JASA ARTICLE

Dynamic 3D MRI of vocal fold oscillations: *In vivo* assessment of vocal fold thickness, contact area, and glottal area waveform across phonation types in comparison with high-speed imaging

Louisa Traser,[1,2,a)] Fiona Stritt,[1,2] Paula L. Jordan,[2,3] Marie Köberlein,[4,5] Jonas Kirsch,[4] Stefanie Rummel,[6] Zhaoyan Zhang,[7] Bernhard Richter,[1,2] Michael Bock,[2,3] Johannes Fischer,[2,3] and Matthias Echternach[4]

[1]*Freiburg Institute for Musicians' Medicine, University of Music Freiburg, Medical Center–University of Freiburg, Freiburg Centre for Music Research and Teaching, Freiburg, Germany*

[2]*Faculty of Medicine – University of Freiburg, Freiburg, Germany*

[3]*Division of Medical Physics, Department of Radiology, Medical Center, University of Freiburg, Freiburg, Germany*

[4]*Department of Otorhinolaryngology, Ludwig-Maximilians-Universität München, Division of Phoniatrics and Pediatric Audiology, Ludwig-Maximilian-University (LMU) Munich Klinikum, Munich, Germany*

[5]*Antonio Salieri Department of Vocal Studies and Vocal Research in Music Education, University of Music and Performing Arts Vienna, Vienna, Austria*

[6]*Institut Rummel, Frankfurt, Germany*

[7]*Department of Head and Neck Surgery, University of California, Los Angeles, Los Angeles 90095, USA*

ABSTRACT:

Conventional endoscopic imaging of vocal fold (VF) vibration provides only a two-dimensional superior view, missing essential vertical dynamics such as VF thickness and vertical displacement. Simulation studies have assigned these parameters important regulatory functions, underscoring the need for *in vivo* methods capturing the three-dimensional (3D) VF geometry and vibratory motion. In this *in vivo* case study, dynamic 3D VF MRI was applied in a professionally trained singer, achieving sub-millimeter spatial and sub-millisecond temporal resolution. Six phonation types were produced according to Estill Voice Training® terminology, each intended to elicit different VF thicknesses and supraglottic configurations. For each type, ten phase-binned 3D datasets of the larynx were reconstructed per oscillatory cycle. Segmentation yielded VF thickness, vertical and horizontal displacement, contact-area dynamics, glottal area waveforms, open quotient (OQ), and supraglottic/subglottic dimensions. These MRI-derived measures showed strong correspondence with those from high-speed imaging and electroglottography acquired from the same subject, indicating that dynamic VF MRI enables reliable *in vivo* quantification of vibratory and structural parameters. Each phonation type was characterized by distinct VF geometry, supraglottic shaping, and oscillatory behavior. Across types, OQ correlated closely with VF thickness, which systematically covaried with supraglottic adjustments. This suggests supraglottic posturing represents an additional dimension of control.

 https://doi.org/10.1121/10.0043582

(Received 13 November 2025; revised 16 February 2026; accepted 28 March 2026; published online 18 May 2026)

[Editor: James F. Lynch] Pages: 4325–4340

I. INTRODUCTION

Voice research and the clinical assessment of patients with voice disorders have historically relied on endoscopic imaging of vocal fold (VF) vibration. Although the rapid oscillation of the VFs can be displayed using high-speed video imaging (HSV) or stroboscopy, this endoscopic approach inherently provides only a two-dimensional superior view of what is, in reality, a complex three-dimensional (3D) movement. Because of the alternatingly convergent-divergent shape change of the glottal channel during opening and closing phases, HSV captures the medial surface dynamics only during the closing phase but not in the opening phase where the medial surface remains largely concealed (Berry *et al.*, 2001; Döllinger *et al.*, 2005). Furthermore, vertical displacement of the superior portions of the VFs—occurring at a magnitude comparable to mediolateral motion (Semmler *et al.*, 2018)—remains inaccessible to endoscopic perspective, thus limiting comprehensive assessment of the 3D vibratory dynamics. These limitations are compounded by the unknown and continuously varying distance between the endoscope and the dynamically

[a)]Email: louisa.traser@uniklinik-freiburg.de

moving larynx (Köberlein *et al.*, 2024), introducing additional variability into quantitative measurements.

Thus, research increasingly included computational and physical simulations (e.g., Alipour and Scherer, 2000; Lehoux and Zhang, 2024; Zhang, 2025) and *ex vivo* studies using excised human or animal hemi-larynges (e.g., Berry *et al.*, 2001; Khosla *et al.*, 2014; Zhang and Chhetri, 2019), which allows analysis of the 3D vibration pattern of the VF medial surface. These studies highlighted the importance of 3D VF dynamics in voice production. In particular, the vertical thickness of the VF medial surface and the configuration of the glottal channel have been shown to critically influence glottal closure patterns, aerodynamic efficiency, and the spectral shape of the produced voice, as summarized in Zhang (2025). It has been shown that thicker VFs, for example, are associated with prolonged glottal closure and enhanced vertical phase differences and can sustain closure under higher subglottal pressures, highlighting functional relevance of medial surface geometry.

Despite its importance, *in vivo* quantification of vertical VF geometry during phonation has remained a challenge and is thus rare (Semmler *et al.*, 2018) as optical methods are restricted to surface observations and lack the ability to capture the 3D characteristics of medial surface dynamics. As a result, the relationship between VF thickness and the produced vibratory patterns, particularly the glottal closure pattern as observed in recent simulations (Zhang, 2016b, 2025), has not yet been directly verified in human subjects. Similarly, although it is known that voice production often involves simultaneous adjustments in the supraglottal configurations (Švec and Zhang, 2025; Bailly *et al.* 2014; Aaen *et al.* 2020; Traser *et al.* 2025), the interaction between supraglottal configuration and VF geometry during the production of different phonation types has not been investigated in humans. Such interactions are of clinical importance as there are indications of functional differences between medio–lateral and anteroposterior supraglottal constrictions (Mayerhoff *et al.*, 2014; Shembel *et al.*, 2025).

The goal of this study is to investigate the relationship between VF vertical thickness, supraglottic shaping, and the resulting VF oscillatory behavior in a human subject. Magnetic resonance imaging (MRI) offers a promising non-invasive approach to address this gap. Recent advances have enabled dynamic *in vivo* imaging of VF oscillations with sufficient spatial and temporal resolution, first in two dimensions (coronal plane) and more recently in three dimensions (Fischer *et al.*, 2020; Fischer *et al.*, 2022; Fischer *et al.*, 2026). These developments allow volumetric assessment of VF motion and adjacent airway structures during phonation.

In the present study, we apply zero echo time (ZTE)–based dynamic 3D MRI (Fischer *et al.*, 2026) to a professionally trained female singer to investigate how different voice production types are associated with changes in vertical VF geometry and supraglottic configuration and how such physiological differences are correlated with differences in the produced VF vibration and glottal closure pattern. In particular, we also address the physiological question of whether distinct supraglottic shaping strategies are accompanied by systematic differences in vertical VF thickness. To contextualize the MRI-derived measures within established assessment techniques, transnasal HSV and electroglottography (EGG) were acquired from the same subject, enabling comparison of oscillatory parameters and identification of potential modality-specific differences.

By focusing on the interplay between supraglottic shaping and vertical VF dynamics in vocal control, this work aims to advance the *in vivo* understanding of voice production mechanisms and to establish dynamic 3D MRI as a complementary research tool for investigating complex phonatory behaviors, with potential relevance for future clinical and pedagogical applications.

II. METHOD

A. Subject and task

This study presents findings from 3D MRI measurements of VF motion during phonation using various voice production mechanisms, performed by a professionally trained singer (co-author S.R.). The figures framework of Estill Voice Training® was used to describe the voice production mechanisms (Steinhauer *et al.*, 2017). Here, "*Thick*" and "*Thin*" refer to the intended thickness of the medial surface of the VF. Thick has so far been associated with a vibratory pattern predominantly produced by thyroarytenoid (TA) muscle activation (M1), whereas Thin has been linked to predominantly cricothyroid (CT) muscle activation (M2) (Steinhauer *et al.*, 2017). *Stiff* refers to a mode where the VFs should be stretched and slightly abduced, leading to a posterior gap between the VFs with different grades of breathy phonation. *Thick & Ventricular Folds (VenF) retract* and *Thick & VenF constrict* combines thick VFs with the ab-/adduction of the VenF. *Thick & aryepiglottic sphincter (AES) narrow* aims additional narrowing of the AES to form a ring-like constriction involving the epiglottis, the aryepiglottic folds, and the arytenoid cartilages with thick VF and with the VenF in a mid position. All phonation types were produced using sustained phonation on the vowel [i], at the fundamental frequency (f_o) = 196 Hz and at the loudness level most typically associated with each type. Only phonatory task productions that were independently evaluated as correct with respect to f_o, loudness, and phonation mechanism by the participant and by two voice experts (co-authors L.T. and F.S.) were included in the analysis.

B. MRI data acquisition and reconstruction

MRI data were acquired on a clinical 3T MRI scanner (PrismaFit, Siemens Healthineers, Erlangen, Germany) using a zero echo time (ZTE) pulse sequence introduced by Fischer *et al.* (2026) A circular radio frequency coil was positioned on the skin of the volunteer at the position of the larynx to acquire the MR signal, achieving an image resolution of 0.8 mm over a scan time of 5 min 20 s. During the MRI acquisition, the singer phonated at a constant f_o with intermittent breathing at will. For tracking shifts in laryngeal

https://doi.org/10.1121/10.0043582

position between phonations, simultaneously acquired MR-navigator signals were used. For motion correction prior to the reconstruction, each navigator is compared to a reference navigator.

For synchronization of VF motion and MR data, acoustic signals were recorded *via* a microphone (MR Confon, Magdeburg, Germany) placed at the volunteer's mouth and processed with a low-pass Butterworth filter to minimize MRI noise interference and extract f_o. For the reconstruction, the VF motion phase during signal acquisition was extracted from the acoustic signal, and MR data were binned into 10 time frames based on the motion phase. This results in one VF vibration depicted by the 10 distinct phases of the oscillatory cycle, in which frame 1 always represents the maximum glottal opening. Image size was upscaled by a factor of 2 using zerofilling, yielding a pixel size of 0.4 mm. Full 3D datasets of 10 time frames in 6 phonation types are included in the supplementary data (Traser *et al.*, 2026) (see phonation type MRI datasets).

To assess temporal stability of the acoustic signal under MRI conditions across the full acquisition duration, each phonation interval was segmented into consecutive analysis windows of 2–4 s. Acoustic measures were extracted using PRAAT (version 6.4.52), including jitter (Relative Average Perturbation, RAP), shimmer (Amplitude Perturbation Quotient over 3 periods, APQ3), and the spectral measure H1–H2 (difference in amplitude between the first and second harmonics). For visualization, these parameters were plotted alongside the corresponding spectrograms (window size: 32 768 samples). All analyses were performed separately for each phonation type. The resulting time-resolved acoustic measures are provided in the supplementary data (Traser *et al.*, 2026) (voice analysis MRI plots).

C. MRI segmentation and measures

Air–tissue boundaries were segmented in each of the ten frames of the resulting image data, and geometric shape parameters were quantified using threshold-based pixel analysis. For each phonation type, 10 complete 3D surface reconstructions were segmented [all 60 STL datasets are included in the supplementary data (Traser *et al.*, 2026)]. representing the temporal course of the oscillatory cycle. Geometric parameters and subvolumes are defined in Fig. 1.

1. Reference lines

(A) In the mid-sagittal plane at the instant of maximum VF contact, reference line A was placed along the inner surface of the thyroid cartilage and extended superiorly beyond the region of VF attachment [Fig. 1(I)].
(B) In the mid-coronal plane at the instant of maximum VF contact reference line B was placed horizontally at the cranial VF edge when the VFs are at their lowest vertical position [Fig. 1(II)].
(C) In the same mid-coronal plane at the instant of the smallest vertical extent of the closed VF, a contact line was marked vertically [Fig. 1(II)]. When no VF contact was present, two vertical lines adjacent to the maximum medial extent of the VFs were used.
(D) In the same mid-coronal plane during maximum VF opening, a horizontal line passing through the superior edge of the VF at maximum opening [Fig. 1(III)].

2. Geometric measures

The following geometric measurements were performed on the 2× upsampled grayscale image data using 3D SLICER:

(1) VF length (VFl) was defined in mid-sagittal plane as the distance from reference line A at the level of the inferior VF edge to the air–tissue boundary at the posterior end of the VF/arytenoid cartilage complex [Fig. 1(I)] obtained during maximum VF contact. It should be noted that this measure does not represent the length of the oscillating portion of the VF as the transition between the membranous VF and the vocal process could not be reliably differentiated based on the available image data.
(2) VF thickness (VFth): The thickness of right VF thickness (VFthr) and left VF thickness (VFthl) was defined according to (Lehoux and Zhang, 2024) as the vertical extent of the VF measured 0.5 mm lateral to the medial VF edge or maximum VF contact as defined by refence line C [Fig. 1(II)] during maximum VF contact. The mean of VFthr and VFthl was then calculated as VFth.
(3) Vertical VF displacement (vVFD) was defined as the vertical distance between reference lines B and D.
(4) Maximum glottal opening (MGO) was defined as the maximal medial distance between the VFs in the transverse plane at maximum glottal open phase [Fig. 1(III)].
(5) Horizontal VF displacement (hVFD) was calculated as MGO divided by two.
(6) Ventricular fold distance (VenFD) was defined as the medial distance between the left and right VenF edge during maximum glottal open phase.
(7) Laryngeal ventricle diameter (LVD) was defined as the maximum lateral dimension of the laryngeal ventricle during maximum glottal open phase.
(8) Subglottic diameter (SGD) was defined as the maximal width at the inferior margin of the caudal attachment point of the VFs,
(9) *Aditus laryngis*–coronal (ALC) was defined as the maximal coronal width of the *aditus laryngis* during maximum glottal open phase.
(10) *Aditus laryngis*–sagittal (ALS) was defined as the maximal anteroposterior depth of the *aditus laryngis* during maximum glottal open phase.

3. Volumetric subdivisions

The subdivision of laryngeal volumes is shown in Figs. 1(VI) and 1(VII) and includes the following:

(1) Laryngeal vestibule: The 3D volume of the laryngeal vestibule was defined as the space extending from the

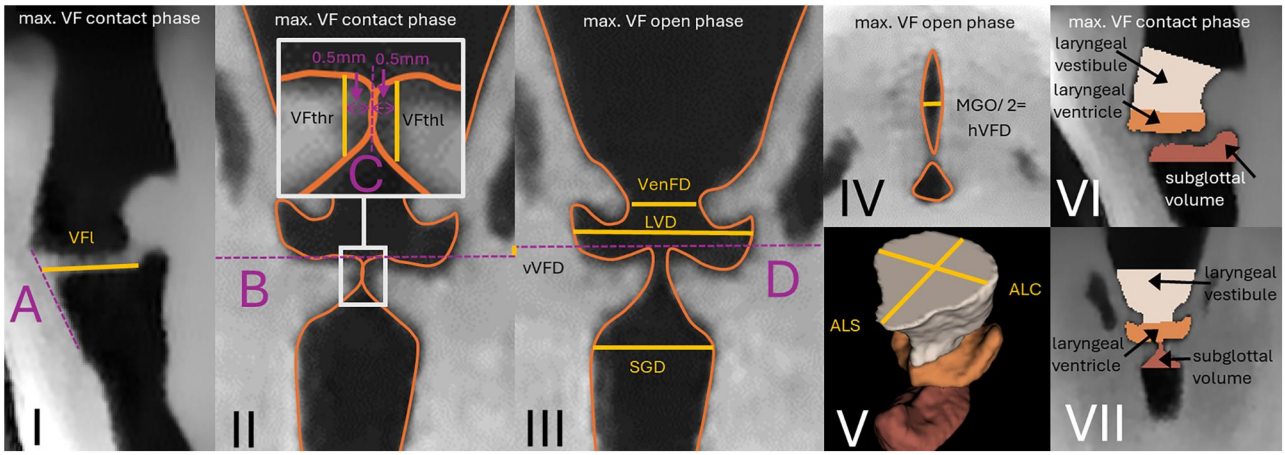


FIG. 1. Definition of reference lines and geometric parameters. (I) Mid-sagittal plane at maximum VF contact: VFl, defined as the distance from reference line A at the inferior VF edge to the air–tissue boundary at the posterior VF end. (II) Mid-coronal plane at maximum VF contact: reference line B (smallest vertical VF extent), reference line C (vertical extent of closed VF), and VF thickness (VFthr, VFthl), measured 0.5 mm lateral to the medial VF edge or maximum VF contact; the mean value was calculated as VFth. (III) Mid-coronal plane at maximum VF open phase: reference line D (greatest vertical VF extent), VenFD, LVD, vVFD (vertical distance between lines B and C), and SGD. (IV) Transverse plane at maximum VF open phase: MGO, hVFD = MGO/2. (V) Three-dimensional renderings of the *aditus laryngis* entrance showing ALC and ALS dimensions. (VI) Subdivision of volumes in the mid-sagittal plane at the instant of maximum VF contact. (VII) Subdivision of volumes in the mid-coronal plane at the instant of maximum VF contact

entrance tangential to the aryepiglottic fold to a horizontal plane intersecting the medialmost point of the VenF during maximum VF contact.

(2) Laryngeal ventricle: The laryngeal ventricle was defined cranially by a horizontal plane intersecting the medialmost point of the VenF and caudally by a horizontal plane intersecting the medialmost point of the VF during maximum VF contact.

(3) Subglottal space: The subglottic space was defined as the volume extending from the horizontal plane intersecting the medialmost point of the VF to the horizontal plane at their inferior border.

Furthermore, the VF contact area was quantified in the midsagittal plane. A region of interest encompassing the contact region was manually defined, and the contact area in each frame was then segmented using grayscale thresholding. In the next step, all pixels that were present in every frame (i.e., the static component) were subtracted from the total segmented area to obtain the dynamic component of the VF contact area.

Analogous to the HSV-derived glottal area waveform (GAW), an MRI-based GAW was extracted from the VF MRI data. This was computed by summing all non-tissue-occupied pixels within the 3D region of the glottis during all ten time frames using a threshold-based approach. In a subsequent step, only those pixels were retained for which the glottal opening remained completely unobstructed by the VFs across all depth planes (see Fig. 2), analogous to the view of an endoscopic camera positioned at infinite distance. This approach was used to derive a GAW that approximates the analysis of a two-dimensionally projected camera image.

Whereas conventional GAWs are subject to variation in viewing angle and distance (Veltrup *et al.*, 2023), the MRI-derived GAW enables *in vivo* quantification of the glottal area from a superior view during VF oscillation. From this GAW, an MRI-based OQ was subsequently calculated in analogy to the HSV signal by normalizing the amplitude and then using a tolerance threshold of 10%. Thus, all MRI GAW values >10% from the normalized baseline (i.e., GAW area of the fully open glottis) were considered to indicate an open glottis, whereas the condition GAW ≤ 10% was used as an indication for a closed glottis.

D. High speed imaging and EGG

In a separated session, transnasal endoscopy of the larynx was performed on the same subject, phonating the same voice production mechanisms using HSV at 20 000 fps with an image resolution of 386 × 320 pixels as described by (Echternach *et al.*, 2017). To promote task consistency across HSV and MRI sessions, identical task instructions were used, target pitch was specified before each task, and phonation was based on well-defined laryngeal configurations from Estill Voice Training®. The participant's certification as an Estill Mentor and Course Instructor, representing the highest level of proficiency within the Estill Voice Training system, supports a high reliability in reproducing these configurations across measurement sessions.

Acoustic output (DPA Microphones, model 4061, Kokkedal, Denmark) and EGG signals (EG2-PCX2, Glottal Enterprises, Syracuse, NY) were recorded simultaneously. From the audio signal, the sound pressure levels (SPLs), indicated in in A-weighted decibels [dB(A)], was calculated after a calibration with a SPL meter (DSL-331, Tecpel, Taipei, Taiwan) using SOPRAN software (Svante Granqvist, Karolinska, Stockholm, Sweden). For further analysis, the GAW was obtained following the procedure described by Echternach *et al.* (2017). First, the honeycomb pattern introduced by the optics of the flexible endoscope was suppressed using a frequency-selective filter in the Fourier domain. The image sequences were then rotated to align the glottal

https://doi.org/10.1121/10.0043582

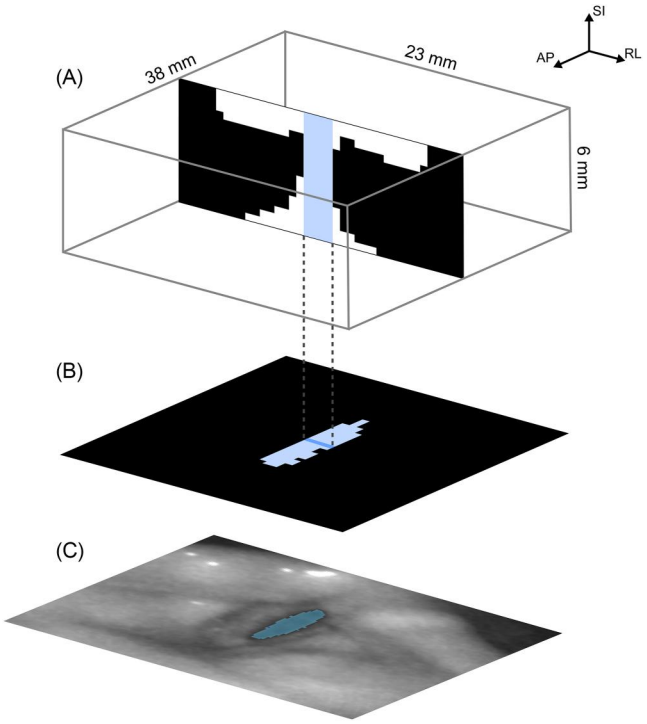


FIG. 2. Differences in the definition of the GAW derived from MRI versus HSV. (A) Based on three-dimensional VF MRI data, a region of interest centered on the VFs was extracted for each of the ten time frames. Intensity thresholding segmented the volume into air voxels (white) and tissue voxels (black). (B) For each time frame, all voxel columns along the superior–inferior direction containing exclusively air voxels were identified (blue region). The number of such columns multiplied by the voxel area in the anteroposterior and right–left plane defined the GAW for that frame. (C) For comparison, the imaging field of view of HSV is shown together with the segmented GAW area (blue region), illustrating differences in definition between MRI- and HSV-based GAW estimation.

midline vertically with respect to the image frame and cropped to a region of interest containing the vibrating VFs. Subsequently, glottal segmentation—i.e., semi-automatic extraction of the time-varying medio–lateral VF deflections—was performed using the Glottis Analysis Tools software (Inwald *et al.*, 2011). This is performed by using a semi-automatic, seed-based region-growing approach, where the dark glottal area is segmented based on user-defined gray-scale intensity thresholds. The resulting binary glottal mask is then used to compute the GAW by integrating the segmented area over time along a defined midline.

The GAW was then normalized in amplitude. The GAW's closed phase between two peaks was defined using a threshold value. That threshold was chosen for each closed phase as 10% of the greater of the two surrounding GAW peaks.

The closing quotient is defined from the glottal area waveform as the ratio of the closing phase duration—i.e., from maximum glottal opening to complete closure—to the total cycle period (Patel *et al.*, 2014).

E. Acoustic analysis of HSV and MRI audio

Because of the inherent noise of the HSV camera, a relevant noise floor is present in the HSV audio recording. For the analysis of the audio captured during the MRI session, the simultaneously recorded signal from an optical, MRI-compatible microphone (FOMRI I, MR Confon, Magdeburg, Germany) was used. Representative audio samples are included in the supplementary data (Traser *et al.*, 2026) (audio recordings). A short segment preceding the onset of the actual MRI scanning sequence was selected for analysis as it contains the general background noise generated by the machine's helium pump but is not affected by the acoustic output of the MRI system during data acquisition.

From both audio recordings of the 6 phonation types, long-term average spectral measures (LTAS) were created with a window length of 50 ms and 45 ms overlap (= 5 ms hop size). To avoid edge effects at the window borders, a Hamming window was applied to each segment. Finally, the LTAS's amplitude was normalized to the range [0,1] between 100 and 5000 Hz.

On each of those sections, a spectrum was calculated with a FFT length of 2p where $p = \lceil \log2\ N \rceil$ (the $\lceil\ \rceil$ symbols denote the ceiling function, N is the number of samples in the window). Finally, the spectrum's amplitude values were converted to dB. The spectrograms were generated for 250 ms of the MRI and the HSV audio recordings.

III. RESULTS

A. Stability of voice production during MRI recording

To assess the stability of voice production during the prolonged MRI acquisition (>5 min), time-resolved acoustic perturbation measures were analyzed together with the relation between the amplitudes of the first two harmonics (H1–H2) and the corresponding spectrograms [see supplementary data (Traser *et al.*, 2026), voice analysis MRI plots].

Frequency perturbation was lowest for *Thick* and *Thick & AES narrow* phonation, with jitter values predominantly below 1%, whereas *Thin*, *Stiff*, and *Thick & VenF constrict* exhibited higher jitter with increased variability, typically ranging between 2% and 5%. In contrast, amplitude perturbation (shimmer) was lowest for Thin and Stiff phonation (<3%) and highest for *Thick* and *Thick & AES narrow* (approximately 3–10%), with intermediate values observed for the remaining phonation types.

The H1–H2 ratio was largest for *Thin*, *Stiff*, and *Thick & VenF retract* (approximately 10–20 dB), smaller for *Thick & VenF constrict* (approximately 0–10 dB), and lowest for *Thick* and *Thick & AES narrow* (approximately −10 to 5 dB). Across phonation types, both the absolute values and the dispersion of the acoustic measures remained largely stable over time. One notable exception was *Thick* phonation, for which an increase in jitter accompanied by a concurrent increase in H1–H2 was observed approximately halfway through the recording.

B. Spectral differences across phonation types in MRI and HSV

Representative 250 ms segments of sustained phonation for each of the six phonation types were analyzed in both HSV and MRI recordings [see Fig. 3 and supplementary data

(Traser *et al.*, 2026), audio recordings]. Spectrograms revealed consistent spectral patterns across modalities. *Stiff* phonation showed low spectral energy and a pronounced noise floor, whereas *Thin* exhibited clear harmonic structure but weak high-frequency amplitude (>2 kHz). *Thick* and *Thick & AES narrow* displayed stronger acoustic energy, with *Thick & AES narrow* showing the most prominent harmonic definition and spectral energy between 2.5 and 5 kHz. *Thick & VenF retract* and *Thick & VenF constrict* presented reduced spectral energy in this region, however, greater than in the *Thin* and *Stiff* conditions.

C. EGG and GAW analysis of OQ, SPL, and glottal dynamics across phonation types

OQ values derived from EGG and GAW showed consistent agreement across methods with distinct differences between phonation types (see Fig. 4).

Stiff phonation yielded the highest OQ, reflecting a prolonged open phase, whereas *Thick & VenF constrict* displayed the lowest values with short glottal opening. Intermediate OQ values were observed in descending order for *Thin*, *Thick & VenF retract*, *Thick*, and *Thick & AES narrow*. A similar distribution was found for the closing quotient. SPL was greatest in *Thick & AES narrow*, slightly reduced in *Thick*, and about 10 dB lower across the remaining conditions.

Visual analysis of HSV with simultaneous EGG and GAW displays confirmed these parameter-based findings [see the supplementary data (Traser *et al.*, 2026), video S1_HSV_EGG_GAW.mp4]. *Thick* phonation showed full-amplitude oscillations, complete closure, and strong mucosal wave propagation, corresponding to narrow GAW peaks and broad EGG peaks. *Thin* displayed reduced oscillatory amplitude but complete closure. *Stiff* exhibited closure restricted to the anterior third and a sinusoidal GAW contour. *Thick & AES narrow* demonstrated pronounced anteroposterior supraglottic narrowing and convex VenF shaping, accompanied by large oscillation amplitude and mucosal wave displacement. *Thick & VenF constrict* showed stronger medio–lateral narrowing of the VenF and reduced oscillation amplitude, whereas *Thick & VenF retract* resembled *Thin* supraglottically but produced broader EGG peaks similar to *Thick*.

Exemplary GAW curves derived from HSV (Fig. 5) further illustrate these systematic differences. *Stiff* produced the largest oscillatory amplitudes without reaching complete closure. Thick showed greater amplitude than *Thin*, whereas *Thick & AES narrow* and *Thick & VenF retract* yielded intermediate values, and *Thick & VenF constrict* exhibited the smallest excursions. The temporal profiles highlight phonation-specific distinctions in open–close symmetry and peak amplitude, consistent with the OQ, closing quotient, and SPL measures.

D. Geometric and volumetric analysis of 3D VF MRI segmentations

Geometric parameters of the supraglottic and subglottic compartments differed systematically across phonation types (see Fig. 6).

VenFD was largest in *Stiff* and *Thin*, followed by *Thick & VenF retract*. During *Thick & AES narrow* VenFD was slightly larger compared to *Thick*, whereas *Thick & VenF constrict* showed by far the smallest distance. The LVD was widest in *Thick & VenF retract*, slightly smaller in *Thin* and *Stiff*, and progressively reduced in *Thick*, *Thick & AES narrow*, and most constricted in *Thick & VenF constrict*. Subglottal diameters were also largest in *Thin* and *Thick & VenF retract*, slightly smaller in *Stiff* and *Thick*, and narrowest in *Thick & AES narrow* and *Thick & VenF constrict*. The *aditus laryngis* showed patterns comparable to the VenFD, with the smallest sagittal (anteroposterior) dimension observed in the *Thick & AES narrow* condition and the smallest coronal (medio-lateral) dimension in the *Thick & VenF constrict* condition. Volumetric analysis confirmed these patterns (see Figs. 7 and 8).

The largest overall laryngeal volumes were observed in *Thin* and *Thick & VenF retract*, both showing expanded subglottal, ventricular, and epilaryngeal compartments. *Stiff* exhibited a slightly smaller total volume. *Thick* occupied an intermediate range, whereas *Thick & AES narrow* showed a clear reduction in all compartments. The smallest overall volume occurred in *Thick & VenF constrict*, reflecting markedly reduced ventricular volume and *aditus laryngis*. Across sub-compartments, the laryngeal ventricle was particularly expanded in *Thin*, *Stiff*, and *Thick & VenF retract*, markly reduced in *Thick* and *Thick & AES narrow* and smallest in *Thick & VenF constrict*.

The laryngeal vestibule followed a similar trend, whereas the subglottal space was largest in *Thin* and *Thick & VenF retract*, but followed by *Thick*, *Stiff*, *Thick & VenF constrict* and smaller in *Thick & AES narrow*.

Three-dimensional reconstructions (see Fig. 8) confirmed these findings: *Thin, Stiff* and *Thick & VenF retract* were characterized by cranial pouching (laryngeal saccule) and lateral expansion of the ventricle, *Thick* and *Thick & AES narrow* showed compact morphologies, and *Thick & VenF constrict* exhibited the most compact configuration.

E. Dynamic and geometric VF parameters derived from 3D MRI

Dynamic and geometric parameters varied systematically across phonation types as shown in Fig. 9 [supplementary data (Traser *et al.*, 2026), video S2 dynamic 3D surface reconstruction.mp4 and video S3_slices animation.gif]. VF cross-sectional shape was predominantly rectangular in *Thick*, *Thick & AES narrow*, and *Thick & VenF constrict*, associated with pronounced thickness and stronger inferior bulging of the medial surface. In contrast, *Thin* and *Stiff* exhibited a triangular configuration with slender folds and only slight inferior bulging. *Thick & VenF retract* presented intermediate shapes between rectangular and triangular, characterized by moderate VF thickness and less pronounced inferior bulging.

Thickness quantified according to Lehoux and Zhang (2024) revealed the lowest values in *Thin* and *Stiff*; whereas in *Thick & VenF retract*, values were intermediate; *Thick*

https://doi.org/10.1121/10.0043582

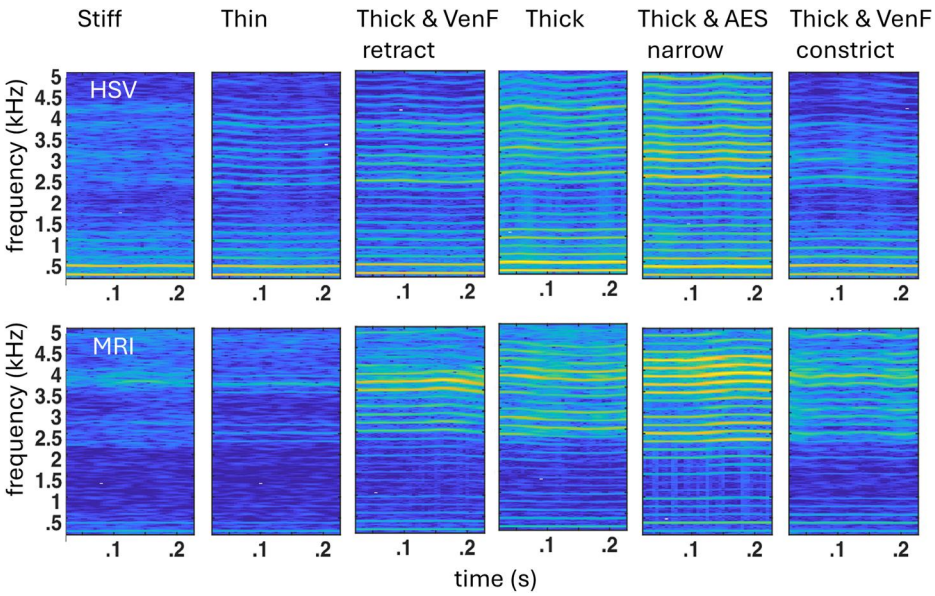
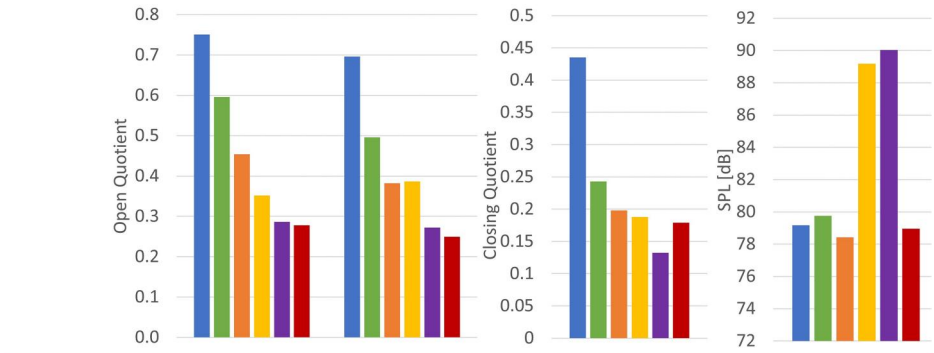


FIG. 3. Spectrograms from HSV and MRI recordings. Spectrograms (0–5 kHz) from sustained phonations across six phonation types as recorded during HSV (top) and MRI (bottom) sessions. Each segment represents 250 ms of stable phonation. Color represents spectral energy, coded from cool (low intensity) to warm (high intensity) colors.



	OQ GAW	OQ EGG	Closing Quotient GAW	SPL
Stiff	0.80	0.70	0.44	79
Thin	0.60	0.50	0.24	80
Thick & VenF retract	0.45	0.38	0.20	78
Thick	0.35	0.39	0.19	89
Thick & AES narrow	0.29	0.27	0.13	90
Thick & VenF constrict	0.28	0.25	0.18	79

FIG. 4. OQ derived from EGG, HSV-based GAW, closing quotient (from HSV GAW) and SPL across six phonation types. The phonation types are color coded in *Stiff* (blue), *Thin* (green), *Thick & VenF retract* (orange), *Thick* (yellow), *Thick & AES narrow* (violet), *Thick & VenF constrict* (red).

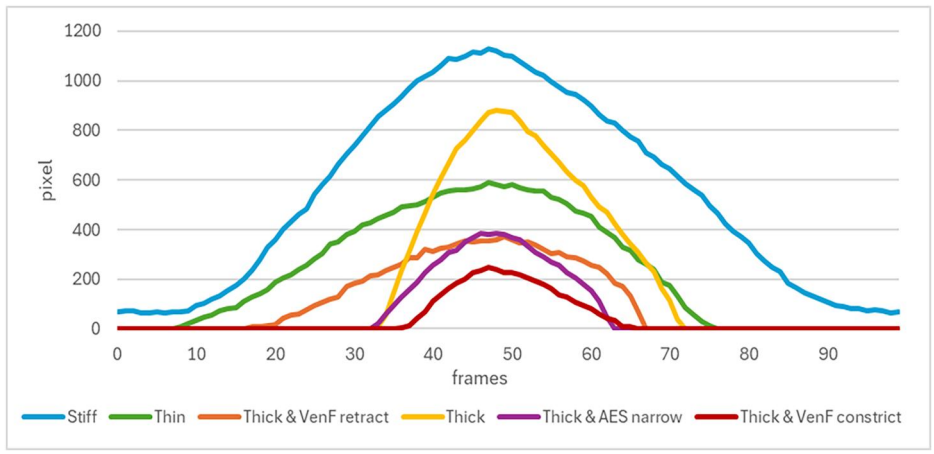
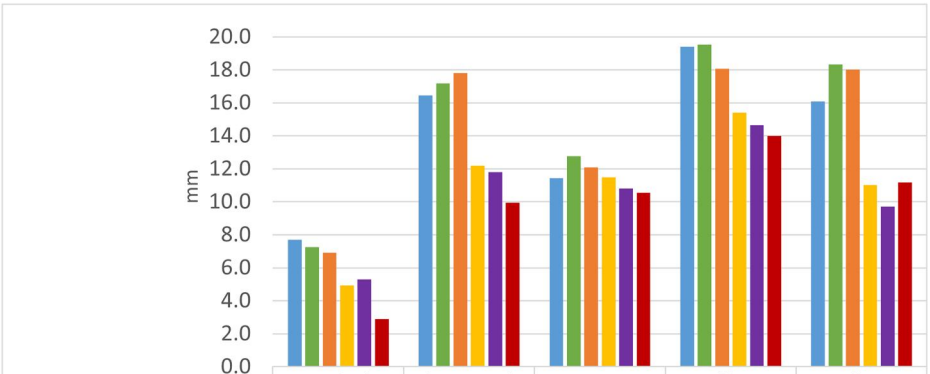


FIG. 5. Exemplary GAWs derived from high-speed imaging for six phonation types. The waveforms were aligned at their maximum values for direct comparison. The curves illustrate systematic differences in oscillatory amplitude and temporal profile across conditions. The phonation types are color coded in *Stiff* (blue), *Thin* (green), *Thick & VenF retract* (orange), *Thick* (yellow), *Thick & AES narrow* (violet), *Thick & VenF constrict* (red).



	ventricular fold distance (VenFD)	laryngeal ventricle diameter (LVD)	subglottal Diameter (SGD)	aditus laryngis coronal (ALC)	aditus laryngis saggital (ALS)
Stiff	7.7	16.5	11.4	19.4	16.1
Thin	7.3	17.2	12.8	19.5	18.3
Thick & VenF retract	6.9	17.8	12.1	18.1	18.0
Thick	4.9	12.2	11.5	15.4	11.0
Thick & AES narrow	5.3	11.8	10.8	14.7	9.7
Thick & VenF constrict	2.9	9.9	10.6	14.0	11.2

FIG. 6. Geometric measures across six phonation types. The figure shows values (in mm) for maximal diameter of LVD, maximal SGD, distance of medial VenFD edges, and the ALC and ALS. The phonation types are color coded in *Stiff* (blue), *Thin* (green), *Thick & VenF retract* (orange), *Thick* (yellow), *Thick & AES narrow* (violet), *Thick & VenF constrict* (red).

was clearly thicker; and *Thick & AES narrow* and *Thick & VenF constrict* exhibited the greatest vertical VF thickness [see Fig. 10(C)]. The VF thickness increased by a factor of 2 between the smallest and the largest condition. The VF length varied by more than 2 mm, displaying an almost mirror-inverted pattern to VF thickness: the longest VF occurred in the *Thin* condition (rather than in *Stiff*), whereas *Thick & AES narrow* and *Thick & VenF constrict* differed only marginally [see Fig. 10(B)].

The phonation types are color coded in Stiff (blue), Thin (green), Thick & VenF retract (orange), Thick (yellow), Thick & AES narrow (violet), Thick & VenF constrict (red). Displacement amplitudes complemented these observations [see Fig. 10(A)]. Vertical excursions were greatest in *Stiff* and *Thick & VenF retract*; moderate in *Thin*, *Thick*, and *Thick & AES narrow*; and lowest in *Thick & VenF constrict*, which exhibited values of only 20% off the Stiff condition. Horizontal displacements followed a similar trend, highest in *Stiff*, intermediate in *Thin* and *Thick & VenF retract*, and smallest in *Thick & VenF constrict*.

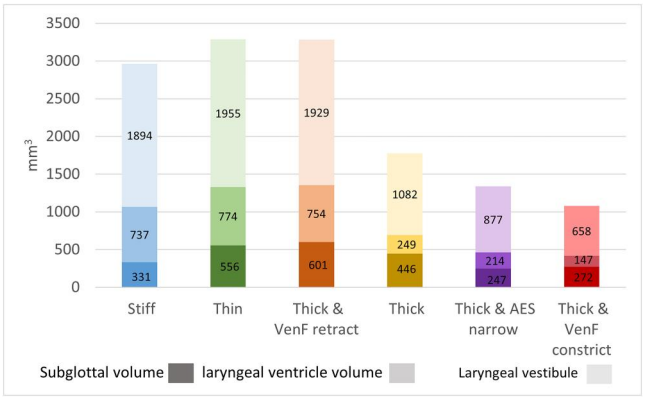


FIG. 7. Volumetric analysis of 3D VF MRI segmentations, separated into three anatomically defined sub-volumes: the subglottal space (lowest section, darkest color), the laryngeal ventricle (middle section, medium color), and the laryngeal vestibule (highest section, lightest color). The phonation types are color coded in *Stiff* (blue), *Thin* (green), *Thick & VenF retract* (orange), *Thick* (yellow), *Thick & AES narrow* (violet), *Thick & VenF constrict* (red).

The combination of long, slender VF with reduced thickness in *Thin* and *Stiff* favored less pronounced convergent–divergent glottal configurations. In contrast, shorter and thicker folds in *Thick* and *Thick & AES narrow* reinforced strong convergent–divergent behavior. The *Thick & VenF constrict* condition, representing the shortest and thickest configuration, showed markedly reduced oscillatory excursions in both vertical and horizontal dimensions. [see supplementary data (Traser *et al.*, 2026), video S3_slices animation.gif].

F. VF contact area

The temporal evolution of the VF contact area differed systematically across phonation types [Fig. 10(E) and supplementary data (Traser *et al.*, 2026), video S4_contact area.gif). For visual analysis of Fig. 10(E) it should be kept in mind, that the VF contact area is plotted in inverse orientation relative to the GAW. *Thick & AES narrow* exhibited the largest maximum contact, peaking in the middle of the oscillatory cycle with a steep rise and decline. *Thick* reached its maximum earlier, showing a pronounced peak in the first half of the cycle followed by a gradual decrease. *Thick & VenF retract* displayed a biphasic pattern with two successive peaks, both occurring later than in *Thick*. In the *Thin* condition, the excursions were reduced by approximately two-thirds compared to the other Thick variants, and both the rising and falling phases were more symmetrical. *Stiff*

https://doi.org/10.1121/10.0043582

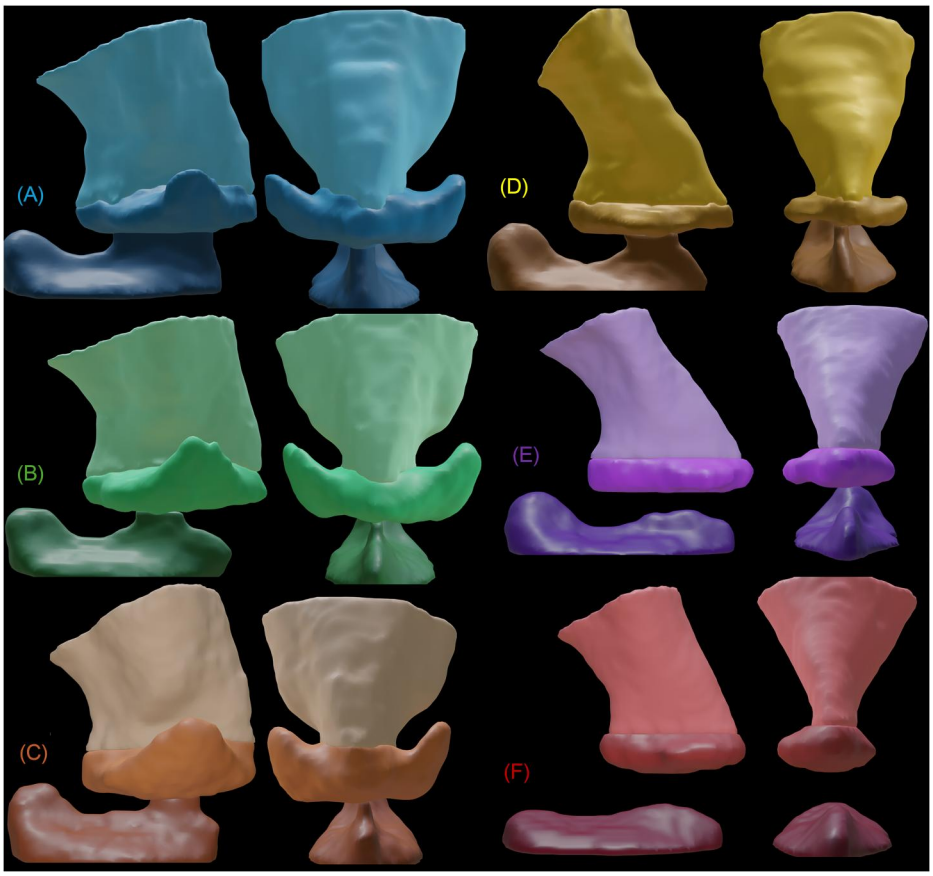


FIG. 8. Three-dimensional reconstructions of vocal tract segmentations derived from 3D VF MRI. Each model comprises three anatomically defined sub-volumes: the laryngeal vestibule (superior), the laryngeal ventricle (middle), and the subglottal space (inferior), rendered in matching hue gradations for each phonation type. Left, lateral view; right, anterior view. The phonation types are color coded in *Stiff* (blue), *Thin* (green), *Thick & VenF retract* (orange), *Thick* (yellow), *Thick & AES narrow* (violet), *Thick & VenF constrict* (red).

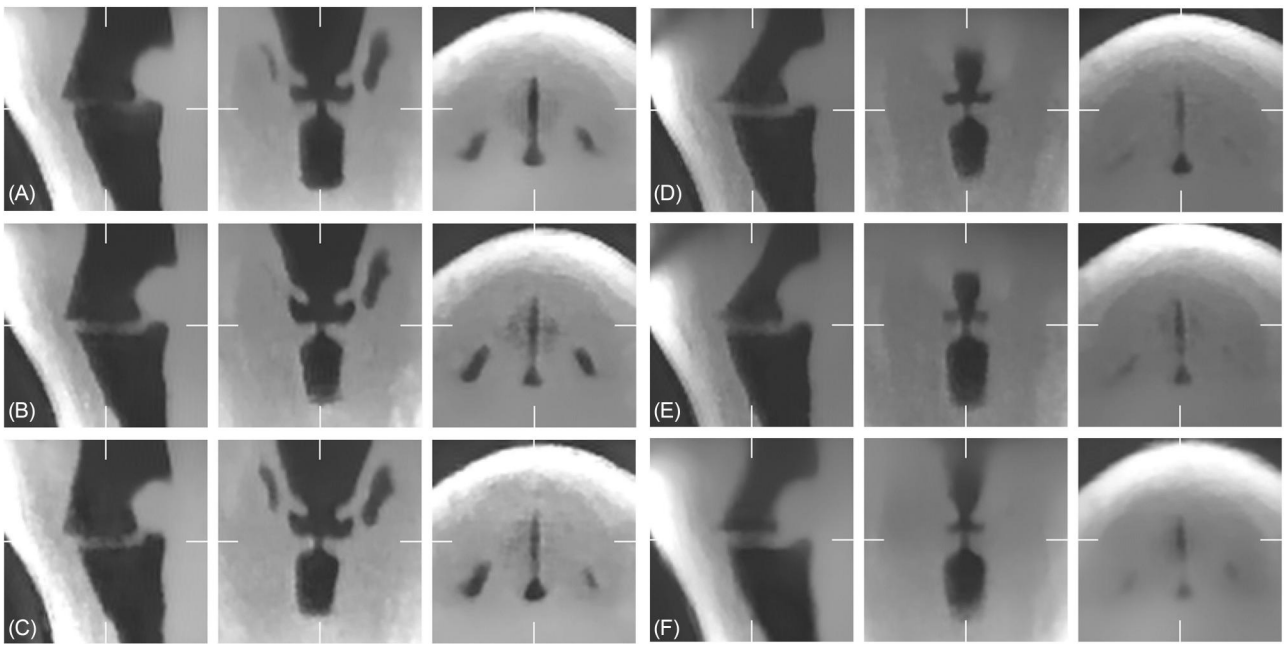


FIG. 9. Sagittal, coronal, and axial views of 3D dynamic VF MRI for six distinct phonation conditions. Rows A–F correspond to the following configurations: *Stiff* (A), *Thin* (B), *Thick & VenF retract* (C), *Thick* (D), *Thick & AES narrow* (E), *Thick & VenF constrict* (F). Each column shows one anatomical plane (left to right: sagittal, coronal, axial). The crosshairs visible in each image indicate the corresponding plane positions in the other two dimensions, providing spatial orientation across views. The image shows the maximal open glottal configuration. A dynamic version is shown in the supplementary data (Traser *et al.*, 2026), video S2_slices animation.gif.

had the lowest overall contact amplitude, reduced by factor 4, reflecting minimal closure dynamics. *Thick & VenF* constriction exhibited a plateau-like pattern, with dynamic contact levels reduced by a factor of two relative to *Thick*. These levels were sustained over most of the cycle, with only a brief opening phase.

G. MRI-based GAW and OQ

MRI-derived GAWs revealed systematic amplitude and temporal differences across phonation types [see Fig. 10(F)]. *Stiff* exhibited the largest glottal opening without complete closure and, compared to *Thick*, reduced peak-to-peak excursion. *Thin* also did not reach full closure, but the glottal area was markedly smaller than in *Stiff*. All other phonation types achieved complete closure in the GAW. Oscillatory amplitudes were slightly greater in *Thick & AES narrow*, *Thick*, and *Thick & VenF retract*, whereas the GAW maximum in *Thick & VenF constrict* showed the smallest peak-to-peak amplitude with a reduction by factor 4 from *Thick & AES narrow*.

OQ values derived from MRI-based GAW analysis (Fig. 10) showed good overall agreement with EGG- and HSV-based measures (Fig. 11), although absolute values were consistently higher compared to EGG/HSV.

Stiff, followed by *Thin*, yielded the highest OQ, reflecting a prolonged open phase, whereas *Thick & VenF constrict* produced the lowest OQ with a short open phase. Intermediate values were observed for *Thin, Thick & VenF retract, Thick*, and *Thick & AES narrow*, matching the order reported for EGG- and HSV-derived data.

IV. DISCUSSION

The present study provides, to our knowledge, the first systematic comparison of six distinct phonation types using dynamic 3D VF MRI (Fischer *et al.*, 2026) integrating morphological, volumetric, and vibratory analyses *in vivo*. MRI-derived parameters were compared with HSV and EGG, which are considered gold standards for assessing VF vibration.

A. Method comparison and acoustic stability considerations

Because simultaneous acquisition of MRI and HSV is not feasible (see Sec. IV D), acoustic recordings were used to verify consistent phonation type production across sessions. The acoustic signatures of the six investigated phonation types were reproducible across modalities (Fig. 3). Overall, MRI-based dynamic and volumetric measures corresponded well with HSV-derived metrics (see Fig. 11).

OQ values from all three methods (EGG, HSV-based GAW, MRI-based GAW) were consistently correlated with VF thickness, with OQ decreasing with increasing VF thickness. OQ also mirrored variation in maximum VF contact area, except under pronounced VenF adduction. Associations

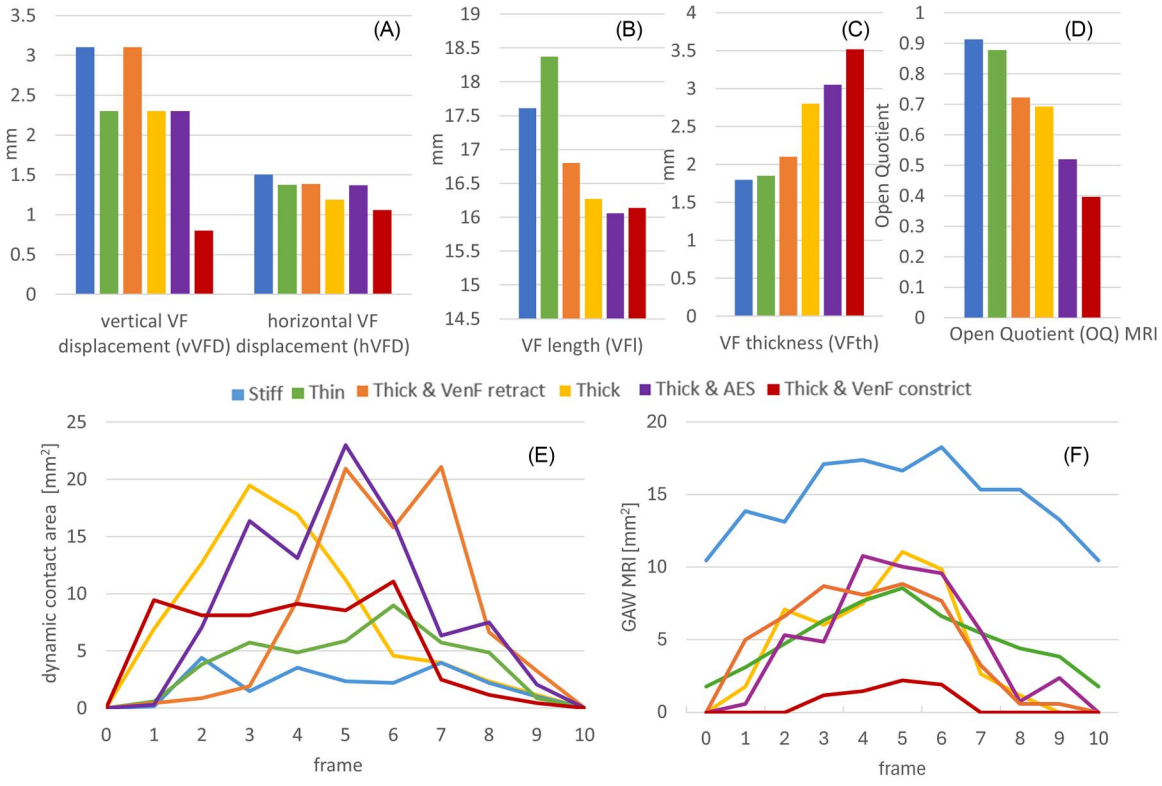


FIG. 10. VF displacement, morphology, contact dynamics, OQ, and GAWs across six phonation types, derived from dynamic 3D MRI. (A–C) Structural and dynamic parameters including vertical and horizontal displacement (A), VF length (B), and VF thickness (C); (D) OQ values derived from MRI-based GAW; (E) dynamic VF contact area over one oscillatory cycle, with static components subtracted to isolate the oscillating portion; (F) MRI-based GAWs across the cycle, illustrating systematic differences in oscillatory amplitude and temporal profile.

https://doi.org/10.1121/10.0043582

between OQ and VF length were less consistent, mostly because of the condition *Thin* (disproportionately long relative to OQ) and the different OQ values between the conditions *Thick & AES narrow* and *Thick & VenF constrict*, despite identical length, which likely reflected differences in vibratory amplitude.

The methodology used to calculate the OQ (i.e., MRI, HSV, or EGG) had only a minor influence on the observed associations, supporting the overall equivalence of these three approaches. However, MRI-derived OQ values were consistently about 0.2 higher. This systematic offset may be related to the segmentation thresholds used, as the grayscale contrast between air and tissue boundaries is more distinct in HSV than in MRI. Such limitations could be mitigated in future work by employing higher spatial resolution and more refined motion-correction algorithms. Interestingly, the association between MRI-derived OQ and VF thickness appeared even more linear than that between GAW- or EGG-based OQ and VF thickness, although this relationship was not analyzed statistically because of the small number of data pairs. Further studies with larger datasets are therefore warranted.

Despite this overall agreement, notable differences emerged in GAW amplitude patterns. HSV-derived GAWs followed the order *Stiff* > *Thick* > *Thin* > *Thick & AES narrow* > *Thick & VenF retract* > *Thick & VenF constrict* (Fig. 5), whereas MRI-derived GAWs ranked *Stiff* > *Thick & AES narrow* > *Thick* > *Thick & VenF retract* > *Thin* > *Thick & VenF constrict* (Fig. 10). When normalized to the minimum, *Stiff* shifted between *Thin* and *Thick & VenF retract*. The observed mismatch in GAW amplitudes between MRI- and HSV-based measures can be attributed to differences between the two approaches: with the explicit goal of mimicking a two-dimensionally defined endoscopic GAW, the MRI-based GAW was intentionally designed to approximate a cranially oriented camera view by summing all non-tissue-occupied pixels in a superior projection. However, HSV-based measurements are inherently affected by camera incidence angle–related distortions. As illustrated in the supplementary data (Traser *et al.*, 2026), video S1_HSV_EGG_GAW.mp4, the endoscopic recordings exhibit a slight left-lateralized viewing angle, which can influence glottal segmentation (Veltrup *et al.*, 2023), whereas the simulated camera position in the MRI-based analysis is centered along the midline. This difference is particularly relevant for phonation types with reduced ventricular volume, where supraglottic constriction—such as at the level of the aryepiglottic sphincter [see supplementary data (Traser *et al.*, 2026), Data 4, HSV video of *Thick & AES narrow*]—may partially obscure the glottal opening in endoscopic images, thereby leading to an underrepresentation of the maximum glottal opening in HSV versus MRI GAW. Such occlusion does not affect the MRI-based GAW, as region-of-interest selection was restricted to the glottis itself. Finally, the distance between the endoscope and the larynx may vary substantially across HSV recordings, which can further contribute to over- or underrepresentation of the glottal area relative to other phonation types. Such variability is excluded in the MRI-based data, where the simulated camera distance is fixed. Comparative analyses of physical VF models using both MRI and HSV, including laser triangulation, are planned to further address and quantify this issue in the future.

As MRI and HSV/EGG were acquired in separate sessions (see Sec. IV D), also small task-related variations cannot be excluded. Nevertheless, the overall correspondence and the systematic divergences observed across modalities highlight both the validity of MRI-derived measures and the potential of combining method-specific insights to refine our understanding of phonatory mechanisms.

Across phonation types, systematic differences in acoustic perturbation measures were observed. Because of the challenging acoustic environment during MRI acquisition, absolute jitter and shimmer values cannot be directly compared to clinical reference thresholds obtained in low-noise recording conditions. Instead, the observed perturbation patterns should be interpreted relative to phonation type and within the context of MRI-specific acoustic constraints. In this regard, softer and breathy phonation types were associated with higher jitter values, whereas louder phonation types exhibited increased shimmer. This pattern is physiologically plausible and consistent with previous reports as frequency perturbation tends to increase under conditions of reduced medial compression and oscillatory stability (Zhang, 2016b), whereas amplitude perturbation is more sensitive to increased SPL and collision forces (Baken and Orlikoff, 2000).

Importantly, acoustic measures remained largely stable over time for most recordings. An exception was observed for *Thick* phonation, where an increase in jitter coincided with a concurrent increase in H1–H2 approximately halfway through the acquisition, suggesting a gradual change in VF oscillatory behavior or resonance characteristics of the vocal tract. Despite this change, VF oscillation remained clearly identifiable in the MRI data and could be robustly analyzed, indicating that the reconstruction approach is tolerant to moderate temporal variations in oscillatory behavior.

In the future, real-time acoustic monitoring may serve as an online quality control tool to provide feedback on phonatory stability and the expected evaluability of a given acquisition. However, additional data will be required to determine critical perturbation thresholds beyond which reconstruction quality becomes substantially compromised.

B. Phonation type comparison

Each phonation type demonstrated a distinct constellation of VF geometry, supraglottic shaping, and oscillatory dynamics. One of the most salient discriminative features across conditions was the 3D configuration of the laryngeal ventricle (see Fig. 11). Specifically, *Thin*, *Stiff*, and *Thick & VenF retract* phonation exhibited a markedly enlarged laryngeal ventricle with pronounced cranial extension of the *sacculus laryngis*, whereas *Thick*, *Thick & AES narrow*, and *Thick & VenF constrict* phonation showed a substantially

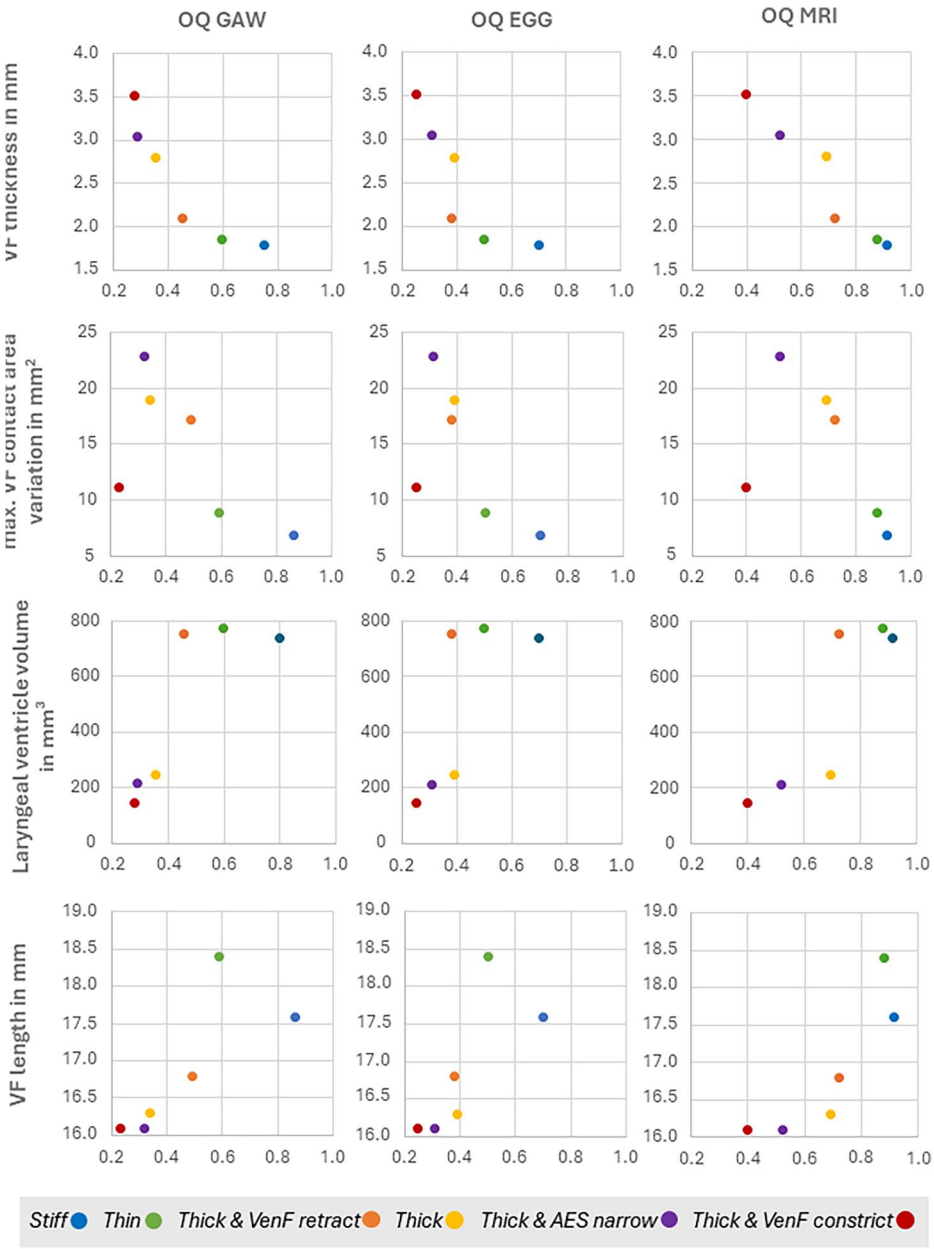


FIG. 11. Relationship between OQ derived from the GAW (left), EGG (middle), and MRI (right) with VF thickness, maximum VF contact area variation, laryngeal ventricle volume, and VF length across six voice production types. Each data point represents the mean value for one phonation type as measured in dynamic VF MRI. The phonation types are color coded in *Stiff* (blue), *Thin* (green), *Thick & VenF retract* (orange), *Thick* (yellow), *Thick & AES narrow* (violet), *Thick & VenF constrict* (red).

reduced ventricular volume without visible sacculus formation (see also Figs. 7 and 8). Differences in supraglottic configuration also appeared to influence VF geometry, suggesting a functional coupling between supraglottic adjustments and glottal biomechanics. Nevertheless, the individual phonation types displayed additional characteristic features reflecting distinct patterns of muscular activation and consistent with specific acoustic-perceptual goals:

1. Expanded laryngeal ventricle phonation types

Thin, *Stiff*, and *Thick & VenF retract* exhibited thinner and longer VF and higher OQ than phonation types from the constricted laryngeal ventricle group. Here, *Thin* and *Stiff* showed particularly long VF with small VF contact area, consistent with a CT-dominant, M2/head-register production (Roubeau *et al.*, 2009; Vahabzadeh-Hagh *et al.*, 2018; Zhang, 2016a). Although *Stiff* phonation exhibited typical features of breathy phonation, including incomplete glottal closure, it resulted in the smallest contact areas, reduced SPL, and limited harmonic richness. In contrast, Thin showed a very symmetric oscillatory pattern with minimal but nearly complete VF contact. This is consistent with the previous literature showing that OQ is primarily determined by VF thickness and further increases with glottal gap size, as demonstrated in simulations (Zhang, 2024) and *in vivo* canine studies (Choi *et al.*, 1993; Chhetri *et al.*, 2012; Luegmair *et al.*, 2014). During *Stiff* phonation, MRI revealed particularly large VF excursions, whereas SPL was relatively low in this condition. Although peak collision pressures may be reduced with incomplete closure (Jiang and Titze, 1994; Zhang, 2024), attempting high SPL in this configuration by elevated subglottal pressure can increase dissipated collision power—a possible predictor of

https://doi.org/10.1121/10.0043582

phonotrauma (Hillman *et al.*, 2020; Motie-Shirazi *et al.*, 2022; Zañartu *et al.*, 2014). Additionally, higher transglottal airflow that accompanies *Stiff* phonation can accelerate epithelial surface dehydration, altering tissue mechanics and increasing phonatory effort (Wu and Zhang, 2017). Thus, if not voluntarily used with sufficient microphone support or acoustic surrounding, *Stiff* may impose significant tissue load.

Thick & VenF retract displayed features consistent with a TA-dominant M1/modal register pattern, even though the laryngeal ventricle was even larger than in *Stiff* phonation. This included shorter VFs with moderate inferior bulging, increased contact area, and reduced OQ compared to *Thin* and *Stiff* while remaining thinner than *Thick* phonation. Thus, *Thick & VenF retract* combined thick-type VF geometry with wide ventricular spacing and reduced VF thickness.

It remains unclear whether expansion of the laryngeal ventricle alone mechanically reduces effective VF thickness through tissue traction or whether coordinated muscular interconnections between supraglottic and glottal structures contribute to fine-tuning of TA activity (Gocht *et al.*, 2023). Although ventricular expansion was associated with a reduction in VF thickness, intrinsic laryngeal muscle activation appears capable of counteracting this effect, highlighting the dynamic interplay between supraglottic configuration and intrinsic glottal control.

2. Constricted laryngeal ventricle phonation types

Thick, *Thick & AES narrow*, and *Thick & VenF constrict* were characterized by shorter, thicker VF, reduced laryngeal ventricle volume, and reduced OQ compared to the expanded laryngeal ventricle group. *Thick* phonation exhibited short, thick, rectangular VFs with pronounced inferior bulging, consistent with dominant TA activation (Vahabzadeh-Hagh *et al.*, 2017; Wu and Zhang, 2019). Functionally, this configuration showed low OQ, broad EGG peaks, and large dynamic contact areas, in line with modal or chest register (M1) production (Roubeau *et al.*, 2009; Vahabzadeh-Hagh *et al.*, 2018; Zhang, 2016a). MRI revealed smaller vertical relative to horizontal displacements, suggesting increased medial thickness/stiffness and an oscillatory pattern that supports efficient harmonic excitation (Zhang, 2016b).

Thick & AES narrow further intensified this biomechanical strategy through marked anteroposterior narrowing of the *aditus laryngis*. This configuration produced the highest SPL, rapid closing dynamics, reduced OQ, large dynamic contact areas, and enhanced high-frequency spectral energy, consistent with belting, twang, or dramatic operatic phonation (Yanagisawa *et al.*, 1989). Notably, supraglottic constriction did not impair vibratory behavior; rather, the combination of thick, compliant VF with strong epilaryngeal narrowing appeared to promote efficient oscillation, reflected in wide vibration amplitudes and increased spectral energy. Such supraglottic shaping has also been reported in clinical observations of trained singers (Mayerhoff *et al.*, 2014).

In contrast, *Thick & VenF constrict* represented the most compact configuration, with minimal VF excursions, the narrowest ventricular gap, and the smallest laryngeal ventricle. Although VF length was comparable to *Thick & AES narrow*, oscillation amplitudes were disproportionately reduced, likely because of the large VF thickness and likely VF medial compression. Reduced SPL and diminished spectral output suggest compromised vocal efficiency despite similar OQ values, resembling patterns described in muscle tension dysphonia (Nemetz *et al.*, 2005).

C. Covariation of supraglottal configuration and VF thickness

The present MRI data demonstrate that supraglottic geometry covaries with VF thickness and vibratory behavior. Reductions in laryngeal ventricle volume were associated with increased VF thickness and lower OQ, whereas increased laryngeal ventricle volume was associated with reduced VF thickness and increased OQ (Fig. 11), highlighting that supraglottic posturing may directly shape VF configuration beyond its effects on impedance matching (Döllinger *et al.*, 2006; Kniesburges *et al.*, 2017; Titze and Story, 1997), glottal jet stability (Xue and Zheng, 2017; Zheng *et al.*, 2009), or spectral reinforcement (Sundberg, 1974; Takemoto *et al.*, 2006). Similar covariation between VF and supraglottal adjustments has also been observed in a recent perception study (Zhang and Steinhauer, 2026), which shows that voices produced with the *Thick & VenF* retract configuration were often perceived as produced by a *Thin* VF configuration.

Importantly, medio–lateral VenF position or AES narrowing alone did not determine the OQ, as shown by the contrast between *Thick & AES narrow* and *Thick & VenF constrict*, suggesting that the supraglottal posturing impacts the voice source indirectly through the covariations of supraglottic shaping with other glottal adjustments. This may explain why numerical simulations report limited source effects of supraglottic changes except under extreme adduction (Yoshinaga and Zhang, 2025; Zhang, 2023) as such models do not capture the muscular co-variations observed *in vivo*.

Recent anatomical work emphasizes the integrated functional anatomy of the *thyroarytenoideus internus (vocalis)*, *thyroarytenoideus externus*, *thyroepiglottic*, and *ventricularis muscles* (Gocht *et al.*, 2023). Given the key role of the *vocalis muscle* in controlling medial VF thickness, its close functional association with the *thyroepiglottic muscle*—a cranial subdivision of the thyroarytenoid complex—is particularly noteworthy. The orientation of this muscle makes it a plausible driver of anteroposterior approximation between the epiglottis and the arytenoid cartilages. It may thus contribute not only to narrowing of the *aditus laryngis* but also to VF shortening and inferior-medial bulging, as observed in the *Thick & AES narrow* condition in the present data. Moreover, muscle fibers

mediating VenF adduction are anatomically linked to this complex. The current findings suggest, however, that their medio–lateral positioning can be controlled independently to some degree. The VenF fibers originate not only from the thyroarytenoid complex but also from the aryepiglottic portion of the oblique arytenoid muscle and may receive dual innervation from both the superior and recurrent laryngeal nerves.

D. Limitations

This study has several limitations, which are the following:

(1) Even if EGG in MRI is possible (Özen *et al.*, 2016) MRI and HSV recordings could not be acquired simultaneously; instead, measurements were performed in separate sessions, and although the participant was highly trained in controlled voice production, identical phonatory conditions across modalities cannot be guaranteed as differences in recording environment, body position (MRI: supine; HSV: upright) can lead to systematic differences (Traser *et al.*, 2014; Traser *et al.*, 2021), and task duration may have influenced phonatory behavior.
(2) The single-subject design precludes assessment of inter-individual variability and does not allow statistical evaluation or generalization beyond this exemplary case.
(3) The current MRI sequence does not allow reliable differentiation of cartilaginous structures or precise assessment of vertical laryngeal position, which is relevant because laryngeal lowering tends to widen supraglottic spaces whereas elevation tends to narrow them (Fleischer *et al.*, 2022).
(4) Tissue–air contrast in MRI is reduced compared to HSV, resulting in less sharply defined glottal boundaries and further limiting the comparability of derived GAW.
(5) Spatial resolution and measurement precision represent an additional limitation, as although the native MRI resolution was 0.8 mm and images were upscaled to a pixel size of 0.4 mm using zero-filling, measurements of VF thickness or displacement at small lateral offsets approach the limits of spatial resolution and may be susceptible to partial volume effects.

Future advances in MRI acquisition speed, spatial resolution, tissue contrast, and multi-subject study designs will be essential to address these limitations and further support clinical translation.

E. Conclusion and future directions

This proof-of-concept study presents the first *in vivo* dynamic 3D MRI analysis of six distinct phonation types and demonstrates strong correspondence with established HSV and EGG measures, validating 3D MRI as a powerful complementary tool in voice research. Although not intended to replace endoscopic imaging techniques central to clinical voice assessment, dynamic 3D VF MRI enables quantitative analysis of oscillatory motion, VF contact patterns, and volumetric characteristics that are inaccessible from a superior endoscopic view. Although current clinical translation is limited by the requirement for prolonged stable phonation and is therefore best suited for trained or healthy voices, the method is non-invasive, implementable on clinical 3T systems with minimal additional hardware, and holds promise for future application in selected clinical scenarios involving vertical VF geometry.

Beyond methodological validation, the presented data provide new physiological insights into voice production. The results demonstrate systematic differences in VF geometry associated with supraglottic configuration: an expanded laryngeal ventricle was consistently associated with thinner and longer VF, whereas a constricted ventricle co-occurred with increased VF thickness and more compact glottal configurations. These findings suggest that supraglottic shaping constitutes an additional dimension of glottal configuration during phonation, alongside intrinsic glottal muscle activation. Importantly, TA- versus CT-dominant activation patterns appeared capable of modulating or partially counteracting supraglottic influences, indicating a dynamic interplay between supraglottic and glottal control mechanisms.

Furthermore, the data support the use of OQ as a robust *in vivo* proxy for medial VF thickness under conditions of complete cartilaginous closure, consistent with prior simulation-based predictions. The strong correspondence between intended phonation types and achieved laryngeal configurations, following the Estill Voice Training® framework, underscores that these complex adjustments can be intentionally and consistently reproduced through training. Although derived from a single expert singer, the findings provide a detailed 3D depiction of laryngeal behavior across phonation types and highlight the potential of dynamic 3D MRI to bridge structural, acoustic, and modeling approaches in future voice science.

ACKNOWLEDGMENTS

L.T. and F.S. contributed equally to this work. J.F. and M.E. share senior authorship. L.T., J.F., and M.E. were supported by the Deutsche Forschungsgemeinschaft (German Research Foundation (Contract Nos. Tr1491/4-1, Fi2803/1-1, Ec409/1-4, and Ec409/5-1). We also acknowledge DFG for financial support (Contract No. INST 380/94-1 FuGG) and the CoreFacility Magnetic Resonance Development and Application Center Freiburg (MRDAC), Department of Radiology, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg, for support in acquisition of the data.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The study followed the ethical principles established by the World Medical Association Declaration of Helsinki and was approved by the ethics committee of the Freiburg University Clinic (23–1042-S1). The participant provided informed consent.

DATA AVAILABILITY

The data that support the findings of this study, including 3D MRI volumes, STL surface reconstructions, synchronized audio recordings, and animation files, are publicly available on Zenodo, https://zenodo.org/records/19629778, and publicly available under a Creative Commons BY 4.0 license.

The supplementary data includes the following:

(1) sixty datasets of complete 3D surface reconstructions for each phonation type (ten time frames per type) (.stl datasets of 3D surface reconstruction);
(2) twelve audio samples from HSV and MRI recording (audio recordings);
(3) ten full 3D MRI datasets of all six phonation types in NRRD (phonation type MRI datasets);
(4) HSV recordings with simultaneous EGG and GAW presentation (supplementary data video S1_HSV_EGG_GAW.mp4);
(5) animation of 3D surface reconstruction in six phonation types (supplementary data video S2: dynamic 3D surface reconstruction.mp4);
(6) animation of sagittal, coronal, and axial views of 3D dynamic VF MRI for six phonation types (supplementary data video S3_slices animation.gif);dynamic VF contact area animation in sagittal slice for six phonation types (supplementary data video S4_contact area.gif); and
(7) six voice analysis MRI plots including spectrogram, jitter (percentage), shimmer (percentage), and H1–H2 plotted over time (supplementary data voice analysis MRI plots).

Aaen, M., McGlashan, J., and Sadolin, C. (**2020**). "Laryngostroboscopic exploration of rough vocal effects in singing and their statistical recognizability: An anatomical and physiological description and visual recognizability study of distortion, growl, rattle, and grunt using laryngostroboscopic imaging and panel assessment," J. Voice **34**, 162.e5–162.e14.
Alipour, F., and Scherer, R. C. (**2000**). "Vocal fold bulging effects on phonation using a biophysical computer model," J. Voice **14**, 470–483.
Bailly, L., Henrich Bernardoni, N., Müller, F., Rohlfs, A.-K., and Hess, M. (**2014**). "Ventricular-fold dynamics in human phonation," J. Speech Lang. Hear. Res. **57**, 1219–1242.
Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Publishing Group, San Diego, CA).
Berry, D. A., Montequin, D. W., and Tayama, N. (**2001**). "High-speed digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547.
Chhetri, D. K., Neubauer, J., and Berry, D. A. (**2012**). "Neuromuscular control of fundamental frequency and glottal posture at phonation onset," J. Acoust. Soc. Am. **131**, 1401–1412.
Choi, H. S., Ye, M., Berke, G. S., and Kreiman, J. (**1993**). "Function of the thyroarytenoid muscle in a canine laryngeal model," Ann. Otol. Rhinol. Laryngol. **102**, 769–776.
Döllinger, M., Berry, D. A., and Berke, G. S. (**2005**). "A quantitative study of the medial surface dynamics of an *in vivo* canine vocal fold during phonation," Laryngoscope **115**, 1646–1654.
Döllinger, M., Berry, D. A., and Montequin, D. W. (**2006**). "The influence of epilarynx area on vocal fold dynamics," Otolaryngol. Head Neck Surg. **135**, 724–729.
Echternach, M., Burk, F., Köberlein, M., Selamtzis, A., Döllinger, M., Burdumy, M., Richter, B., and Herbst, C. T. (**2017**). "Laryngeal evidence for the first and second passaggio in professionally trained sopranos," PLoS One **12**, e0175865.
Fischer, J., Abels, T., Özen, A. C., Echternach, M., Richter, B., and Bock, M. (**2020**). "Magnetic resonance imaging of the vocal fold oscillations with sub-millisecond temporal resolution," Magn. Resonance Med. **83**, 403–411.
Fischer, J., Jordan, P. L., Stritt, F., Richter, B., Traser, L., and Bock, M. (**2026**). "Dynamic 3D MRI for vocal fold oscillation measurements," Z Med. Phys. (published online).
Fischer, J., Özen, A. C., Traser, L., Echternach, M., Richter, B., and Bock, M. (**2022**). "Sub-millisecond 2D MRI of the vocal fold oscillation using single point imaging with rapid encoding," Magn. Reson. Mater. Phys. Biol. Med. **35**, 301–310.
Fleischer, M., Rummel, S., Stritt, F., Fischer, J., Bock, M., Echternach, M., Richter, B., and Traser, L. (**2022**). "Voice efficiency for different voice qualities combining experimentally derived sound signals and numerical modeling of the vocal tract," Front. Physiol. **13**, 1081622.
Gocht, A., Lüers, G., and Schumacher, U. (**2023**). "Amount and spatial arrangement of muscle fibers in the human laryngeal *Musculus ventricularis*," Clin. Anat. **36**, 1138–1146.
Hillman, R. E., Stepp, C. E., Van Stan, J. H., Zañartu, M., and Mehta, D. D. (**2020**). "An updated theoretical framework for vocal hyperfunction," Am. J. Speech. Lang. Pathol. **29**, 2254–2260.
Inwald, E. C., Döllinger, M., Schuster, M., Eysholdt, U., and Bohr, C. (**2011**). "Multiparametric analysis vocal fold vibrations healthy disordered voices using high-speed imaging," J. Voice **25**, 576–590.
Jiang, J. J., and Titze, I. R. (**1994**). "Measurement of vocal fold intraglottal pressure and impact stress," J. Voice **8**, 132–144.
Khosla, S., Oren, L., Ying, J., and Gutmark, E. (**2014**). "Direct simultaneous measurement of intraglottal geometry and velocity fields in excised larynges," Laryngoscope **124**, S1–S13.
Kniesburges, S., Birk, V., Lodermeyer, A., Schützenberger, A., Bohr, C., and Becker, S. (**2017**). "Effect of the ventricular folds in a synthetic larynx model," J. Biomech. **55**, 128–133.
Köberlein, M., Hünnemeyer, P., Burk, F., Burdumy, M., Richter, B., Echternach, M., and Traser, L. (**2024**). "Vocal tract configurations of professional operatic singers during sustained phonation," J. Voice **38**, 1238–1247.
Lehoux, S., and Zhang, Z. (**2024**). "A methodology to quantify the effective vertical thickness of prephonatory vocal fold medial surface," J. Voice. **38**, 1465–1473.
Luegmair, G., Chhetri, D. K., and Zhang, Z. (**2014**). "The role of thyroarytenoid muscles in regulating glottal closure in an *in vivo* canine larynx model," Proc. Mtgs. Acoust. **22**, 060007.
Mayerhoff, R. M., Guzman, M., Jackson-Menaldi, C., Munoz, D., Dowdall, J., Maki, A., Johns, M. M., Smith, L. J., and Rubin, A. D. (**2014**). "Analysis of supraglottic activity during vocalization in healthy singers," Laryngoscope **124**, 504–509.
Motie-Shirazi, M., Zañartu, M., Peterson, S. D., Mehta, D. D., Hillman, R. E., and Erath, B. D. (**2022**). "Collision pressure and dissipated power dose in a self-oscillating silicone vocal fold model with a posterior glottal opening," J. Speech Lang. Hear. Res. **65**, 2829–2845.
Nemetz, M. A., De Lima Pontes, P. A., Pedrosa Vieira, V., and Kazuo Yazaki, R. (**2005**). "Vestibular fold configuration during phonation in adults with and without dysfonia," Rev. Bras. Otorrinolaringol. **71**, 6–12.
Özen, A. C., Traser, L., Echternach, M., Dadakova, T., Burdumy, M., Richter, B., and Bock, M. (**2016**). "Ensuring safety and functionality of electroglottography measurements during dynamic pulmonary MRI," Magn. Reson. Med. **76**, 1629–1635.
Patel, R. R., Dubrovskiy, D., and Döllinger, M. (**2014**). "Measurement of glottal cycle characteristics between children and adults: Physiological variations," J. Voice **28**, 476–486.
Roubeau, B., Henrich, N., and Castellengo, M. (**2009**). "Laryngeal vibratory mechanisms: The notion of vocal register revisited," J. Voice **23**, 425–438.

Semmler, M., Döllinger, M., Patel, R. R., Ziethe, A., and Schützenberger, A. (**2018**). "Clinical relevance of endoscopic three-dimensional imaging for quantitative assessment of phonation," Laryngoscope **128**, 2367–2374.
Shembel, A. C., Morrison, R. A., McDowell, S., Smeltzer, J. C., Crocker, C., Bell, C., and Mau, T. (**2025**). "Relationships between laryngoscopic analysis metrics of supraglottic compression and vocal effort in primary muscle tension dysphonia," J. Voice **39**, 1571–1578.
Steinhauer, K., Klimek, M., and Estill, J. (**2017**). *The Estill Voice Model: Theory and Translation* (Estill Voice International, Pittsburgh, PA).
Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant,'" J. Acoust. Soc. Am. **55**, 838–844.
Švec, J. G., and Zhang, Z. (**2025**). "Application of nonlinear dynamics theory to understanding normal and pathologic voices in humans," Philos. Trans. R. Soc. B Biol. Sci. **380**, 20240018.
Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., and Honda, K. (**2006**). "Acoustic roles of the laryngeal cavity in vocal tract resonance," J. Acoust. Soc. Am. **120**, 2228–2238.
Titze, I. R., and Story, B. H. (**1997**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.
Traser, L., Burdumy, M., Richter, B., Vicari, M., and Echternach, M. (**2014**). "Weight-bearing MR imaging as an option in the study of gravitational effects on the vocal tract of untrained subjects in singing phonation," PLoS One **9**, e112405.
Traser, L., Fleischer, M., Priegnitz, D., Stritt, F., Köberlein, M., Kirsch, J., Fischer, J., Herzel, H.-P., Mürbe, D., Richter, B., and Echternach, M. (**2025**). "*In vivo* insights into irregular voice production as a complex nonlinear system—a case study," J. R. Soc. Interface **22**, 20250086.
Traser, L., Schwab, C., Burk, F., Özen, A. C., Burdumy, M., Bock, M., Richter, B., and Echternach, M. (**2021**). "The influence of gravity on respiratory kinematics during phonation measured by dynamic magnetic resonance imaging," Sci. Rep. **11**, 21524.
Traser, L., Stritt, F., Jordan, P., Köberlein, M., Kirsch, J., Rummel, S., Zhang, Z., Richter, B., Bock, M., Fischer, J., and Echternach, M. (**2026**). Supplementary Dataset for "Dynamic 3D MRI of vocal fold oscillations: In vivo assessment of vocal fold thickness, contact area, and glottal area waveform across phonation types in comparison with high-speed imaging," Zenodo, V. **3**, Dataset. https://doi.org/10.5281/zenodo.19629778.
Vahabzadeh-Hagh, A. M., Zhang, Z., and Chhetri, D. K. (**2017**). "Quantitative evaluation of the *in vivo* vocal fold medial surface shape," J. Voice **31**, 513.e15.
Vahabzadeh-Hagh, A. M., Zhang, Z., and Chhetri, D. K. (**2018**). "Hirano's cover-body model and its unique laryngeal postures revisited," Laryngoscope **128**, 1412–1418.
Veltrup, R., Kniesburges, S., and Semmler, M. (**2023**). "Influence of perspective distortion in laryngoscopy," J. Speech Lang. Hear. Res. **66**, 3276–3289.
Wu, L., and Zhang, Z. (**2017**). "A computational study of vocal fold dehydration during phonation," IEEE Trans. Biomed. Eng. **64**, 2938–2948.
Wu, L., and Zhang, Z. (**2019**). "Voice production in a MRI-based subject-specific vocal fold model with parametrically controlled medial surface shape," J. Acoust. Soc. Am. **146**, 4190–4198.
Xue, Q., and Zheng, X. (**2017**). "The effect of false vocal folds on laryngeal flow resistance in a tubular three-dimensional computational laryngeal model," J. Voice **31**, 275–281.
Yanagisawa, E., Estill, J., Kmucha, S. T., and Leder, S. B. (**1989**). "The contribution of aryepiglottic constriction to "ringing" voice quality—A videolaryngoscopic study with acoustic analysis," J. Voice **3**, 342–350.
Yoshinaga, T., and Zhang, Z. (**2025**). "Effects of false vocal fold adduction and aryepiglottic sphincter narrowing on the voice source in a three-dimensional voice production model," J. Acoust. Soc. Am. **157**, 2408–2421.
Zañartu, M., Galindo, G. E., Erath, B. D., Peterson, S. D., Wodicka, G. R., and Hillman, R. E. (**2014**). "Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction," J. Acoust. Soc. Am. **136**, 3262–3271.
Zhang, Z. (**2016a**). "Mechanics of human voice production and control," J. Acoust. Soc. Am. **140**, 2614–2635.
Zhang, Z. (**2016b**). "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model," J. Acoust. Soc. Am. **139**, 1493–1507.
Zhang, Z. (**2023**). "The influence of source-filter interaction on the voice source in a three-dimensional computational model of voice production," J. Acoust. Soc. Am. **154**, 2462–2475.
Zhang, Z. (**2024**). "Interaction effects in laryngeal and respiratory control of the voice source and vocal fold contact pressure," J. Acoust. Soc. Am. **156**, 4326–4335.
Zhang, Z. (**2025**). "Vocal fold vertical thickness in human voice production and control: A review," J. Voice **39**, 1183–1191.
Zhang, Z., and Chhetri, D. K. (**2019**). "Effect of changes in medial surface shape on voice production in excised human larynges," J. Acoust. Soc. Am. **146**, EL412–EL417.
Zhang, Z., and Steinhauer, K. (**2026**). "Perceiving physiology from the voice: Evidence for physiological coupling between laryngeal and epilaryngeal adjustments," J. Voice (published online).
Zheng, X., Bielamowicz, S., Luo, H., and Mittal, R. (**2009**). "A computational study of the effect of false vocal folds on glottal flow and vocal fold vibration during phonation," Ann. Biomed. Eng. **37**, 625–642.