



# OPEN Biomechanics of sound production in high-pitched classical singing

Matthias Echternach<sup>1,10</sup>✉, Fabian Burk<sup>2</sup>, Marie Köberlein<sup>1</sup>, Michael Döllinger<sup>3</sup>, Michael Burdumy<sup>4</sup>, Bernhard Richter<sup>5</sup>, Ingo R. Titze<sup>6</sup>, Coen P. H. Elemans<sup>7</sup> & Christian T. Herbst<sup>8,9,10</sup>✉

Voice production of humans and most mammals is governed by the MyoElastic-AeroDynamic (MEAD) principle, where an air stream is modulated by self-sustained vocal fold oscillation to generate audible air pressure fluctuations. An alternative mechanism is found in ultrasonic vocalizations of rodents, which are established by an aeroacoustic (AA) phenomenon without vibration of laryngeal tissue. Previously, some authors argued that high-pitched human vocalization is also produced by the AA principle. Here, we investigate the so-called “whistle register” voice production in nine professional female operatic sopranos singing a scale from C6 ( $\approx 1047$  Hz) to G6 ( $\approx 1568$  Hz). Super-high-speed videolaryngoscopy revealed vocal fold collision in all participants, with closed quotients from 30 to 73%. Computational modeling showed that the biomechanical requirements to produce such high-pitched voice would be an increased contraction of the cricothyroid muscle, vocal fold strain of about 50%, and high subglottal pressure. Our data suggest that high-pitched operatic soprano singing uses the MEAD mechanism. Consequently, the commonly used term “whistle register” does not reflect the physical principle of a whistle with regard to voice generation in high pitched classical singing.

Operatic solo singing requires electronically un-amplified vocal sound production at sound levels that are suitable to compete with large orchestras and choirs. This kind of voice production for artistic purposes extends the fundamental frequency ( $f_0$ ) range well beyond what is used in human speech communication, where the average  $f_0$  is at about 120 Hz and 200 Hz for adult males and females, respectively<sup>1</sup>. The entire singing  $f_0$  range can only be covered if the different laryngeal production mechanisms available to the human voice—often termed voice “registers”<sup>2,3</sup>—are utilized.

The two main laryngeal mechanisms are mechanism **M1** (also frequently termed the “chest” or “modal” register, and typically used in speech and often in singing) and **M2** (also frequently termed “falsetto” or “head” register, mainly used in singing, but sometimes also in speech). The most extreme upper musical pitch range of operatic sopranos—typically sung by adult human females, who are generally known to phonate at higher frequencies than males due to their shorter vocal fold length—extends the voice to an  $f_0$  range of about 1000–1600 Hz, or about three octaves above the  $f_0$  of speech. This range, which is regularly accessible to professionally trained classical soprano singers, is acoustically characterized by a strong fundamental and weak overtones in comparison to other laryngeal mechanisms<sup>4,5</sup>, justifying a classification into a separate mechanism, **M3**. This **M3** voice production mechanism is commonly being called the “whistle register” (German: *Pfeifstimme*—cf.<sup>6</sup>) in singing voice pedagogy<sup>7</sup>. There is no clear definition of the actual range and mechanism of that register: Garnier et al.<sup>8</sup> found a transition to the whistle register already between the pitches D#5 and D6, while some authors only speak about a transition at ca. C6, and Titze would only call it whistle register from and above F6<sup>9</sup>. Mechanism **M3** is hypothesized to be distinguishable from **M2** due to differences (1) in the voice source, i. e., the laryngeal

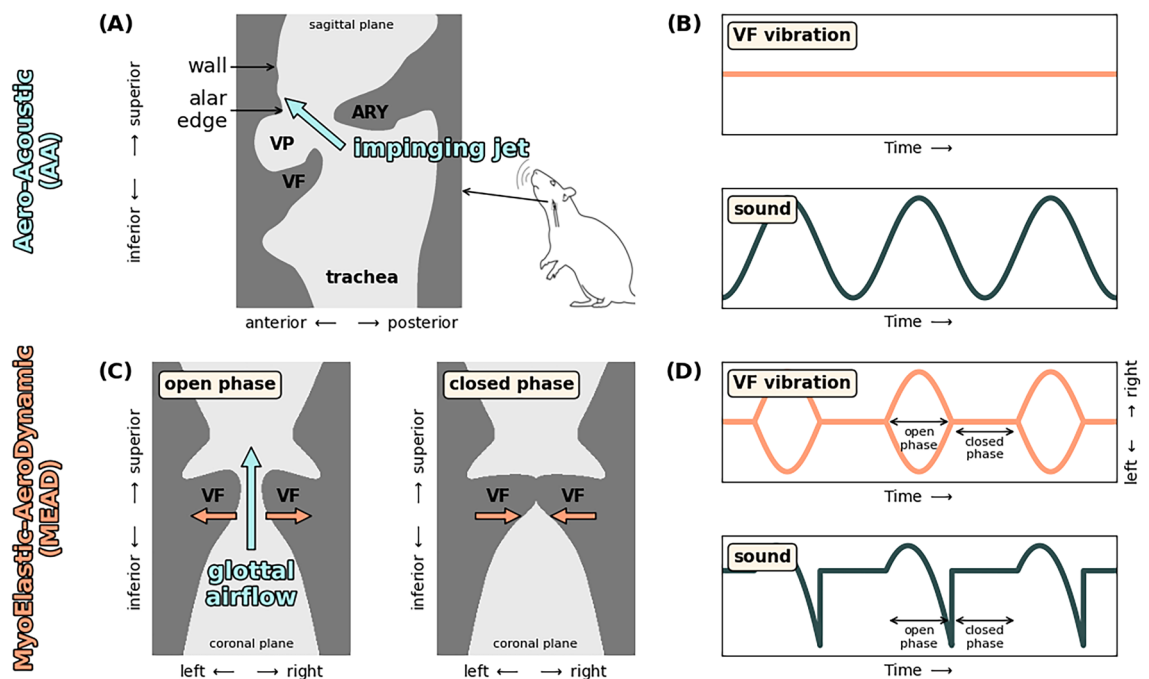
<sup>1</sup>Division of Phoniatics and Pediatric Audiology, Department of Otorhinolaryngology, LMU University Hospital, Marchioninstr. 15, 81377 Munich, Germany. <sup>2</sup>Department of Otorhinolaryngology and Plastic Surgery, SRH Wald-Klinikum Gera, Strasse des Friedens 122, Gera, Germany. <sup>3</sup>Division of Phoniatics and Pediatric Audiology, Department of Otorhinolaryngology Head and Neck Surgery, University Hospital Erlangen, Waldstr. 1, 91054 Erlangen, Germany. <sup>4</sup>Department of Medical Physics, Department of Radiology, Faculty of Medicine, Medical Center-University of Freiburg, Breisacher Str. 60, 79106 Freiburg, Germany. <sup>5</sup>Institute of Musicians’ Medicine, Freiburg University Medical Center and Faculty of Medicine Freiburg University, Elsäßer Str. 2m, 79110 Freiburg, Germany. <sup>6</sup>Utah Center for Vocology, 240 S 1500 E, Room 206, Salt Lake City, UT 84112, USA. <sup>7</sup>Vocal Neuromechanics Lab, Sound Communication and Behavior Group, Department of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. <sup>8</sup>Department of Behavioural and Cognitive Biology, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria. <sup>9</sup>Janette Ogg Voice Research Center, Shenandoah Conservatory, Winchester, VA, USA. <sup>10</sup>These authors contributed equally: Matthias Echternach and Christian T. Herbst. ✉email: matthias.echternach@med.uni-muenchen.de; info@christian-herbst.org

mechanism, (2) in the resonances or tuning strategy, (3) of interactions of the voice source and the resonances. Detailed essays on this can be found elsewhere<sup>2,3,10,11</sup>. The utilized terminology would suggest an aeroacoustic production mechanism, where a rigid structure combined with certain resonance causes airflow instabilities that produce such high frequencies.

In contrast, humans typically produce voice through the MyoElastic-AeroDynamic (MEAD) principle<sup>12–14</sup>. There, the vocal folds enter a state of self-sustained oscillation. The ensuing medio-lateral oscillation of the vocal folds, successively facilitating partial or full closure of the laryngeal airway, causes cyclic modulation of the exhalatory airflow. Those airflow fluctuations translate to pressure variations which are the main constituent of the generated vocal sound<sup>15,16</sup> (see Fig. 1C,D). In particular, the main acoustic excitation event is set up at the moment of airflow cessation during each cycle<sup>17</sup>, when the vocal folds are maximally approximated in their membranous part, often resulting in full vocal fold collision. The corresponding medio-lateral vibratory mode, resulting in air flow modulation variation, is a crucial requirement for the MEAD mechanism and thus distinguishes it from an aeroacoustic production principle.

MEAD is the predominant mechanism of sound production in mammals, extending across a range of body sizes and  $f_o$ , spanning more than four orders of magnitude from 10 Hz to 120 kHz. However, the murine rodents—which with 1400 species comprise about 25% of all 5400 mammal species—have adapted a completely different physical mechanism of sound production to extend their high frequency vocalization range. Murine rodents—including mice and rats—produce ultrasonic vocalizations (USV) with an aeroacoustic mechanism<sup>18</sup>. Several aeroacoustic mechanisms have been proposed to explain USVs in rodents that differ in their local flow conditions and acoustic feedback properties: (a) wall-impinging jets<sup>19,20</sup> (i.e., focused airflow that strikes an opposing surface—see Fig. 1A,B); (b) edge impinging jets (resulting from successive oscillation of an airflow jet to alternate sides of a ridge that is struck by the airflow)<sup>21</sup> and (c) cavity whistles (generated through air vortex oscillation within the cavity)<sup>20</sup>. In laboratory rats and mice, the wall-impinging whistle drives USV production, but alternative mechanisms may be found in the large number of rodent species<sup>19,20</sup>. Interestingly, in these species, the respective USV production mechanism co-exists with the “conventional” MEAD mechanism, the latter being exclusively used for vocalization at lower and thus humanly audible  $f_o$ .

Given the highly conserved laryngeal anatomy across mammals<sup>23</sup>, it might be possible that the operatic sopranos vocalizations at very high  $f_o$  are also produced by a special aeroacoustic mechanism. In point of fact, a number of authors describe the general possibility for an alternative aeroacoustic sound production mechanism in humans. This was hypothesized to be achieved by means of “chink tones” analogous to whistling<sup>24</sup> in the larynx and subsequent cavity resonance<sup>12</sup>, or vortex-induced vibration of the folds<sup>25–28</sup>, possibly involving interactions between the voice source and the vocal tract<sup>29,30</sup>. Such mechanisms are sometimes referenced to as M4 or “glottal whistle” for the  $f_o$  range of 1–3 kHz and above<sup>31</sup>. In such an aeroacoustic sound production mechanism it could be expected that the frequency of the whistle  $f_{m1}$  becomes  $f_o$ . Other studies showed that high-pitched human vocalization can be produced with the MEAD mechanism<sup>5,6,8,28,32–35</sup>. However, all those



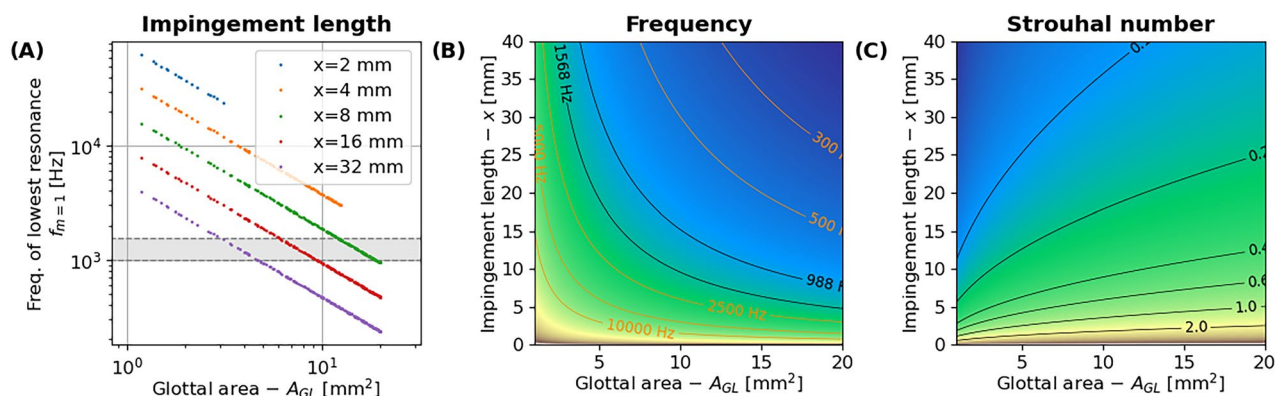
**Figure 1.** Schematic overview of Aero-Acoustic (AA) and MyoElastic-AeroDynamic (MEAD) voice production mechanisms. (A) Mid-sagittal view of rodent larynx, illustrating the formation of the impinging jet in AA sound production; (B) schematic display of the lack of tissue vibration in the AA mechanism and the resulting sinusoidal sound source; (C) coronal view of human larynx, illustrating the open and the closed phase of medio-lateral vocal fold vibration; (D) vocal fold displacement pattern and resulting prototypical acoustic voice source<sup>22</sup> for the MEAD mechanism.

studies come with certain limitations: They either (a) had a limited number of participants for laryngoscopic examination, i. e.,  $n = 1^{8,32-34}$ ,  $n = 2^{35}$ ,  $n = \text{unknown}^{28}$ ; (b) employed a limited data acquisition methodology which was either indirect, using electroglottography<sup>5</sup>, or did not allow direct observation of the vocal folds along their entire antero-posterior length<sup>32</sup>; and/or (c) had a limited temporal resolution (with the Nyquist frequency below  $f_o$ ), thus resulting in aliasing and preventing adequate time-resolved within-cycle documentation of the sound production mechanism<sup>6,8,34</sup>. Even though there is evidence for both principles, it remains uncertain which sound production mechanism and configuration are responsible for operatic singing in a range of  $f_o$  that resides at the bottom of the so-called whistle register range, which typically exceeds fundamental frequencies of 1000 Hz. In addition, a confusion regarding the terminology and the associated mechanisms and application in different music genres remains widespread up until today. In this study, we address the issue by providing the first comprehensive documentation of high-pitched soprano singing with super-HSV at 20,000 fps, investigating a larger cohort of professional operatic sopranos.

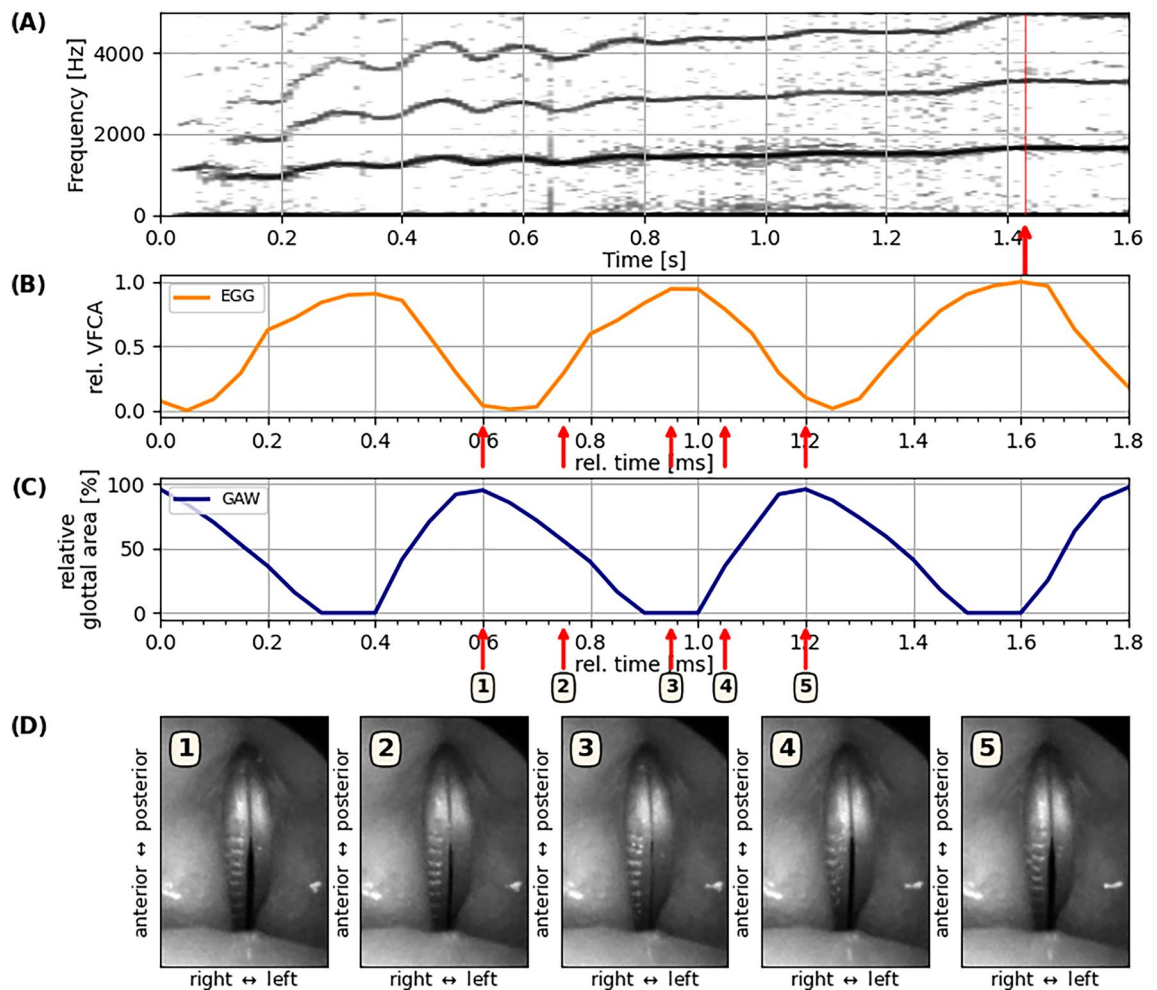
## Results

First, we tested whether the wall-impinging jet model developed for rodents<sup>20</sup> also applies to human voice production anatomy and physiology. Notably, our data (Fig. 2) suggest that such an aeroacoustic mechanism is hypothetically possible for high-pitched female operatic singing. The three independent parameters of that model—i. e., glottal area, impingement length (the length of the airflow jet that strikes the opposing surface), and volumetric airflow—could be gradually controlled by singers through glottal adduction, epiglottis tilt (i. e., a backwards rocking motion of the epiglottis, reducing the volume of the space just above the vocal folds), laryngeal constriction, and muscular adaptations of the pulmonary apparatus, thus theoretically allowing for gradual control of the emerging phonatory frequency in artistic contexts.

Contrary to these results from the simple aeroacoustic model, empirical data from the nine investigated professional singers strongly suggest that the aeroacoustic mechanism is not the origin of high-pitched soprano singing. As compelling supporting evidence for the MEAD mechanism, we found vocal fold vibration and collision in all nine participants (see supplementary materials for HSV samples from all participants). In all investigated sopranos, the vibratory frequency of the medio-lateral vocal fold oscillation corresponded to the  $f_o$  of the radiated sound (see Suppl. Fig. S1). This suggests that the tissue oscillation is causal to sound generation, which is highly indicative of the MEAD principle. This phenomenon is documented exemplarily in Fig. 3: The HSV still images in Fig. 3D document full glottal closure along the entire visible vocal fold length. The electroglottographic (EGG) signal (panel C) shows clear cyclic variation of vocal fold contact along the sagittal glottal plane. The glottal area, i. e., the opening between the left and right vocal folds co-varied in synchrony with the EGG signal. The resulting glottal area waveform (panel B) reached a maximum when the relative vocal fold contact area, as retrieved by the EGG signal, was at a minimum. This is in good agreement with prototypical human voice production in the modal register (e. g. in speech), where the vocal fold contact assumes a maximum when the glottis is closed), and vice-versa<sup>38</sup>. Without exception, the acoustic signals captured from the singers resembled a harmonic series with a number of noteworthy harmonics (between two and seven) throughout the examined vocal range, from pitch C6 ( $f_o \approx 1047$  Hz) to G6 ( $f_o \approx 1568$  Hz) in all investigated sopranos, with participant S3 achieving phonation at musical pitch B6 ( $f_o \approx 1975$  Hz—see Suppl. Fig. S2, out of the regular experimental protocol). The sound level differences (H/-H2) between the first and second harmonic of the radiated acoustic signal were in the range of 23.47 ( $\pm 8.21$ ) dB for the lowest  $f_o$  and 17.20 ( $\pm 4.48$ ) dB for the highest investigated  $f_o$  of each participant (see Suppl. Fig. S3,S4 for details).



**Figure 2.** Simulation of hypothetical frequencies for impinging jet sound production in humans, using the model presented in<sup>20</sup>. The resulting frequencies scale linearly as a function of volumetric airflow  $V$ . The figure illustrates the case of  $V = 150$  ml/sec, i. e., a default flow value seen in human singing<sup>36</sup>. (A) and (B) isoparametric curves for emerging mode-1 frequencies (i. e., lowest possible stable frequency of a whistle) as a function of glottal area ( $A_{GL}$ ) and impingement length ( $x$ ). The gray area in (A) depicts the fundamental frequency region of the “whistle” register in female operatic singing<sup>37</sup>. (C) Strouhal number  $St$  (a dimensionless quantity describing the oscillatory flow mechanism). Stable vortex whistles are expected at  $d/x < St < 1$ , where  $d$  is the jet diameter.



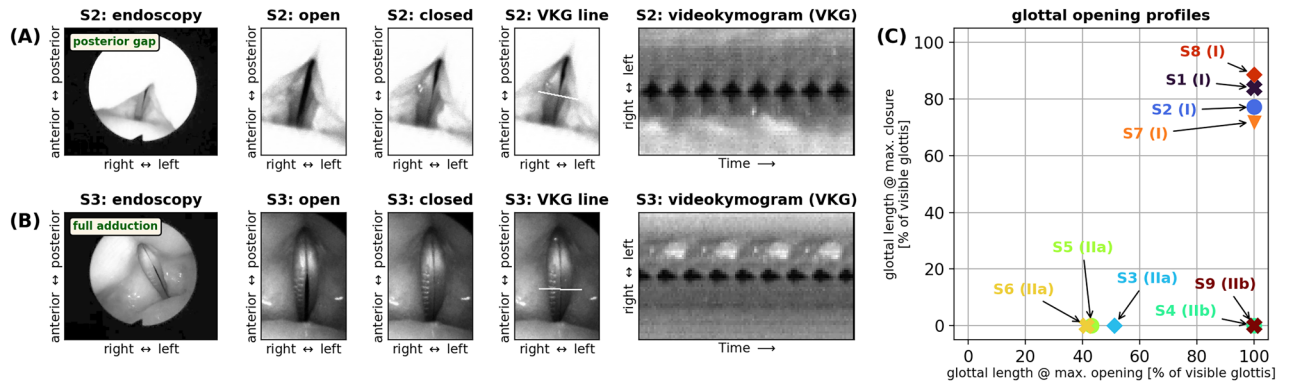
**Figure 3.** Example of high-pitched phonation of S3. (A) Acoustic spectrogram; (B) three cycles of laryngeal oscillation at  $t \approx 1.43$  s, showing vocal fold contact area documented with electroglottography (EGG); (C) time-varying glottal area waveform (GAW), as documented by high-speed video (HSV) recording at 20,000 frames per second. The arrows indicate the still HSV frames shown in panel d; (D) HSV frames extracted at the incidents indicated by the arrows in panels B and C. Note the full glottal closure in the third out of the five displayed video frames.

In contrast to the “default” laryngeal configuration in classical singing, which requires a moderately low vertical laryngeal position, phonation in the high-pitched soprano range was invariably facilitated by a raised larynx and moderate to extreme medialization of the ventricular folds (see Suppl. Fig. S5,S6). This suggests that the respective  $f_0$  could only be achieved with pronounced larynx elevation and/or constriction. Overall, we found the following stereotypical glottal configurations at the highest examined pitches, which are documented in Fig. 4 and Suppl. Fig. S5:

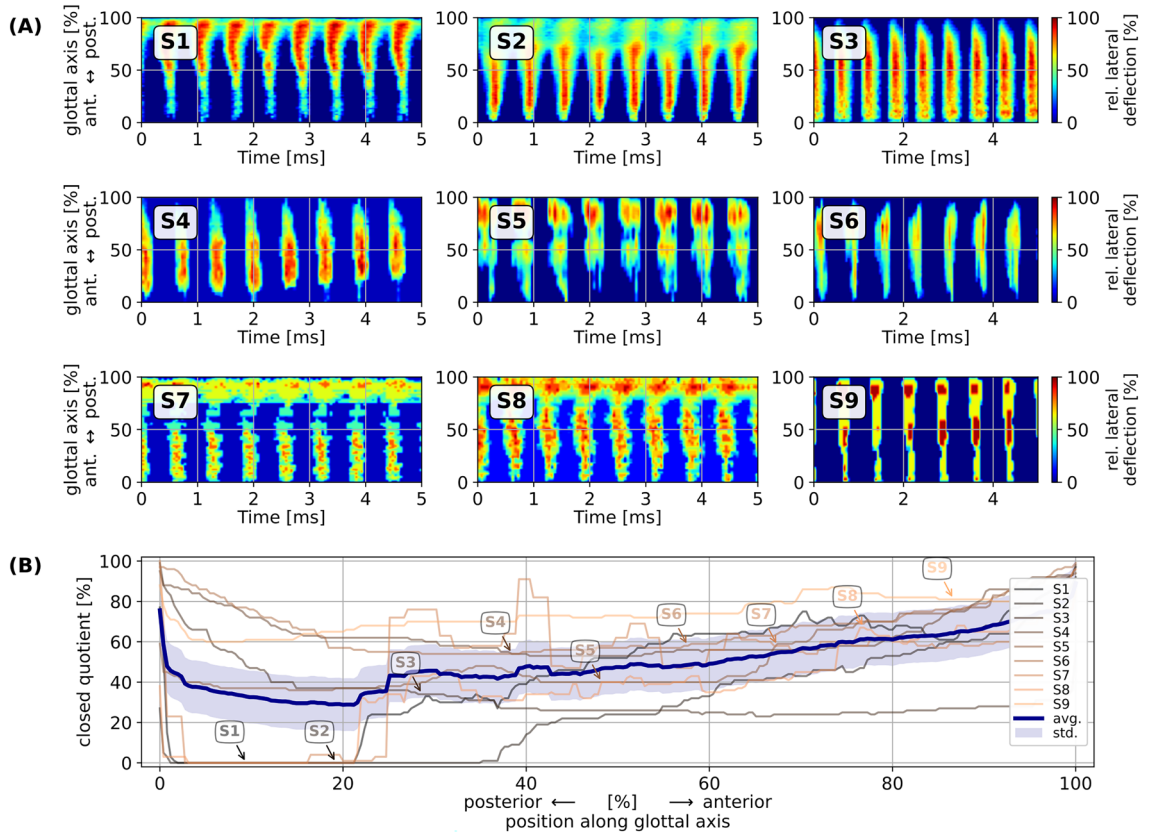
Four participants (S1, S2, S7, and S8) phonated with a posterior glottal gap (denoted as **glottal configuration I** throughout the remainder of this manuscript), suggesting incomplete vocal fold adduction (see Fig. 4A for an example). They had completely separated vocal folds (along the entire visible anterior-posterior length) in the open phase and a partially closed glottis in the closed phase (i.e., the duration of the oscillatory cycle where the vocal folds are in contact, temporarily stopping or at least greatly reducing the laryngeal air flow), with vocal fold contact along 44 % to 75 % of the visible glottal length.

In contrast, the five other participants (S3, S4, S5, S6, and S9) phonated with full glottal closure (100 % vocal fold contact) in the closed phase, but with different configurations during the open phase. Three participants (S3, S5, and S6) had only a partial opening of the vocal folds, occurring along 40 % to 50 % of the visible vocal fold length (**glottal configuration IIa**)—see Fig. 4B for an example. The other two participants (S4 and S9) phonated with a fully opened visible glottis in the open phase (**glottal configuration IIb**).

Respective documentation for all sopranos is provided in the supplementary materials (Suppl. Fig. S5,S6) and in Fig. 5. These data clearly corroborate the observation made in the HSV footage: The blue areas in the detail panels of Fig. 5 for S1 through S9 (A) are indicative of glottal closure and vocal fold collision, which occurred



**Figure 4.** Stereotypical glottal configurations in high-pitched operatic soprano singing. Two main strategies emerged: (A) **glottal configuration I**: phonation with incomplete vocal fold adduction, resulting in a posterior glottal gap during vibration (even in the “closed” phase); and (B) **glottal configuration II**: greatly increased adduction of the arytenoids, supported by medialization of the ventricular folds; (C) glottal opening profiles for all strategies—see supplementary materials S5 and S6 for details.

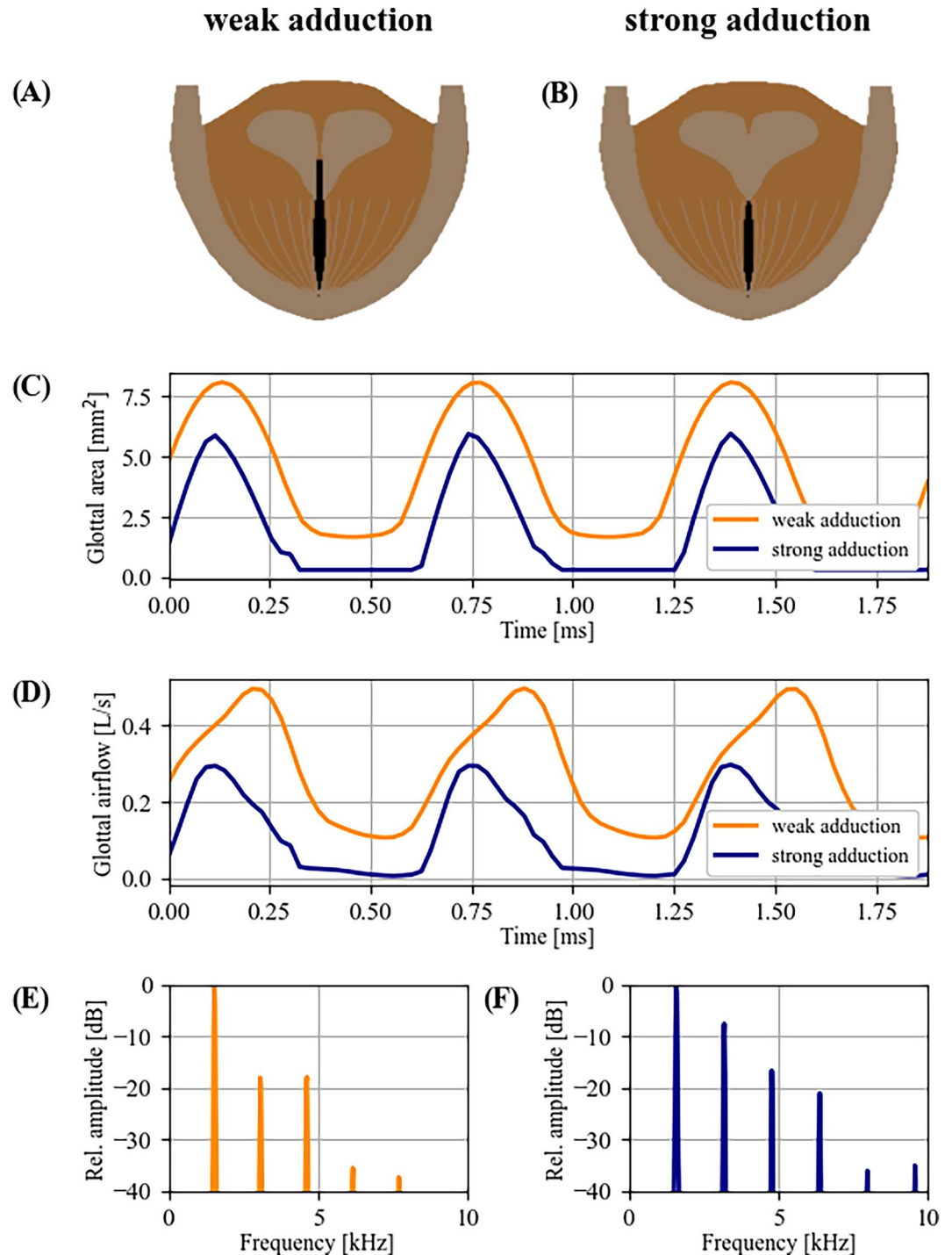


**Figure 5.** Vocal fold vibration analysis for phonation at pitch G6 (ca. 1568 Hz). (A) Glottovibrograms (GVG) for all investigated sopranos (S1 through S9); (B) individual and averaged glottal closed quotients along the antero-posterior axis (cf. Figure 4 in<sup>39</sup>).

either partially (S1, S2, S7, and S8) or along the entire antero-posterior length of the visible glottis (S3, S4, S5, S6, and S9). Due to the observed vocal fold collision, the closed quotient (CQ), i. e., the relative duration of glottal closure over one vibratory cycle, was non-zero in most instances. Averaging all computed CQ values across the entire antero-posterior glottal length across all nine participants resulted in a median CQ value of 47.6%, with 5 and 95 percentiles at 29.6% and 73.0%, respectively.

We successfully reproduced the high-pitched soprano voice production with a finite difference model of vocal fold tissue vibration with string-like restoring forces. The two glottal configurations I and II were simulated with “weak” and “tight” vocal fold adduction, regulated via the pre-phonatory distance between the vocal processes of the arytenoid cartilages (with  $d=0.6$  mm and  $d=0.1$  mm for weak and tight adduction, respectively). Phonation

with weak adduction resulted in a posterior glottal gap (Fig. 6A,B), all other parameters being equal across the two conditions. The emerging  $f_0$  was 1,540 Hz and 1597 Hz for weak and tight adduction. Results showed that string-like vocal tissue layers (mucosa and ligament), both with a fiber stress of 0.9 MPa, produced self-sustained vocal fold oscillation, again corroborating the MEAD production mechanism. This was the case for both weak and tight adduction scenarios. With weak adduction, the larger time-varying glottal area (Fig. 6C) caused larger airflow rates that were non-zero in the “closed” phase (Fig. 6D). This resulted in a reduced strength of the second harmonic in the frequency spectrum (Fig. 6F), as compared to phonation with strong adduction (Fig. 6E).



**Figure 6.** Computer simulation with a low-order finite difference model of vocal fold tissue vibration. (A) and (B) two pre-phonatory glottal configurations, resembling strong and weak adduction; (C) and (D) resulting glottal area and glottal airflow for high-pitched simulations with both pre-phonatory glottal configurations; (E) and (F) normalized spectra of glottal airflow resulting from strong and weak adduction.

## Discussion

This study investigates high-pitched operatic phonation in the so-called whistle register above C6 using super high-speed laryngoscopy and computational modeling. Our data suggest that high-pitched soprano operatic singing is not produced by an aerodynamic whistle. Rather, we found medio-lateral vocal fold vibration synchronous to the variation of the radiated acoustic pressure in all investigated sopranos when phonating at  $f_0 \approx 1.6$  kHz. The observed medio-lateral vibratory component—in five out of nine singers resulting in full glottal closure—is a fundamental requirement for voice production according to the MEAD principle<sup>15</sup>. This medio-lateral vibratory component, resulting in a cyclical variation of the glottal area at the observed fundamental frequencies would be clearly detrimental to sound production with an aeroacoustic phenomenon. This is because a time-varying glottal area—causing time-varying airflow rates at the rate of the fundamental frequency—would introduce a considerable amount of frequency modulation (FM) into the putatively emerging aeroacoustic sound, thus violating the requirement to produce voice at quasi-stationary  $f_0$  conditions in artistic singing. Furthermore, the acoustic signals contained a well-defined harmonic structure, with the second harmonic having a sound level that is only about 20 dB lower than that of the fundamental. This clearly contrasts true aeroacoustic sound production like rodent ultrasonic vocalization<sup>40</sup> or human lip whistling<sup>41</sup>, where the second harmonic's level is 40 dB or more below that of the fundamental.

For these reasons, an aeroacoustic sound production phenomenon can be clearly ruled out for the investigated high-frequency operatic singing style. Consequently and in agreement to previous studies<sup>5,6,8,28,32–35</sup>, the frequently used term “whistle register”—while potentially applicable to other types of ultra-high-pitched voice production—does not reflect the physiologic voice production mechanism for classical/operatic singing at these high frequencies. Further, register names are commonly deducted rather from perceptive factors, such as “head register” and “chest register”, which are not necessarily scientifically appropriate but still established. However, it is precisely when using the term “whistle register” for the pitch range analyzed in the presented study that the underlying physiological principles should be correctly classified.

Female speech occurs at an average  $f_0$  of approximately 200 Hz<sup>1</sup>. The lowest pitches of adult females are found at about 135 Hz<sup>1</sup>, and the highest pitches of operatic soprano singing, investigated here, occur at about 1.57 kHz (and ca. 2 kHz in one case, see Suppl. Fig. S2), thus covering a range of almost four octaves. This is in good agreement with the predicted  $f_0$  range for different mammalian species at large<sup>42</sup>.

It is remarkable that the observed vibratory characteristics of the vocal folds documented here (recall Fig. 5B, showing median CQ values of 47.6%) closely resemble those reported for singing voice production in the M2 (“falsetto”) register found at relatively lower  $f_0$ . For M2 phonation, Henrich et al.<sup>43</sup> documented EGG contact quotients in the range of 5–50%, Herbst et al.<sup>44</sup> reported videokymographic closed quotients in the range of zero to about 50%, and Echternach et al.<sup>45</sup> reported high speed videolaryngoscopically derived closed quotients of the glottal area from 0 to 50%.

Furthermore, during the review process of the presented manuscript, Kato et al. published a quite comparable study in non-professional-singers subjects, analyzing pitches from C6 to A6, however using rigid transoral laryngoscopy during high speed recordings<sup>46</sup>. Although a transoral rigid endoscopy might have affected vocal fold tensions and vocal tract/voice source interactions, also these authors documented vocal fold oscillations in all of their 6 subjects.

Considering these findings, we conclude that the laryngeal vibratory phenomena of high-pitched operatic soprano singing are comparable to what is seen in the M2 mechanism. The auditory perceptual distinction of the investigated type of voice production (commonly termed the M3 mechanism) from the lower-pitched M2 vocal register is likely caused by influences of the vocal tract, as suggested by previous research<sup>47–50</sup>.

## Types of vocal fold closure

Five of the nine participants phonated with a fully adducted posterior glottis, and in three of these the vocal folds were partially in contact along the antero-posterior mid-line at the moment of maximum glottal opening. This might be indicative of a “damping” phenomenon in analogy to violin playing—shortening the vibrating portion of a violin string with finger pressure—as previously proposed by some authors<sup>51–54</sup>. In such a “damping” mechanism, control would be facilitated by adjustment (and, specifically, shortening) of the vocal fold portion that is in vibration, brought about by high degrees of vocal fold adduction and arytenoid compression. It is, however, unlikely that a medial (adductory) pressure can establish a fixed boundary without excessively constricting the entire glottis. Further, the “pinning” force would be in the wrong direction. It is not perpendicular to the vibration, as in pinning a violin string, but rather in the direction of motion. This would establish a “fuzzy” boundary point, unlikely to be controllable over a wide pitch range.

The damping concept was also not supported by our computational model. For a “damping” phenomenon to occur, the physical boundaries of vocal fold vibration would have to be varied in antero-posterior direction, which was not observed in vivo and could also not be reproduced in silico. Furthermore, contraction of the thyroarytenoid (TA) muscle, which may normally contribute to medial compression during voice production in M1 and M2<sup>2</sup>, does not have much effect when counteracted by extremely high ligament stiffness. We therefore propose the following alternative hypothesis: MEAD phonation at the investigated range requires unusually high activity in the cricothyroid (CT) muscle in order to influence the ligament stiffness that is required for the targeted  $f_0$ . However, all else being equal, such a maneuver may lead to abduction of the posterior glottis<sup>55</sup>, which is detrimental to voice source strength (recall Fig. 6E). Consequently, singers try to counteract this tendency to vocal fold abduction by increasing glottal closure through substantial degrees of laryngeal medialization (i. e., adduction of arytenoids, vocal folds, and ventricular folds—recall Suppl. Fig. S5,S6), in order to maximize the achievable degrees of glottal closure and thus increase the amplitudes of the higher harmonics of the voice source.

The observed glottal configurations (I, IIa, and IIb) were thus most likely the result of individual anatomical and physiological predisposition.

### Findings in relation to other types of high-pitched singing

Summarizing, our investigation was concerned with investigation of high-pitched operatic soprano singing. However, ultra-high-pitched singing of both female and male singers is also found in other singing styles. Specifically, some contemporary commercial music (CCM) singers regularly phonate in the  $f_o$  range of 2–3 kHz<sup>47</sup>. Furthermore, performers singing in the “extranormal voice” style have been reported to phonate at  $f_o$  of up to 6 kHz<sup>31</sup>, and it was speculated that this “M4” phonation would be produced with a “vortex whistle”<sup>56</sup>. Notably in this context, Tsai et al.<sup>26</sup> suggested a diffuser jet with periodic vorticity bursts in the larynx for phonation at  $f_o \approx 4$  kHz. That respective phonation was investigated with ultrasound Doppler imaging, revealing a vocal fold vibratory amplitude of 0.1 mm, i.e., only barely visible to the naked eye. This is in agreement with Di Corcia et al.<sup>28</sup>, who reported that their participants’ “stop closure whistle” was produced during total absence of a mucosal wave and thus vocal fold vibration. These findings suggest that further investigation of the ultra-high-pitched phonations of humans is required. While—based on our data and findings—we can conclusively show that high-pitched operatic soprano singing is produced according to the MEAD principle, we cannot altogether rule out an aeroacoustic production mechanism for other human singing styles, for example CCM, contemporary experimental music, non-professional classical singing, or folk styles, at higher fundamental frequencies. In this respect, in contrast to western classically trained singing, in CCM, amplification techniques are commonly used for high-pitched singing. If an aeroacoustic mechanism would be present, it could be expected that the radiated sound would be rather weak. However, this could be counteracted by amplification. If an aeroacoustic production mechanism could be empirically documented in non-classical high-pitched singing or any other human voice production, it would be of interest to explore in future experiments if it is possible to drive the vocal folds as a passive, coupled resonator with such a sound generating mechanism.

### Limitations

The transnasal endoscopic data acquisition might have caused some irritation for the participants and it cannot be excluded that it led to tightening reflexes or increased muscle tension. However, none of the participants reported such irritation and none of them canceled the ongoing procedure. The advantage over exclusively electroglottographic measurements is the detailed observation of the vocal folds’ configuration and movement. In addition, for very high pitches, the larynx is often raised and the vocal tract tube is narrowed. This might cause slipping of the electrodes and corruption of the signal. It could be of interest how singers estimate their voice production at such high pitches, i.e. if they use AA or MEAD and if there is a full closure for the MEAD. It is a limitation of the present study that no systematic evaluation of the participants’ proprioceptive feedback concerning register transitions and phonation mechanism was performed. Such systematization of proprioception linked to actual physiology could, however, be a subject of future research.

Finally, the acquired voice signals contained aliasing artifacts above 4 kHz. However, given that the investigated fundamental frequencies were typically well below 2000 kHz, it was safe to compute both the  $f_o$  and H1-H2 metrics of those signals (see Methods for an in-depth discussion).

### Materials and methods

All methods were carried out in accordance with relevant guidelines and regulations. Furthermore, all experimental protocols were approved by the Freiburg University Ethical Committee (Nr 380/12). Informed consent was obtained from all subjects.

### Participants and phonatory task

Following approval from the local ethical committee, nine female professional operatic soprano singers were investigated. All participants had at least 4 years of professional training in western classical singing. Limited biographic data and a taxonomic classification according to the scheme proposed by Bunch and Chapman<sup>57</sup> are provided in Table 1. At the time of the experiment none of the participants complained of any vocal symptoms, and vocal pathologies were excluded based on videostroboscopy and/or high-speed digital imaging evidence.

The participants were asked to sing an ascending major scale on the vowel [a:] from musical pitch C6 (ca. 1047 Hz) to G6 (ca. 1568 Hz), avoiding extensive vibrato. Each note should last approximately one second. As pointed out in the introduction, there is no consensus at which musical pitch the so-called whistle register would start. However, there seems to be a general agreement that the musical pitch G6 falls within the whistle register. This motivated the choice of the pitch range in the presented study. The vowel [a:] was chosen because it could be expected that above  $f_o$  of 700–800 Hz classical singers will only produce this vowel quality: When  $f_o$  reaches the center frequency of the lowest vocal tract resonance ( $f_{R1}$ ), singers tend to avoid a crossing of  $f_o$  and  $f_{R1}$ , thus raising  $f_{R1}$  as  $f_o$  further increases<sup>50,58</sup>. All except two participants (S6 and S9) successfully accomplished the phonatory task, while S6 and S9 could only reach the musical pitch F6 ( $f_o \approx 1397$  Hz).

### Data acquisition

Analogous to previous investigations<sup>45,59–61</sup>, High Speed Digital Videolaryngoscopy (HSV) was performed using trans-nasal endoscopy. A flexible endoscope (ENF GP, Fa. Olympus, Hamburg, Germany) was mounted on a 38 mm C-mount adapter (Karl Storz, Tuttlingen, Germany), connected to a Photron high-speed camera (Fastcam SA-X2, Photron, Tokyo, Japan) and a 300 W light source (Karl Storz, Tuttlingen, Germany) which was operated at a frame rate of 20,000 frames per second (fps) and a spatial resolution of 386 × 320 pixels.



Subject	Age	Taxonomy (Bunch–Chapman)	Highest pitch sung in study	Glottal configuration	Glottal configuration (detail)
S1	21	7.2	G6 (1568 Hz)	I	intercartilaginous posterior glottal gap (PGG), potentially reaching into membranous vocal fold portion
S2	29	7.1/4.1	G6 (1568 Hz)	I	PGG reaching into membranous vocal fold portion
S3	25	7.2	(B6 (1975 Hz) beside the protocol)	IIa	complete glottal closure, arytenoids and vocal processes pressed together, visible partial vocal fold contact along antero-posterior axis even in open phase
S4	27	7.2	G6 (1568 Hz)	IIb	complete glottal closure, arytenoids and vocal processes pressed together (but most posterior glottis not visible)
S5	36	3.1a, 4.5	G6 (1568 Hz)	IIa	complete glottal closure, arytenoids and vocal processes pressed together, visible partial vocal fold contact along antero-posterior axis even in open phase
S6	30	4.1/4.5	F6 (1397 Hz)	IIa	complete glottal closure, arytenoids and vocal processes pressed together, visible partial vocal fold contact along antero-posterior axis even in open phase
S7	31	3.1a, 3.1c	G6 (1568 Hz)	I	PGG, potentially reaching into membranous vocal fold portion
S8	46	3.1a, 3.1c	G6 (1568 Hz)	I	PGG
S9	23	7.2	F6 (1568 Hz)	IIb	complete glottal closure, arytenoids and vocal processes pressed together

**Table 1.** Overview of participants, indicating their proficiency level (taxonomy according to Bunch and Chapman<sup>57</sup>), investigated phonatory  $f_0$ , and observed glottal configuration.

Simultaneously with the HSV recordings, time-synchronous acoustic and electroglottographic (EGG) signals were captured with a National Instruments (Austin, Texas, USA) DAC USB-6251 interface at a sampling rate of 20,000 Hz. The DAC was automatically triggered (i.e., switched on and off) via a TTL signal emitted by the HSV camera. The acoustic signal was captured with an omnidirectional microphone (Behringer ECM-8000, Behringer, or Sennheiser ME 62, Sennheiser, Wedemark, Germany) and a Mackie 802VLZ4 (Bothell WA, USA) preamplifier. The EGG signal was acquired using a dual-channel EGG device (EG2-PCX2, Glottal Enterprises, Syracuse, NY, USA). Due to the lack of a hardware low-pass filter and because of the relatively low sampling frequency of 20,000 Hz (with a Nyquist frequency of 10,000 Hz), the acoustic signals contained mild traces of aliasing artifacts. Spectrographic inspection of these acoustic signals using a spectrogram dynamic range of 90 dB revealed that the aliasing artifacts were on average above 5267 Hz, with a standard deviation of about 900 Hz. This suggests that—while the spectral acoustic data needs to be considered with care—it was safe to utilize the acoustic signals for both fundamental frequency estimation and computation of the intensity relation of the first to the second harmonic found in the signals. That latter approach is justified because even for the highest investigated musical pitch (G6) at about 1568 Hz, the second harmonic was well below the lowest frequency where aliasing artifacts were observed in the acoustic signals.

In order to verify the accuracy of the signals' synchronization, a custom-made rotating disk with a printed black and white pattern was synchronously recorded with both HSV and a simple electric circuit containing a photodiode that monitored the rotating disk's light intensity. The output of the photodiode current was routed to the acoustic channel of the DAC. The digitally computed light intensity of the respective HSV recording was compared with the photodiode current variation seen in the DAC's input channel, and a perfect temporal agreement was found.

### HSV pre-processing and data analysis

The segmentation of the visible glottis and the medio-lateral vocal fold deflections—as documented by HSV—required several pre-processing steps that were accomplished through scripts implemented in the Matlab framework (R2014b, MathWorks Inc., Natick, MA, USA)<sup>45,59–61</sup>:

- In some cases, a honeycomb structure introduced by the endoscope optics was visible in the HSV recordings. This artifact was removed with a frequency-selective FFT-filter by transforming the images into the frequency domain via a 2-D discrete Fourier transform. Therein, the periodic noise appeared as 2 major peaks, apart from the center frequency peak. These two peaks were identified via an adaptable threshold binarization, and the areas were slightly increased via an opening filter and set to 0. The images were then transformed back into the image domain.
- Because the angle of the glottis could change with respect to the orientation of the HSV field of view during the recordings, the HSV footage had to be spatially rotated in order to align the glottis with the vertical dimension in the recordings. To this end, an approximate mask of the glottal opening was found in every image by means of time-difference images. In these ellipsoid-shaped masks, the orientation of the main axis of the glottis-opening could be calculated. Here, we assumed that the glottis shows the most movement between discrete images in the video. Hence, we calculated difference images from pictures with an adaptable time offset (typically set to 5 frames). Via various image processing methods, such as binarization, opening filter, and search for connected components, the largest object with the greatest movement was identified. The major axes of the resulting elliptical object were then determined, and the whole image

- could then be rotated so that the glottis was in vertical alignment. In most cases, some manual adjustments of the resulting angle graphs were necessary in order to remove sudden and improbable changes.
- (c) In a last pre-processing step, a bounding box was manually drawn around the glottis at key frames, and all images of a sequence were correspondingly cropped. As a result, the glottis now appeared both vertically and horizontally centered in every frame of the HSV sequence.

Glottis segmentation—and thus determination of the time-varying medio-lateral vocal fold displacement along the antero-posterior glottal dimension—was performed using the custom-made Glottis Analysis Tools (Denis Dubrovskiy and Michael Döllinger, Erlangen University, Germany), as described previously<sup>62</sup>.

For further analysis, the glottal area waveform (GAW) was computed, producing the time-varying visible area of the glottis, indicated in pixels. The glottal segmentation data was also utilized to compute phonovibrograms<sup>63</sup>, i. e., a visualization procedure that extracts vocal fold vibrations from HSV data and transfers the motion information into a set of displacement data for both the left and the right vocal fold<sup>63</sup>. The PVG information pertaining to the individual vocal folds was then combined to a glottovibrogram<sup>64</sup> (GVG), representing the time-varying glottal width in pixels along the antero-posterior glottal dimension. The GVG data was then used to compute the glottal closed quotients (i. e., the relative duration of vocal fold collision per vibratory cycle, expressed in percent) along the antero-posterior glottal axis. These last two processing steps, as well as the generation and assembly of all figures in this manuscript, were achieved with scripts written in the Python programming language.

### Analysis of voice signals

After recording and segmentation of the glottis from the HSV material all voice signals (GAW, EGG signal and the audio signal, respectively) were analyzed concerning  $f_0$  using an auto-correlation method within the custom made MultiSignalAnalyzer Software<sup>65</sup>. In order to avoid irregularities occurring during possible  $f_0$  transitions, only the stable part of each phonation was analyzed using a time window of 100 ms at the temporal midpoint of each phonatory (musical) pitch (midpoint  $\pm 50$  ms). The relative sound level difference between the first and the second harmonic (H1–H2, expressed in decibel (dB)<sup>66</sup>) was computed for all phonations at lowest and highest phonations, attempted at musical pitches C6 and G6. Because of some aliasing phenomena in the audio signals, calibration of the sound pressure level was considered problematic.

### Aeroacoustic model

For the aeroacoustic model used for computing the data shown in Fig. 2 we applied the following reasoning. When, in a hypothetical aeroacoustic sound source without vocal fold vibration, the glottal air flow separates from the glottis and a jet is formed, small instabilities in this glottal air jet can become entrained at certain frequencies due to a feedback loop between these downstream-traveling flow structures and acoustic waves traveling upstream<sup>20</sup>. For an impinging jet model, the emerging frequencies have been estimated with a previously<sup>20</sup> described model established by  $f_n = n \cdot u/x_{wall}$ , where is the  $f_n$  frequency of the  $n^{\text{th}}$  possible whistle frequency,  $n$  is the mode number,  $x_{wall}$  is the jet length, and  $u$  is the mean convection speed of downstream moving coherent structures (i.e., the air flow speed). The mean convection  $u$  is approximated by  $u = V/A_{gl}$ , where  $V$  is the volumetric air flow rate and  $A$  is the glottal constriction area. For the lowest possible frequency (mode-1), the model is reduced to  $f_{mode-1} = u/x_{wall}$ .

### Computer simulation

A finite difference model of vocal fold tissue vibration was used to generate a high  $f_0$  with string-like restoring forces. However, a single string is insufficient to produce self-sustained oscillation with air passing over its surface. It requires multiple coupled strings with ribbon-like flexing of the medial surface of the vocal folds. Alternating convergent and divergent glottal shapes can then produce an aerodynamic push–pull on the vocal folds for sustained oscillation<sup>67</sup>. The vocal folds also have tissue depth in the medial–lateral direction that allows edge movement but restrains movement into the deep muscular layer. A rectangular parallelepiped with 90 coupled masses was sufficient to meet the boundary conditions and the tissue properties. The vocal fold length was 0.945 cm, the thickness 0.3 cm, and the depth 0.45 cm. In the anterior–posterior direction, 5 masses allowed string-like motion with fixed boundary conditions at both ends. Along the vocal fold thickness, 3 masses provided the ribbon-like flexure and bi-stable nature of voice registration<sup>68</sup>, and 6 masses were used laterally to allow vibration to dissipate to zero. The masses were coupled with fiber stresses of 0.9 MPa along the vocal fold length in the first two medial–lateral layers that represented the mucosa and ligament, respectively. In the 3–6 medial–lateral layers, a muscle fiber stress 5 kPa was selected, a value in the mid-range of measured thyroarytenoid muscle stress<sup>69</sup>. For shear coupling between the fibers, a gel shear modulus of 1.0 kPa was chosen according to measurement<sup>70</sup>. The damping ratio for the vibrating tissue needed to be 0.04, lower than the 0.1 value typically chosen for speech-like fundamental frequencies<sup>71</sup>. With this low damping ratio, it required a 4 kPa lung pressure to obtain self-sustained oscillation. For the 90 masses, 180 first-order differential equations were solved with a 4-th order Runge–Kutta solver<sup>14</sup>. With a simple string formula based on ligament stress and vocal fold length, the natural frequency of oscillation was predicted to be 1587 Hz.

Vocal fold adduction was controlled with one variable, the superior-posterior glottal width, which is the distance between the vocal processes of the arytenoid cartilages (i.e., the posterior cartilaginous boundary of the membranous vocal fold portion). This width was chosen to be 0.1 mm for tight adduction and 0.6 mm for weak adduction. The width varied linearly to zero at the anterior commissure. However, the glottal width did not vary vertically along the thickness of the vocal folds. In other words, the pre-phonatory glottis was neither convergent nor divergent, but rectangular.

The aero-acoustic solution was obtained for an [a:] vowel with a simplified Navier–Stokes approach as described in<sup>68</sup>. The airway geometry, from the trachea to the lips, was taken from MRI data obtained by Story et al.<sup>72</sup>.

### Ethical votum

Freiburg University 380/12.

### Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 1 August 2023; Accepted: 20 May 2024

Published online: 07 June 2024

### References

1. Baken, R. J. & Orlikoff, R. F. *Clinical Measurement of Speech and Voice*. (Singular Thomson Learning, 2000).
2. Herbst, C. T. The snake pit of voice pedagogy part I: Proprioception, perception, and laryngeal mechanisms. *J. Sing.* **77**, 173–188 (2020).
3. Echternach, M. Vocal registers. In *European Manual of Medicine* (eds Zehnhoff-Dinnesen, A. et al.) (Springer Verlag, 2020).
4. Walker, J. S. An investigation of the whistle register in the female voice. *J. Voice* **2**, 140–150 (1988).
5. Miller, D. G. & Schutte, H. K. Physical definition of the ‘flageolet register’. *J. Voice* **7**, 206–212 (1993).
6. Keilmann, A. & Michek, F. Physiologie und akustische Analysen der Pfeifstimme der Frau. *Folia Phoniatr. Et Logop* **45**, 247–255 (1993).
7. Kiesgen, P. Voice Pedagogy: Registration. *J. Sing* **62**, 537–539 (2006).
8. Garnier, M. et al. Glottal behavior in the high soprano range and the transition to the whistle register. *J. Acoust. Soc. Am.* **131**, 951–962 (2012).
9. Titze, I. *Principles of voice production* (Prentice Hall, 1994).
10. Henrich, N. Mirroring the voice from Garcia to the present day: Some insights into singing voice registers. *Logop. Phoniatr. Vocology* **31**, 3–14 (2006).
11. Herbst, C. T. 2021 Registers—The snake pit of voice pedagogy. Part 2: mixed voice, vocal tract influences, Individual Teaching Systems. *J. Sing* **77**, 345 (2021).
12. Van den Berg, J. Myoelastic-aerodynamic theory of voice production. *J. Speech Hear. Res.* **1**, 227–244 (1958).
13. Herbst, C. T., Elemans, C. P. H., Tokuda, I. T., Chatziioannou, V. & Švec, J. G. Dynamic system coupling in voice production. *J. Voice* <https://doi.org/10.1016/J.JVOICE.2022.10.004> (2023).
14. Titze, I. R. *The Myoelastic Aerodynamic Theory of Phonation*. (The National Center for Voice and Speech, 2006).
15. Švec, J. G., Schutte, H. K., Chen, C. J. & Titze, I. R. integrative insights into the Myoelastic-Aerodynamic theory and acoustics of phonation. Scientific tribute to Donald G. Miller. *J. Voice* **37**, 305–313 (2023).
16. Schoder, S. et al. Aeroacoustic sound source characterization of the human voice production—Perturbed convective wave equation. *Appl. Sci.* **11**, 2614 (2021).
17. Fant, G. Glottal source and excitation analysis. *STL-QPSR* **1**, 85–107 (1979).
18. Roberts, L. H. The rodent ultrasound production mechanism. *Ultrasonics* **13**, 83–88 (1975).
19. Mahrt, E., Agarwal, A., Perkel, D., Portfors, C. & Elemans, C. P. H. Mice produce ultrasonic vocalizations by intra-laryngeal planar impinging jets. *Curr. Biol.* **26**, R880–R881 (2016).
20. Håkansson, J. et al. Aerodynamics and motor control of ultrasonic vocalizations for social communication in mice and rats. *BMC Biol.* <https://doi.org/10.1186/s12915-021-01185-z> (2022).
21. Riede, T., Borgard, H. L. & Pasch, B. Laryngeal airway reconstruction indicates that rodent ultrasonic vocalizations are produced by an edge-tone mechanism. *R Soc. Open Sci.* **4**, 170976 (2017).
22. Klatt, D. & Klatt, L. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* **87**(2), 820–857. <https://doi.org/10.1121/1.398894> (1990).
23. Harrison, D. F. N. *The anatomy and physiology of the mammalian larynx* (Cambridge University Press, 1995).
24. Schultz, P. Über einen Fall von willkürlichem laryngealen Pfeifen beim Menschen. *Arch. f. Physiol., Arch. f. Physiol. Suppl.* **523**, (1902).
25. Berry, D. A., Herzel, H., Titze, I. R. & Story, B. H. Bifurcations in excised larynx experiments. *J. Voice* **10**, 129–138 (1996).
26. Tsai, C. G., Shau, Y. W., Liu, H. M. & Hsiao, T. Y. Laryngeal mechanisms during human 4-kHz vocalization studied with CT videostroboscopy and color Doppler imaging. *J. Voice* **22**(3), 275–82 (2008).
27. Edgerton, M. The extra-normal voice. In *The Oxford Handbook of Singing* (eds Howard, D. M. et al.) 729–750 (Oxford University Press, 2013). <https://doi.org/10.1093/oxfordhb/9780199660773.013.007>.
28. Corcia, A. D. & Fussi, F. Whistle register and M3: A preliminary HSDI investigation by visualization and acoustics in male and female singers. In *Normal and abnormal vocal folds kinematics: high speed digital phonoscopy (HSDP), optical coherence tomography (OCT) & narrow imaging* (ed. Izdebski, K.) 267–272 (CreateSpace Independent Publishing Platform, 2016).
29. Herzel, H. & Reuter, R. Whistle register and biphonation in a child’s voice. *Folia Phoniatr Logop* **49**, 216–224 (1997).
30. Döllinger, M., Berry, D. A., Luegmair, G., Hüttner, B. & Bohr, C. Effects of the Epilarynx area on vocal fold dynamics and the primary voice signal. *J. Voice* **26**, 285–292 (2012).
31. Edgerton, M. E. et al. Pitch profile of the glottal whistle (M4). *Malays. J. Sci.* **32**, 78–85 (2013).
32. Švec, J. G., Sundberg, J. & Hertegård, S. Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography. *J. Acoust. Soc. Am.* **123**, 347–353 (2008).
33. Echternach, M., Döllinger, M., Sundberg, J., Traser, L. & Richter, B. Vocal fold vibrations at high soprano fundamental frequencies. *J. Acoust. Soc. Am.* **133**, EL82–EL87 (2013).
34. Izdebski, K., Di Lorenzo, E., Yan, Y. & Blanco, M. What we have learned about ingressive phonation and whistle voice production from HSDP. In *Normal and Abnormal Vocal Folds Kinematics: High Speed Digital Phonoscopy (HSDP), Optical Coherence Tomography (OCT) & Narrow Imaging* 301–306 (ephonoscope, 2016).
35. Sakakibara, K.-I. Production mechanism of voice quality in singing. *J. Phon. Soc. Jpn.* **7**, 27–39 (2003).
36. Lã, F. M. B., Sundberg, J. & Granqvist, S. Augmented visual-feedback of airflow: Immediate effects on voice-source characteristics of students of singing. *Psychol. Music* **50**, 933–944 (2022).
37. Miller, D. G. Registers in Singing. Empirical and Systematic Studies in the Theory of the Singing Voice. *Doctoral Dissertation* (University, 2000).

38. Childers, D. G., Naik, J. M., Larar, J. N., Krishnamurthy, A. K. & Moore, G. P. Electroglottography, speech, and ultra-high speed cinematography. In *Vocal Fold Physiology and Biophysics of Voice 202–220* (eds Titze, I. R. & Scherer, R.) (Denver Center of Performing Arts, 1983).
39. Lohscheller, J., Švec, J. G. & Döllinger, M. Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects. *Logoped. Phoniatr. Vocol.* **38**, 182–192 (2013).
40. Brudzynski, S. M., Kehoe, P. & Callahan, M. Sonographic structure of isolation-induced ultrasonic calls of rat pups. *Dev Psychobiol* **34**(3), 195–204 (1999).
41. Shadle, C. Experiments on the acoustics of whistling. *Phys. Teach.* **21**(3), 148–154 (1983).
42. Titze, I., Riede, T. & Mau, T. Predicting achievable fundamental frequency ranges in vocalization across species. *PLoS Comput. Biol.* **12**(6), e1004907 (2016).
43. Henrich, N., d'Alessandro, C., Doval, B. & Castellengo, M. Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *J. Acoust. Soc. Am.* **117**(3), 1417–1430 (2005).
44. Herbst, C. T., Schutte, H. K., Bowling, D. L. & Švec, J. G. Comparing chalk with cheese—The EGG contact quotient is only a limited surrogate of the closed quotient. *J. Voice* **31**(4), 401–409. <https://doi.org/10.1016/j.jvoice.2016.11.007> (2017).
45. Echternach, M. *et al.* Laryngeal evidence for the first and second passaggio in professionally trained sopranos. *PLoS One* **12**(5), e0175865 (2017).
46. Kato, H., Lee, Y., Wakamiya, K., Nakagawa, T. & Kaburagi, T. Vocal fold vibration of the whistle register observed by high-speed digital imaging. *J. Voice* <https://doi.org/10.1016/j.jvoice.2023.08.026> (2023).
47. Titze, I. R. A hypothesis about whistle voice. *J. Sing.* **64**, 473–475 (2008).
48. Garnier, M., Henrich, N., Smith, J. & Wolfe, J. Vocal tract adjustments in the high soprano range. *J. Acoust. Soc. Am.* **127**, 3771–3780 (2010).
49. Echternach, M. *et al.* Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies. *J. Acoust. Soc. Am.* **137**, 2586–2595 (2015).
50. Köberlein, M. *et al.* Investigation of resonance strategies of high pitch singing sopranos using dynamic three-dimensional magnetic resonance imaging. *J. Acoust. Soc. Am.* **150**, 4191–4202 (2021).
51. Pressman, J. J. & Kelemen, G. Physiology of the larynx. *Physiol. Rev.* **35**, 506–554 (1955).
52. Van den Berg, J. Vocal ligaments versus registers. *The NATS Bulletin* 16–31 (1963).
53. Titze, I. R. & Hunter, E. J. Normal vibration frequencies of the vocal ligament. *J. Acoust. Soc. Am.* **115**, 2264–2269 (2004).
54. Thurman, L., Welch, G. F., Theimer, A. & Klitzke, C. Addressing Vocal Register Discrepancies: An Alternative, Science-Based Theory Of Register Phenomena. in *Second International Conference The Physiology and Acoustics of Singing (PAS2)* 64 (National Center for Voice and Speech, 2004).
55. Baken, R. J. & Isshiki, N. Arytenoid displacement by simulated intrinsic muscle contraction. *Folia Phoniatr (Basel)* **29**, 206–216 (1977).
56. Edgerton, M. E. *The 21st-century voice : Contemporary and traditional extra-normal voice* (Scarecrow Press, 2015).
57. Bunch, M. & Chapman, J. Taxonomy of singers used as subjects in scientific research. *J. Voice* **14**, 363–369 (2000).
58. Joliveau, E., Smith, J. & Wolfe, J. Tuning of vocal tract resonance by sopranos. *Nature* **427**(6970), 116–116 (2004).
59. Echternach, M. *et al.* The influence of vowels on vocal fold dynamics in the tenor's passaggio. *J. Voice* **31**(4), 424–429 (2017).
60. Echternach, M., Döllinger, M., Köberlein, M., Kuranova, L. & Kainz, M. A. Influence of loudness on vocal stability in the male Passaggio. *J. Voice* **37**(2), 296.e1–296.e8 (2021).
61. Echternach, M. *et al.* Are source-filter interactions detectable in classical singing during vowel glides?. *J. Acoust. Soc. Am.* **149**(6), 4565 (2021).
62. Inwald, E. C., Döllinger, M., Schuster, M., Eysholdt, U. & Bohr, C. Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging. *J. Voice* **25**, 576–590 (2011).
63. Lohscheller, J. & Eysholdt, U. Phonovibrogram visualization of entire vocal fold dynamics. *Laryngoscope* **118**, 753–758 (2008).
64. Karakozoglou, S.-Z., Henrich, N., d'Alessandro, C. & Stylianou, Y. Automatic glottal segmentation using local-based active contours and application to glottovibrometry. *Speech Commun.* **54**, 641–654 (2012).
65. Schäffner, F. D. *Signalerfassung und Analyse medizinischer Daten: Entwicklung einer Analysesoftware für medizinische Messdaten innerhalb der Stimmforschung* (University Erlangen, 2015).
66. Sundberg, J., Andersson, M. & Hultqvist, C. Effects of subglottal pressure variation on professional baritone singers' voice sources. *J. Acoust. Soc. Am.* **105**(3), 1965–1971 (1999).
67. Titze, I. R. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* **83**, 1536–1552 (1988).
68. Titze, I. R. Bi-stable vocal fold adduction: A mechanism of modal-falsetto register shifts and mixed registration. *J. Acoust. Soc. Am.* **135**, 2091–2101 (2014).
69. Alipour-Haghighi, F., Titze, I. R. & Perlman, A. L. Tetanic contraction in vocal fold muscle. *J. Speech Hear. Res.* **32**, 226–231 (1989).
70. Chan, R. W. & Titze, I. R. Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling. *J. Acoust. Soc. Am.* **107**, 565–580 (2000).
71. Ishizaka, K. & Flanagan, J. L. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst. Tech. J.* **51**, 1233–1268 (1972).
72. Story, B. H. & Titze, I. R. Parametrization of vocal tract areafunctions by empirical orthogonal modes. *J. Phonetics* **26**, 223–260 (1998).

## Acknowledgements

The authors acknowledge a NovoNordisk Interdisciplinary Synergy grant (to CPHE), and grant # Ec409/1-1 and 1-4 by the German Research Foundation, DFG (to ME) and National Institutes of Health, Grant number 1R01 DC 018280 – 05 (to IRT). A pre-print version of this article was published on August 11th 2023 on ResearchSquare, DOI: <https://doi.org/10.21203/rs.3.rs-3222892/v1>.

## Author contributions

M.E.: Conceptualization, Methodology, Performing Experiments, Formal analysis, Funding, Writing—Original Draft F.B.: Methodology, Software, Formal analysis, Performing Experiments, Writing—Editing M.K.: Formal analysis, Data Curation, Writing—Editing M.D.: Methodology, Software, Writing—Editing M.B.: Methodology, Software B.R.: Conceptualization, Writing—Editing I.R.T.: Methodology, Software, Formal analysis, Performing Experiments C.P.H.E.: Conceptualization, Writing—Editing C.T.H.: Conceptualization, Methodology, Software, Formal analysis, Performing Experiments, Writing—Original Draft.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-62598-8>.

**Correspondence** and requests for materials should be addressed to M.E. or C.T.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024