar Lateral Orifices (Neck):

The critical orifices are hypothesized to be located between the upper and lower molars—in the lateral gap of the dental arch:

- When the jaw is partially open, a gap exists between the upper and lower molar rows
- The buccal (cheek) wall forms the outer boundary of this gap
- Air from the supra-lingual cavity would exit through these inter-molar orifices into the buccal space and toward the lips
- The orifice geometry would be controlled by jaw opening (vertical dimension) and cheek/buccal wall tension (aperture area)—analogous to whistling, where lip-cheek tension and jaw position precisely tune Helmholtz neck geometry (Fletcher & Rossing, 1998)

4.3 Current Status in the Literature

Before proceeding, this hypothesis focuses exclusively on **supraglottal tract mechanisms** for *sygyt*'s extreme harmonic selectivity. Unlike *kargyraa* (subharmonic generation via false vocal fold vibration), *sygyt* employs **true vocal folds only** with stable modal phonation and **linear source characteristics** (Bergevin et al., 2020).

Published MRI limitation: All overtone singing studies **focus analysis on midsagittal views**, examining tongue-palate constriction, front/back cavity division, and lip rounding:

Study	MRI Type	Analysis Plane	Lateral Channels?
Bergevin et al. (2020)	Volumetric + Dynamic	3D (all planes); analysis midsagittal	3D data exists; lateral features not analyzed
Barbiera et al. (2022)	Dynamic	Midsagittal	Not reported
Saus (videos)	Dynamic	Midsagittal	"Frenulum covers lateral space" (Saus, oberton.org)
Hefele et al. (2019)	Dynamic	Midsagittal	Not reported

Contrast with speech MRI: 3D dynamic studies reveal lateral channels in /l/ production, where tongue sides lower around central occlusion (Fu et al., 2017; Kim et al., 2012).

Coronal/axial planes visualize these channels clearly—yet remain **unanalyzed in overtone singing** despite 3D data availability (Bergevin et al., 2020).

Conclusion: Supra-lingual lateral orifices remain **invisible in current analyses** (midsagittal focus) but are **acoustically predicted** and **testable via reanalysis of existing 3D data or new coronal/axial MRI** (Prediction 8.1.1).

4.4 Acoustic Function

If present, this structure would constitute a Helmholtz resonator with distinctive properties:

Resonance (Pole): The supra-lingual volume and inter-molar orifice area would determine a resonant frequency per:

$$f = (c/2\pi) \times \sqrt{(A / VL)}$$

where c is sound velocity, A is orifice area, V is cavity volume, and L is effective neck length.

Anti-Resonances (Zeros): The lateral channels would function as side-branches to the main vocal tract pathway. Side-branch resonators introduce zeros (anti-resonances) into the transfer function, suppressing energy at specific frequencies (Stevens, 1998). This provides a candidate mechanism for the extreme neighbor-harmonic attenuation observed.

Combined Pole-Zero Response: A sharp bandpass at the target frequency with notches at adjacent frequencies—explaining both the amplification of the target and suppression of neighbors.

Lateral Channel Asymmetry: Importantly, lateral channel asymmetry would further enhance selectivity. MRI studies of lateral and rhotic consonants reveal that symmetric channels (left = right) produce canceling anti-resonances with minimal spectral effect, while asymmetric channels (one dominant side) generate strong zeros in the F3–F5 range. Zhou et al. (2008) demonstrated this principle for American English /r/, where asymmetric sublingual channels produce stronger anti-resonances than symmetric ones—a principle applicable to lateral structures generally. Szalay and Proctor (2025) confirmed similar effects for /l/. In sygyt, singers may preferentially develop one dominant lateral orifice for maximal neighbor suppression. This would explain why unilateral occlusion degrades but does not abolish production (see Prediction 8.4.2), and why different singers may show left or right dominance based on individual anatomy.

5. Component 2: Vortex-Acoustic Lock-In

5.1 Theoretical Background

Airflow through narrow inter-molar orifices ($A \approx 0.1$ cm², $Re \approx 1{,}500$) produces laminar vortex shedding at tongue-edge intrusions—analogous to edge-tone generation in flutes. The shedding frequency is governed by the Strouhal number:

$$St = fL/U$$

where f is shedding frequency, L is orifice dimension, and U is flow velocity. For most orifice geometries, St ≈ 0.2–0.5.

When shedding frequency approaches an acoustic resonance, **vortex-acoustic lock-in** occurs. Dai et al. (2015) demonstrated that in flow-excited Helmholtz resonators, "the vortical flow excites the acoustic mode of the resonator which in turn provides feedback on the vorticity motion. This vortex-acoustic coupling results in a **self-sustained oscillation** in the resonator."

5.2 Feasibility Check: Can Lock-In Occur Under Physiological Conditions?

A critical question is whether flow conditions during sygyt singing fall within the range where lock-in is possible:

Orifice Dimensions: Gross inter-molar gap measures approximately 0.5–1 cm (vertical height, jaw opening) × 0.5–1 cm (anteroposterior length along molars). However, the lateral tongue edge dynamically intrudes between upper/lower molars on the active side, reducing effective length to approximately 0.1–0.2 cm while one side remains fully occluded via tongue-cheek contact (analogous to /ŋ/ production). This yields A ≈ 0.1 cm^2—precisely matching physiological requirements for lock-in.

Flow Rate: Sustained singing typically involves flow rates of 0.1–0.2 L/s.

Flow Velocity: Through the orifice: U = Flow/Area ≈ 0.15 L/s ÷ 0.1 cm^2 = 1500 cm/s ≈ 15 m/s.

Strouhal Check: For f = 1500 Hz and L = 0.5 cm:

$$St = fL/U = (1500 \times 0.005) / 15 = 0.5$$

This falls at the upper end of the typical range for whistle-type mechanisms (St ≈ 0.2–0.5). The hypothesis is **mathematically plausible** under realistic physiological conditions, though closer to the higher Strouhal values associated with edge-tone rather than pure Helmholtz excitation. Inter-molar geometry (tongue edge intrusion) may provide the necessary edge-tone excitation for lock-in, analogous to flute jet-edge interaction.

5.3 Consequences of Lock-In

If lock-in occurs, several consequences would follow:

- **Regenerative Energy Transfer:** Energy flows continuously from the airstream into the acoustic mode. This *offsets damping losses*, raising the effective Q-factor beyond typical passive limits.
- **Bandwidth Narrowing:** The coupled system exhibits higher effective Q than the passive cavity alone—potentially explaining Q ≈ 40–75.
- **Frequency Stabilization:** The locked frequency remains stable over a range of flow velocities (the "lock-in range"), explaining the dynamic stability of expert production.

- **Enhanced Contrast:** Non-locked frequencies receive no energy input and are passively damped, further increasing the contrast between target and neighbor harmonics.

5.4 The Flute Analogy

This mechanism is not unprecedented. In a flute:

- The jet-edge interaction creates periodic vortex shedding
- This shedding locks onto the tube resonance
- Energy transfers from the airstream to the acoustic mode
- The result is a sustained tone with high Q—far higher than passive tube resonance would produce

We propose that the supra-lingual cavity with inter-molar orifices may create an analogous system within the human vocal tract. Unlike the flute, where jet-edge interaction generates the primary sound, in sygyt the glottal source provides harmonic-rich input; the lateral orifice mechanism selectively amplifies one harmonic via the same lock-in physics—enhancing filter Q rather than replacing the source.

6. Component 3: Multi-Parameter Control Framework

6.1 The Control Space Concept

Pedagogical observation reveals that teachers across traditions—classical voice acoustics (Bozeman), extended techniques (Howell), Western overtone pedagogy (Saus, van Tongeren, Curtet, Hinds, Hefele, Tran Quang Hai), and traditional Tuvan masters—emphasize different articulatory strategies yet achieve similar acoustic outcomes. This suggests that F2-F3 clustering (and Helmholtz tuning) is achieved not by a single configuration but by a **manifold of configurations** satisfying the same acoustic constraint.

6.2 Three Control Parameters

We identify three anatomical adjustments that independently affect **biphonic overtone frequency**. All move frequency in the **same direction** (larger/wider → higher frequency), **making the system monotonic**—enabling "bi-phonic" (two perceived pitches)—strongly compatible with learned biological control systems:

6.2.1 Lip Opening Aperture

- **Mechanism:** Increased lip opening raises radiation impedance, shortens effective front cavity, pushes F2-F3 upward
- **Characteristics:** Coarse control, relatively high acoustic losses, tends to *broaden* bandwidth due to increased radiation damping
- **Skill level:** Accessible to beginners; sufficient for entry into overtone regime but suboptimal for extreme selectivity

6.2.2 Tongue-Uvula Constriction

- **Mechanism:** The posterior constriction between tongue dorsum and uvula/soft palate—corresponding to the uvular/pharyngeal constriction documented in sygyt MRI (Bergevin et al., 2020)—serves dual functions: coupling neck between pharyngeal and oral cavities, and inertive flow conditioning for downstream resonance
- **Characteristics:** Medium precision, well-documented in existing MRI studies
- **Skill level:** Intermediate—requires refined control beyond simple lip adjustment

6.2.3 Lateral Helmholtz Geometry (Proposed)

- **Mechanism:** Bowl depth (volume V) and inter-molar orifice aperture (area A) control Helmholtz frequency per $f \propto \sqrt{(A/VL)}$
- **Characteristics:** Fine control, high Q-factor preserved because energy exits through orifices optimized for lock-in rather than through lossy radiation
- **Skill level:** Expert—this may be what masters learn over years of training, controlling structures (tongue lateral edges, jaw opening, cheek tension) not consciously accessible to novices

6.3 Skill Hierarchy

Level	Primary Control	Mechanism	Typical Bandwidth
Beginner	Lip aperture	Passive F2-F3 convergence	50–100 Hz
Intermediate	Tongue-uvula space	Optimized convergence	30–50 Hz
Expert	Lateral Helmholtz	+ Helmholtz resonance	20–40 Hz
Master	All three + neural predictive control	+ Vortex lock-in + auditory-motor feedback	≤20–30 Hz (approaching narrowest reported)

6.4 Monotonicity and Learnability

All three parameters move frequency in the same direction: larger opening / wider orifice / shallower bowl → higher frequency. This monotonicity has important implications:

- **Self-correcting:** If one parameter drifts, others can compensate
- **Fast learning:** No conflicting feedback signals
- **High stability:** Multiple valid configurations for each target frequency

This is strongly compatible with what is typically found in learned biological control systems, where monotonic mappings support robust learning and stability.

6.4.1 Neural Basis of Mastery

The monotonic control manifold supports rapid learning, but mastery requires **experience-dependent neuroplasticity**. Neuroimaging studies demonstrate enhanced somatosensory cortex, premotor cortex (PMC), and arcuate fasciculus connectivity in expert singers, correlating with practice hours (Kleber et al., 2010; Zarate & Bhagat, 2010). Masters likely develop what Saus (2017) terms **"overtone hearing"**—the ability to perceive formants as discrete pitches rather than timbral coloration—enabling rapid feedback corrections via strengthened auditory-motor loops (posterior superior temporal sulcus → dorsal premotor cortex → primary motor cortex).

This neural infrastructure may be essential for maintaining vortex lock-in despite physiological perturbations. While the anatomical structures provide the *substrate* for high-Q resonance, the trained auditory-motor network provides the *control precision* to keep the system in the lock-in regime. The combination of favorable physics (monotonic control space, regenerative coupling) and refined neural control may explain why Tuvan masters achieve seemingly inhuman precision after years of dedicated practice.

7. What This Framework Explains

7.1 The Q-Factor Paradox

The central puzzle—how $Q \approx 40–75$ is achieved with soft-tissue resonators typically modeled at $Q \approx 10–20$—receives a candidate answer:

- **Passive filtering alone** would not typically achieve this Q (damping too high)
- **Regenerative energy injection** via vortex-acoustic lock-in could offset damping losses
- The system would function not as a passive filter but as an **active oscillator**—like a flute, not like a bottle

7.2 Extreme Neighbor-Harmonic Attenuation

The 35–65 dB attenuation of neighboring harmonics could result from multiple synergistic mechanisms:

- **Anti-resonances (zeros):** Lateral channels as side-branches actively suppress energy at specific frequencies
- **Lock-in selectivity:** Only the locked frequency receives regenerative energy; others are passively damped
- **Contrast enhancement:** Active gain at target + active/passive suppression at neighbors
- **Asymmetry enhancement:** Expert singers may favor one dominant lateral channel over symmetric pairing, as asymmetric side-branches produce stronger anti-resonances than symmetric ones that cancel each other out (Szalay & Proctor, 2025)

7.3 Expert-Novice Differences

The multi-parameter framework explains what experts may be learning:

- Beginners achieve overtones via **lip control**—coarse but accessible

- Experts master **lateral Helmholtz control**—fine, high-Q, requiring control of structures not consciously accessible
- Masters achieve **lock-in conditions**—coordinating geometry and flow for regenerative coupling

Years of training would involve discovering and refining control over structures (lateral tongue edges, inter-molar orifice geometry, cheek tension) that are invisible in midsagittal imaging and not part of normal speech articulation.

7.4 Pedagogical Variability

Different teachers emphasize different parameters. The multi-parameter framework suggests this represents legitimate path diversity—multiple routes through the same control space to the same acoustic target—rather than contradiction.

8. Testable Predictions

A hypothesis is valuable only if it can be **falsified**. The following predictions, if disconfirmed, would weaken or refute the framework.

8.1 Anatomical Predictions

1. **Coronal/Axial MRI:** Imaging in the coronal and/or axial planes during sygyt production should reveal: (a) concave supra-lingual air space, (b) lateral tongue-palate contact forming U-shaped seal, (c) inter-molar orifices. If these structures are absent in expert singers, the anatomical hypothesis is falsified.
2. **Expert vs. Novice Comparison:** Experts should show more defined lateral seal formation and more controlled inter-molar orifices than novices producing weaker overtones.
3. **Dynamic Correlation:** Changes in overtone pitch should correlate with changes in inter-molar orifice aperture and/or bowl volume.

8.2 Acoustic Predictions

1. **Transfer Function Poles and Zeros:** Detailed acoustic analysis should reveal both poles (resonances) and zeros (anti-resonances) attributable to lateral structures, distinct from both midsagittal F2-F3 convergence and piriform fossa zeros (which affect F3-F4 but are not position-dependent on lateral occlusion). Lateral zeros should shift or disappear with inter-molar occlusion while piriform zeros remain stable.
2. **Bilateral Manipulation:** Asymmetric blocking of inter-molar orifices (e.g., one side only) should produce predictable, asymmetric spectral changes.
3. **Bandwidth-Orifice Correlation:** Bandwidth should correlate with lateral orifice geometry—conditions favoring lock-in should produce narrower bandwidth.
4. **Channel Asymmetry:** Spectral analysis across expert singers should reveal consistent left/right dominance correlating with stronger neighbor suppression, paralleling /l/ findings where asymmetric channels outperform symmetric ones (Szalay & Proctor, 2025).

8.3 Aeroacoustic Predictions

1. **Flow Velocity Measurements:** Hot-wire anemometry or particle image velocimetry at inter-molar orifices should detect periodic flow fluctuations consistent with vortex shedding during overtone production. If no periodic flow structure is detected, the lock-in hypothesis is weakened.
2. **Lock-In Range:** Systematic pressure variation should reveal a frequency-stable "lock-in range" followed by mode transitions outside this range.
3. **Strouhal Number:** The measured St = fL/U should fall within 0.2–0.5 during overtone production.

8.4 Interventional Predictions

1. **Orifice Occlusion:** External pressure closing the inter-molar orifices should abolish or severely degrade overtone production even with midsagittal configuration maintained.
2. **Unilateral vs. Bilateral:** Blocking the dominant lateral orifice should produce greater spectral degradation than blocking the weaker side, while bilateral blocking abolishes production entirely. This follows from /l/ studies where one asymmetric channel dominates anti-resonance generation (Szalay & Proctor, 2025).
3. **Artificial Enhancement:** Prosthetic devices creating artificial lateral channels with appropriate dimensions might facilitate overtone production in less-trained individuals.

9. Discussion

9.1 Relationship to Existing Research

This framework builds on rather than replaces existing research:

Demetrio Stratos (1945–1979), a pioneer who helped introduce overtone singing techniques to European audiences through his collaborations with researchers at the University of Padua (Ferrero et al., 1980) and with Tran Quang Hai, demonstrated 'flautofonia'—flute-like sounds via supraglottic resonance without vocal fold vibration. The present hypothesis proposes a complementary mechanism: analogous flow-acoustic coupling preserving modal phonation, achieving spectral focusing through vortex lock-in rather than replacing the glottal source.

- Bergevin et al. (2020) and others correctly explain **frequency selection** via F2-F3 convergence
- Our hypothesis addresses **bandwidth and amplitude**—the energetic profile—which passive linear models may not fully account for
- We propose that linear filtering is **necessary but may not be sufficient** for the most extreme examples

9.2 Relationship to Saus's Sublingual Model

Saus (2009) proposed that the sublingual cavity (*below* the tongue), extending laterally around the frenulum, contributes to overtone production. The present supra-lingual hypothesis (cavity *above* the tongue) is **complementary, not contradictory**. Both structures may contribute:

- The sublingual cavity may contribute to F3 control as Saus suggests
- The supra-lingual cavity with inter-molar orifices may provide the regenerative mechanism for bandwidth narrowing

Different singers may emphasize one or both, potentially explaining the variability observed across performers.

9.3 Methodological Implications

The focus on midsagittal imaging in published overtone singing studies represents a gap worth addressing. We recommend:

- Coronal and axial MRI sequences during overtone production
- Three-dimensional volumetric acquisition with analysis of lateral structures
- Dynamic (real-time) imaging in multiple planes
- Comparison of expert, intermediate, and novice performers

9.4 Limitations and Epistemic Humility

This hypothesis may be wrong. Several limitations must be acknowledged:

- The proposed structures have not been analyzed in overtone singing imaging—lateral features may exist in published 3D data (Bergevin et al., 2020) but remain unexamined
- Flow velocities during overtone singing are not directly measured in the published literature
- **Compliant tissue damping:** Flow-excited Helmholtz resonators in the engineering literature assume rigid walls (Dai et al., 2015). Mucosal viscoelasticity may prevent or weaken lock-in by broadening resonances, requiring empirical validation in biological tissue
- The author is not a physicist, acoustician, or otolaryngologist—though the hypothesis emerged from extended personal practice and teaching of overtone singing. Expert review is essential
- Individual anatomical variation (dental arch width, molar spacing, tongue size, cheek elasticity) may limit generalizability—some individuals may lack suitable geometry for lateral Helmholtz resonance
- False vocal fold involvement: While sygyt employs modal phonation without false fold vibration (unlike *kargyraa*), dynamic coronal imaging would be needed to definitively rule out any supraglottic contribution

The framework is offered not as established fact but as a structured hypothesis worthy of empirical investigation.

Preliminary Behavioral Validation: Self-observations confirm that unilateral production (one lateral channel active) yields superior spectral focus (narrower bandwidth, stronger neighbor attenuation) compared to symmetric bilateral production. Complete bilateral tongue-cheek occlusion (analogous to /ŋ/ production) blocks lateral airflow, eliminating the hypothesized mechanism. This supports Prediction 8.4.2 and asymmetric channel physics (Szalay & Proctor, 2025).

10. Conclusion

The Q-factor paradox—expert overtone singers achieving Q ≈ 40–75 with soft-tissue resonators typically modeled at Q ≈ 10–20—suggests that purely passive linear models, while correctly describing frequency selection, may not fully account for the energetic profile. The **hybrid linear-nonlinear framework** presented here proposes that:

- A **supra-lingual Helmholtz cavity** with **inter-molar lateral orifices** may provide an additional resonant structure with pole-zero characteristics
- **Vortex-acoustic lock-in** at these orifices could provide regenerative energy transfer that offsets tissue damping, enabling Q-factors beyond typical passive limits
- A **multi-parameter control space** explains why different articulatory strategies achieve similar acoustic outcomes, and why years of training are required to master the most refined techniques

Rigorous empirical testing is needed to confirm or refute these proposals. If confirmed, this framework would represent a step toward understanding not just what overtone singers achieve, but how they achieve it—with implications for voice science, pedagogy, and the broader study of biological flow-acoustic systems.

We offer this hypothesis in a spirit of scientific humility, deeply grateful for the substantial body of research on which it builds, and hopeful that it may contribute to understanding this extraordinary human vocal capability.

Acknowledgments

The author gratefully acknowledges the foundational contributions of researchers across voice science, ethnomusicology, and vocal pedagogy whose work made this synthesis possible, including Theodore Levin, Michael Edgerton, Seiji Adachi, Malte Kob, Wolfgang Saus, Mark van Tongeren, Tran Quang Hai, Johanni Curtet, Stuart Hinds, Anna-Maria Hefele, Ingo Titze, Brad Story, Kenneth Bozeman, Ian Howell, Christopher Bergevin, and the pioneering vocal explorations of Demetrio Stratos, among many others who have advanced our understanding of voice acoustics and overtone singing. Any errors or overreach in this work are entirely the author's responsibility.

References

Adachi, S., & Yamada, M. (1999). An acoustical study of sound production in biphonic singing, Xöömij. Journal of the Acoustical Society of America, 105(5), 2920-2932.

Barbiera, F., Lo Casto, A., Murmura, B., Bortoluzzi, G., Orefice, I., & Gucciardo, A. G. (2022). Dynamic Fast Imaging Employing Steady State Acquisition Magnetic Resonance Imaging of the Vocal Tract in One Overtone Male Singer. Journal of Voice, 36(2), 170-175.

Bergevin, C., Narayan, C., Williams, J., Mhatre, N., Steeves, J. K., Bernstein, J. G., & Story, B. (2020). Overtone focusing in biphonic Tuvan throat singing. eLife, 9, e50476.

Bloothooft, G., Bringmann, E., van Cappellen, M., van Luipen, J. B., & Thomassen, K. P. (1992). Acoustics and perception of overtone singing. Journal of the Acoustical Society of America, 92(4), 1827-1836.

Bozeman, K. W. (2013). Practical Vocal Acoustics: Pedagogic Applications for Teachers and Singers. Pendragon Press.

Dai, X., Jing, X., & Sun, X. (2015). Flow-excited acoustic resonance of a Helmholtz resonator: Discrete vortex model compared to experiments. Physics of Fluids, 27(5), 057102.

Dai, X. (2016). Vortex convection in the flow-excited Helmholtz resonator. Journal of Sound and Vibration, 371, 292-309.

Fant, G. (1960). Acoustic Theory of Speech Production. Mouton.

Fletcher, N. H., & Rossing, T. D. (1998). The Physics of Musical Instruments (2nd ed.). Springer. [Chapter 7: Whistling & edge-tones]

Ferrero, F. E., Croatto, L., & Accordi, M. (1980). Descrizione elettroacustica di alcuni tipi di vocalizzo di Demetrio Stratos. Rivista Italiana di Acustica, 4(3), 229-258.

Fu, M., Barlaz, M. S., Holtrop, J. L., Perry, J. L., Kuehn, D. P., Shosted, R. K., ... & Sutton, B. P. (2017). 3D dynamic MRI of the vocal tract during natural speech. Magnetic Resonance in Medicine, 77(4), 1590-1599.

Howell, I. (2019). Parsing the Spectral Envelope: Toward a General Theory of Vocal Tone Color. Doctoral dissertation, New England Conservatory.

Hefele, A.-M., Eklund, R., & McAllister, A. (2019). Polyphonic overtone singing: An acoustic and physiological (MRI) analysis and a first-person description of a unique mode of singing. Proceedings of FONETIK 2019, Stockholm. doi:10.5281/zenodo.3246011

Kim, Y. C., Narayanan, S. S., & Nayak, K. S. (2012). Improved imaging of lingual articulation using real-time multislice MRI. Journal of Magnetic Resonance Imaging, 35(4), 943-948.

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. Cerebral Cortex, 20(5), 1144-1152.

Klingholz, F. (1993). Overtone singing: Productive mechanisms and acoustic data. Journal of Voice, 7(2), 118-122.

Kob, M. (2004). Analysis and modelling of overtone singing in the sygyt style. Applied Acoustics, 65(12), 1249-1259.

Levin, T. C., & Edgerton, M. E. (1999). The throat singers of Tuva. Scientific American, 281(3), 80-87.

Saus, W. (2009). Karlheinz Stockhausen's STIMMUNG and vowel overtone singing. Universitas Carolina Pragensis, 471-478.

Saus, W. (n.d.). MRI of overtone singing. oberton.org. Retrieved 2025. [Describes sublingual space: "frenulum covers lateral space"]

Saus, W. (2017). Obertongesang: Das Phänomen der Obertöne in Gesang und Klang. Traumzeit Verlag. [Describes 'overtone hearing' as perceptual skill]

Stevens, K. N. (1998). Acoustic Phonetics. MIT Press.

Story, B. H. (2002). An overview of the physiology, physics and modeling of the sound source for vowels. Acoustical Science and Technology, 23(4), 195-206.

Stratos, D. (1978). Cantare la voce. Cramps Records. [Demonstrates flautofonia: flute-like multiphonics via supraglottic resonance without vocal fold vibration]

Sundberg, J. (1987). The Science of the Singing Voice. Northern Illinois University Press.

Sundberg, J., Lindblom, B., & Hefele, A. M. (2021). Voice source, formant frequencies and vocal tract shape in overtone singing: A case study. Logopedics Phoniatrics Vocology, 1-13.

Szalay, T., & Proctor, M. (2025). Articulatory and acoustic characteristics of lateral consonants: Asymmetric channel effects on anti-resonance generation. Journal of Phonetics, 108, 101-118.

Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., & Honda, K. (2006). Acoustic roles of the laryngeal cavity in vocal tract resonance. Journal of the Acoustical Society of America, 120(4), 2228-2238.

Titze, I. R. (1994). Principles of Voice Production. Prentice Hall.

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: Theory. Journal of the Acoustical Society of America, 123(5), 2733-2749.

van Tongeren, M. (2002). Overtone Singing: Physics and Metaphysics of Harmonics in East and West. Fusica.

Xu, M. B., Selamet, A., & Kim, H. (2010). Dual Helmholtz resonator. Applied Acoustics, 71(9), 822-829.

Zarate, J. M., & Bhagat, N. (2010). Neural mechanisms of pitch and error correction in singers. Annals of the New York Academy of Sciences, 1169, 270-274.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched'

American English /r/. Journal of the Acoustical Society of America, 123(6), 4466-4481. [Demonstrates asymmetric channel effects on anti-resonances]