Physics of Fluids ARTICLE pubs.aip.org/aip/pof

Evaluating the accuracy of one-dimensional glottal flow model in predicting voice production: Comparison to experiments and three-dimensional flow simulations

Cite as: Phys. Fluids **37**, 116113 (2025); doi: 10.1063/5.0292598
Submitted: 23 July 2025 · Accepted: 15 October 2025 ·
Published Online: 10 November 2025

View Online Export Citation CrossMark

Tsukasa Yoshinaga[1,a] and Zhaoyan Zhang[2]

AFFILIATIONS

[1]Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan
[2]School of Medicine, University of California, Los Angeles, Los Angeles, California 90095-1794, USA

Note: This paper is part of the Special Topic, Flow and Phonation.
[a]**Author to whom correspondence should be addressed:** yoshinaga.tsukasa.es@osaka-u.ac.jp

ABSTRACT

The glottal flow is often simplified as one-dimensional (1D) in phonation models to reduce computational cost. Although previous studies showed that a 1D flow model can predict voice production by a three-dimensional (3D) flow combined with a simplified two-mass vocal fold model, its validity in voice production involving more realistic 3D vibrations remains unclear. The goal of this study is to investigate the accuracy of the 1D flow model in predicting vocal fold vibration and voice production in a vocal fold model exhibiting a more realistic 3D vibration pattern, by comparing its prediction to that from a mechanical experiment and a 3D Navier–Stokes compressible flow model. The results showed that the 1D flow model predicted overall vibratory pattern similar to that observed in experiment and simulations based on the 3D flow model. However, the 1D flow model predicted slightly larger displacements and greater glottal flow fluctuations than the 3D flow model. The 3D flow model revealed strong variations in surface pressure along the anterior–posterior direction, particularly during the closing phase, which was not captured by the 1D flow model. Despite these differences, the 1D flow model adequately reproduced major aerodynamic and vibratory features under typical normal phonatory conditions, supporting its use in phonation models for efficient voice simulations.

I. INTRODUCTION

Human phonation is produced by self-sustained oscillations of the vocal folds excited by airflow from the lungs. Vocal fold oscillation repeatedly opens and closes the glottis and interrupts the airflow through the glottis, which is crucial for producing the rich harmonic characteristics of the human voice.[1] It is also known that the airflow surrounding the vocal folds exhibits complex three-dimensional (3D) phenomena,[2] including jet flapping, vortex shedding from the glottal jet, and vortex propagation toward the vocal tract.

When modeling such airflow in numerical simulation, one must consider that the glottal opening is on the order of 1 mm, and the resulting jet flow can reach peak velocities[2] of approximately 40 m/s. Accurately resolving such flow features requires fine spatial grids, which are computationally expensive. As a result, many previous studies have employed simplified one-dimensional (1D) flow models to reduce computational cost.[3–14] The 1D flow model predicts the intraglottal air pressure based on Bernoulli's equation up to the point where the glottal flow separates from the vocal fold surface. Early studies often assumed flow separation at the glottis exit to simply the calculation.[3–6] More refined 1D flow models have also been proposed, in which boundary layer equations are solved to better capture flow separation point.[7,8] Two-dimensional (2 D) Navier–Stokes based simulations have been conducted to compare the flow-induced pressure distribution on the vocal fold surface[15] and phonation threshold pressure with those predicted by 1D flow models.[16] Furthermore, comparisons between 1D flow models and experimental setups featuring realistic 3D flow structures have provided insights into the applicability and limitations of 1D flow models .[17,18]

More recent studies aimed to numerically capture the full 3D flow field.[19–22] Due to substantial computational costs, most of these

simulations assumed an incompressible glottal flow. To estimate the resulting sound, acoustic wave propagation was separately calculated from the aeroacoustic source in the flow field. Thus, the potential effects of vocal tract acoustics on vocal fold vibration and the glottal flow cannot be investigated in these studies.

In our recent studies, we developed a computational model that combines a 3D compressible fluid simulation with an immersed boundary method, enabling us to investigate the relationship between vocal fold vibration and the generation of aeroacoustic sound.[23–25] Using a two-mass model of the vocal folds, we have demonstrated that the 1D flow model can reproduce left–right asymmetric vibrations in the 3D flow under certain asymmetric vocal fold conditions.[23] However, because of the simplified vocal fold dynamics in the two-mass model, which, while capable of capturing vertical phase differences, assumes a uniform shape along the anterior–posterior direction, it remains unclear how accurately the 1D flow model can predict vocal fold vibration and voice production in a more realistic 3D vocal fold model. The actual vocal folds often show greater vibration amplitude near the mid-membranous region than at the anterior or posterior ends, resulting in airflow that not only moves vertically but also along the anterior–posterior axis.[26] This leads to a more complex pressure distribution on the vocal fold surface,[27] which may in turn influence the dynamics of vocal fold vibration.

In this study, we aim to evaluate the accuracy of a 1D flow model coupled with a 3D continuum vocal fold model by comparing its predictions to those observed from both mechanical experiment and simulations solving the 3D compressible Navier–Stokes equations. To enable acoustic simulation in the 1D flow, we adopt a simplified version of the equivalent-circuit model of the vocal tract originally proposed by Ishizaka and Flanagan.[3] The approach of coupling a 1D flow model to a 3D vocal fold model has been increasingly adopted in many recent studies,[12,13] highlighting the need to better understand the accuracy of this approach in predicting voice production. In this study, we aim to clarify to what extent 1D models reproduce the surface pressure distribution and the resulting vibration dynamics, by comparing them with the 3D flow simulations.

II. METHODS

A. Vocal fold model

The three-dimensional vibratory dynamics of the vocal folds were simulated using transient response analysis based on mode superposition. This approach significantly reduces computational cost by considering only low-order vibration modes, while still capturing the characteristic vibratory behavior of the vocal folds.[28]

The vocal folds were modeled as a linear viscoelastic continuum and discretized using the finite element method (FEM). The equation of motion for each node can be expressed as

$$\mathbf{M}\ddot{\boldsymbol{\delta}} + \mathbf{C}\dot{\boldsymbol{\delta}} + \mathbf{K}\boldsymbol{\delta} = \mathbf{F}, \tag{1}$$

where, **M**, **C**, and, **K** are mass, damping, and stiffness matrices, respectively, and $\boldsymbol{\delta}$ is the nodal displacement vector. The external force vector **F** includes aerodynamic pressure and contact forces between the vocal folds.

In the mode superposition, the nodal displacement $\boldsymbol{\delta}$ is represented as a linear combination of eigenmode $\boldsymbol{\phi}_i$ and modal displacements q_i in the modal space:

$$\boldsymbol{\delta} = \sum_{i=1}^{N} \boldsymbol{\phi}_i q_i. \tag{2}$$

Here, i denotes the mode number, and the summation is taken up to the Nth mode. The eigenmodes $\boldsymbol{\phi}_i$ and the corresponding natural angular frequencies ω_i are obtained by solving the characteristic equation derived from the undamped, unforced (*in vacuo*) system equation:

$$-\omega^2\mathbf{M} + \mathbf{K} = 0. \tag{3}$$

Solving this eigenvalue problem yields the set of mode shapes $\boldsymbol{\phi}_i$ and their associated frequencies ω_i. In general, high-order modes typically lie beyond the frequency range of interest and have negligible influence on voice production. Thus, only a limited number of low-order modes are retained to reduce the number of degrees of freedom in the simulation. Following previous work,[28] we considered up to the 20th mode in this study, which is the minimal number of modes required to obtain reasonable predictions, resulting in eigenmodes up to 200 Hz being included in the simulation. However, more recent studies[29] have reported that considering up to 100 modes may provide a more accurate representation of the detailed vibratory characteristics of the vocal folds (see details in Appendix A).

By expressing the displacement vector as a superposition of eigenmodes, the equation of motion in Eq. (1) can be diagonalized due to the orthogonality of the eigenmode matrix, yielding

$$m_i\ddot{q}_i + c_i\dot{q}_i + k_i q_i = f_i. \tag{4}$$

Here, m_i, c_i, and k_i are the modal mass, modal damping, and modal stiffness, respectively, and the external force is given by $f_i = \boldsymbol{\phi}_i^T F$. The angular frequency satisfies the relationship $k_i = \omega_i^2 m_i$. Normalizing each eigenmode with respect to the modal mass as $\tilde{\boldsymbol{\phi}}_i = \sqrt{m_i}\boldsymbol{\phi}_i$, the force term as $\tilde{f}_i = \tilde{\phi}_i^T F$, and defining the modal damping ratio as $\zeta_i = c_i/\left(2\sqrt{m_i k_i}\right)$, Eq. (4) becomes

$$\ddot{q}_i + 2\omega_i\zeta_i\dot{q}_i + \omega_i q_i = \tilde{f}_i. \tag{5}$$

By specifying the damping ratio ζ_i, the modal displacement q_i can be computed for each mode. The nodal displacement at each time step can then be reconstructed using Eq. (2) by superposing the eigenmodes. It should be noted that this method assumes linear superposition of modal displacements, and therefore does not account for nonlinear modal interactions that may occur during large-amplitude vibrations. However, nonlinear effects induced by external forces, such as subharmonics and transition to chaos, can still be represented.[29] In this study, we assumed a constant damping ratio of $\zeta = 0.06$ for all modes and performed time integration of Eq. (5) using a fourth-order Runge–Kutta method.

The geometry of the vocal folds and vocal tract was simplified as shown in Fig. 1. In this study, the x, y, and z axes corresponded to the inferior–superior, medial–lateral, and anterior–posterior directions. The vocal tract and subglottal airway were modeled as rectangular ducts. Their cross-sectional area was set to $17 \times 16.8\,\mathrm{mm}^2$, and their lengths were 175 and 150 mm, respectively, based on typical adult male anatomy.[30] An inlet chamber, representing the lungs, was placed upstream of the subglottal airway, and the steady inflow conditions were applied at the entrance.

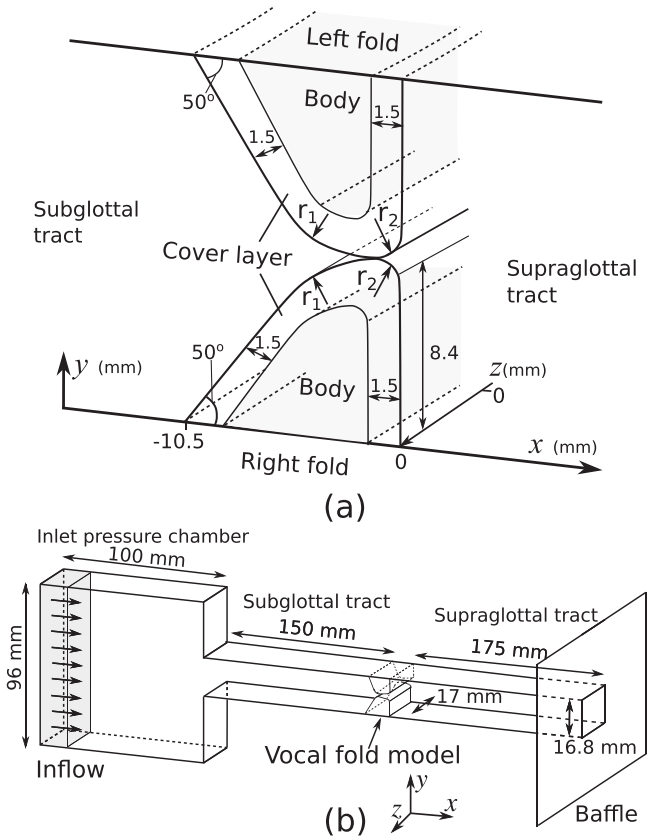

FIG. 1. Vocal fold model and vocal tracts. (a) The vocal fold geometry was based on M5 model.[31] (b) The vocal fold model was placed at the middle between the subglottal tract and supraglottal (vocal) tract. The *x*, *y*, and *z* axes corresponded to the inferior–superior, medial–lateral, and anterior–posterior directions, respectively.

The vocal fold geometry is based on the widely used M5 model.[31] The geometric parameters r_1 and r_2, as shown in the figure, were set to 6 and 0.987 mm, respectively, following the experimental setup.[32] The width of the vocal fold at the lateral wall was 10.5 mm. The glottal length L_g was set to 17 mm, corresponding to the typical length in adult males.[30] The vocal fold was placed between the vocal tract and the subglottal airway. The lateral and anterior–posterior surfaces that were attached to the walls were modeled as fully fixed.

The material properties and geometrical parameters used in this study are summarized in Table I. The stiffness and mass matrices were computed based on experimentally measured material properties (Young's modulus, density, and Poisson's ratio) of the silicone resin used in previous experiments.[33] The eigenvalue analysis was conducted using COMSOL Multiphysics 6.1 (COMSOL Inc.). The vocal fold geometry was discretized using 5650 finite elements. The contact force between the vocal folds was computed using the penalty method.[34]

B. 3D flow model

In the 3D flow model, the airflow around the vocal folds was predicted by solving the Navier–Stokes equations. To account for both flow and sound generation, we solved the governing equations for a compressible fluid using the finite-difference method. To handle the movement of the vocal fold surface within a structured grid, the immersed boundary method was employed. Specifically, we adopted the volume penalization (VP) method,[35] in which a penalty term was added as an external force to model the effect of vocal fold surface on the glottal flow.

TABLE I. The material properties and geometrical parameters for the vocal fold model.

Young's modulus (kPa)	Body layer	10.4
	Cover layer	4.9
Density (g/cm^3)		1.07
Poisson's ratio		0.49
Damping coefficient ζ		0.06
Glottal length L_g (mm)		17
Vocal fold lateral thickness D (mm)		10.5

The governing equations are given by

$$\dot{\mathbf{Q}} + \frac{\partial}{\partial x}(\mathbf{F} - \mathbf{F}_\nu) + \frac{\partial}{\partial y}(\mathbf{G} - \mathbf{G}_\nu) + \frac{\partial}{\partial z}(\mathbf{H} - \mathbf{H}_\nu) = \mathbf{V}, \quad (6)$$

$$\mathbf{V} = -(1/\phi - 1)\chi \begin{pmatrix} \partial \rho u_i / \partial x_i \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (7)$$

where $\mathbf{Q}$ represents the conservative variables, $\mathbf{F}, \mathbf{G}, \mathbf{H}$ are the inviscid flux vectors, and $\mathbf{F}_\nu, \mathbf{G}_\nu, \mathbf{H}_\nu$ are the viscous flux vectors. The external force $\mathbf{V}$ corresponds to the penalty term in the VP method, and a mask function χ was used to distinguish between fluid and wall regions. x_i denotes $x_i = (x, y, z)$. In regions where $\chi = 1$, the medium was treated as a porous medium. By setting porosity at $\phi = 0.25$, the acoustic reflectivity of the walls was confirmed to be above 99%.[36]

The mask function near the moving wall was computed as

$$\chi = \begin{cases} 1 & (\text{inside object}), \\ |d/\Delta \mathbf{x}| & (\text{moving boundary}), \\ 0 & (\text{outside object}), \end{cases} \quad (8)$$

where d denotes the distance from the wall surface to the computational grid point, and $\Delta \mathbf{x}$ is the grid spacing. By defining the mask function to vary smoothly with distance from the wall, the wall surface can move smoothly across the grid points.[37]

To accurately compute both the airflow and acoustic pressure fields, a sixth-order accuracy compact scheme was used for spatial differentiation. In addition, a tenth-order accuracy spatial filter was applied to perform implicit large eddy simulation (LES), which models the viscosity of turbulent vortices at the subgrid scale.[38] For time integration, a third-order accuracy Runge–Kutta method was employed. Details of the methodology are described in the previous study.[39]

Since the governing equations for airflow were solved using the finite difference method, a structured computational grid was constructed throughout the computational domain. Figure 2 shows the computational mesh near the vocal folds. The minimum grid spacing in this region was 0.025 mm, which has been shown to be sufficient for resolving the airflow near the vocal folds in a prior grid convergence

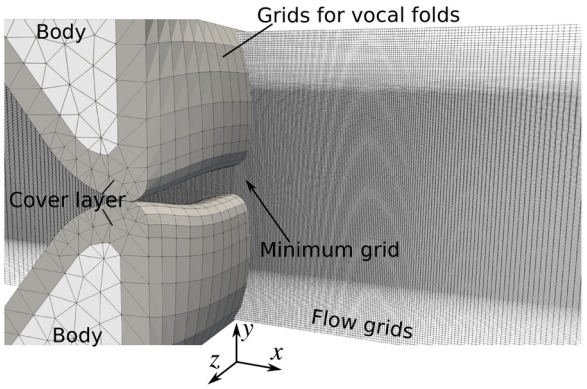

FIG. 2. Computational grids for the 3D flow model.

study.[23] The grid spacing was increased with distance from the vocal folds to reduce computational cost. The total number of grid points used in the fluid simulation was approximately 1.55×10^8. The mesh used in the finite element modeling of the vocal folds was relatively coarse compared to that of the fluid simulation to reduce computational costs. Nevertheless, we confirmed that this mesh resolution had negligible effect on the overall vibratory characteristics (see Appendix B).

Figure 3 illustrates the boundary conditions used in the 3D flow model. In the inlet chamber simulating the lungs, a constant pressure condition was applied, with $p_{\text{lung}} = 1000$ Pa simulating normal phonation and $p_{\text{lung}} = 1800$ Pa simulating loud phonation.[40] While the glottal flow region had a relatively fine grid resolution, the vocal tract region had a coarser grid to reduce computational cost. Nevertheless, the maximum grid spacing in the vocal tract was set to 4.86 mm so that sound wave propagation up to approximately 10 kHz can be resolved with at least eight points per wavelength.

A buffer zone was placed outside the acoustic region to prevent reflected waves from reentering the vocal tract. In this zone, the grid

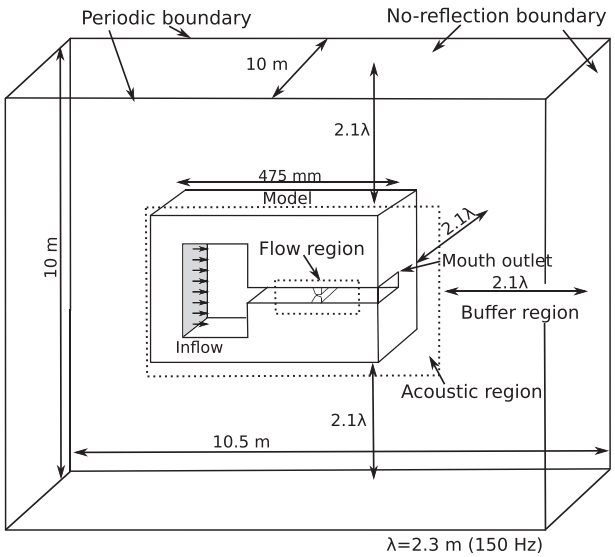

FIG. 3. Boundary conditions.

spacing was further increased, and a damping force was applied to attenuate the outgoing waves.[41] At the outermost boundaries of the domain, non-reflecting and periodic boundary conditions were imposed.

At each time step, the pressure on the vocal fold surface was used to compute the aerodynamic force. Based on this force, the vocal fold displacement and velocity were updated using Eqs. (2) and (5). These updated surface displacement and velocity values were then imposed on the flow field. The time integration was carried out by repeating this procedure. The time step for both the flow and vocal fold oscillation solvers was set to 0.25×10^{-7} s, for both the flow and vocal fold models, ensuring that the Courant number remained below 0.4 based on the speed of sound at 20 °C (=343 m/s).

C. 1D flow model

The 1D flow model was computed using an equivalent electrical circuit model, as described in detail by Ishizaka and Flanagan,[3] in which components such as resistors, inductors, and capacitors are used to represent the flow resistance, inertia, and compressibility, respectively. The cross-sectional area A, circumference S, and length l of the subglottal tract and vocal tract were used to calculate the inductance $L = \rho l/2A$, resistance $R = \alpha(Sl/A^2)\sqrt{\rho\mu\omega/2}$, and capacitance $C = lA/\rho c^2$, where ρ is the air density, ω is the natural frequency, and c is the speed of sound. The resistance was set in the subglottal tract, and the scaling factor of the resistance α was adjusted to $\alpha = 4.00$ under $p_{\text{lung}} = 1000$ Pa and $\alpha = 2.94$ under $p_{\text{lung}} = 1800$ Pa, consistent with the previous study.[23] Here, α represents the pressure resistance at the sudden contraction from the inlet pressure chamber to the subglottal tract and is tuned to match the amplitude of the subglottal pressure.

The pressure on the vocal fold surface was computed at discrete points from the inferior to the superior point (on the x-axis) using the subglottal pressure p_0, based on Bernoulli's principle with an added viscous loss term[7]

$$p_i = p_{i-1} + \frac{1}{2}\rho U_g^2\left(\frac{1}{a_{i-1}^2} - \frac{1}{a_i^2}\right) - \frac{12\mu\Delta x}{L_g h_{eq}^3}U_g. \tag{9}$$

Here, U_g is the volume flow rate through the glottis, a_i is the cross-sectional area at the ith point in the medial–lateral ($y - z$) plane, μ is the air viscosity, and Δx is the distance between discrete points. h_{eq} represents the height difference between adjacent points and was computed as $h_{eq} = (a_i + a_{i+1})/2L_g$. A total of 22 points were placed on the surface of the vocal folds to calculate the pressure distribution.

For simplicity, the flow separation point on the vocal fold surface was assumed to be located at the position of the minimum glottal area. The pressure up to this point was calculated using Eq. (9), while the pressure in the downstream region was assumed to be constant and no pressure recovery was considered.

At the inlet, the lung pressure p_{lung} was set to 1000 and 1800 Pa as an input voltage in the same way as the 3D flow model. At the outlet of the vocal tract, the transmission line was terminated by a radiation load of a duct with an infinite baffle. At each time step, the flow rate was computed based on the minimum glottal area and the upstream and downstream pressures in the equivalent circuit model. The pressure on the vocal fold surface was then calculated using Eq. (9), from which the aerodynamic force acting on each point was obtained. The vocal fold displacement was updated by solving Eqs. (2) and (5) using this force. This process was repeated at each time step. The time step size was set

to 1.0×10^{-5} s. For the 3D flow model, the simulation was performed on a supercomputer using 100 CPU nodes in parallel, which required approximately 450 h for each case. In contrast, the 1D flow model can be computed on a single CPU of a laptop in about 8 s.

III. RESULTS AND DISCUSSION

A. Comparison with experimental measurements

We first evaluated to what extent the computational vocal fold model replicated the vocal fold vibration pattern observed in a silicone vocal fold model experiments. We compared the simulation results with the superior view images of the vibrating silicone vocal folds recorded using a high-speed camera.[32] In this study, the computational model was constructed to have the same geometry and material properties as those in the experiment. For this comparison, the 1D flow model was used, and no vocal tract was included as in the experiment.

Figure 4 shows the superior view images of the vocal folds during one oscillation cycle as observed in both the experiment and the simulation. The fundamental frequency f_0 in the experiment was approximately 133 Hz, while f_0 was about 131 Hz in the simulation. Thus, in both models, one vibration cycle lasted around 7.5 ms. In both sets of images, time zero corresponds to the moment of minimum glottal opening area, and after 3 ms, the vocal folds reached their maximum glottal opening. At this time, the vocal fold vibration amplitude was larger in the middle compared to the anterior and posterior regions. The mean closed quotient (CQ) was 0.18 in the experimental measurements, whereas CQ was 0.11 in the simulation. The glottal closure duration of both the experiment and simulation was slightly shorter than that typically observed in human vocal folds, which is probably due to the relatively small vocal fold thickness[42] and the use of isotropic materials, which were different from the anisotropic characteristics of actual vocal fold tissue.

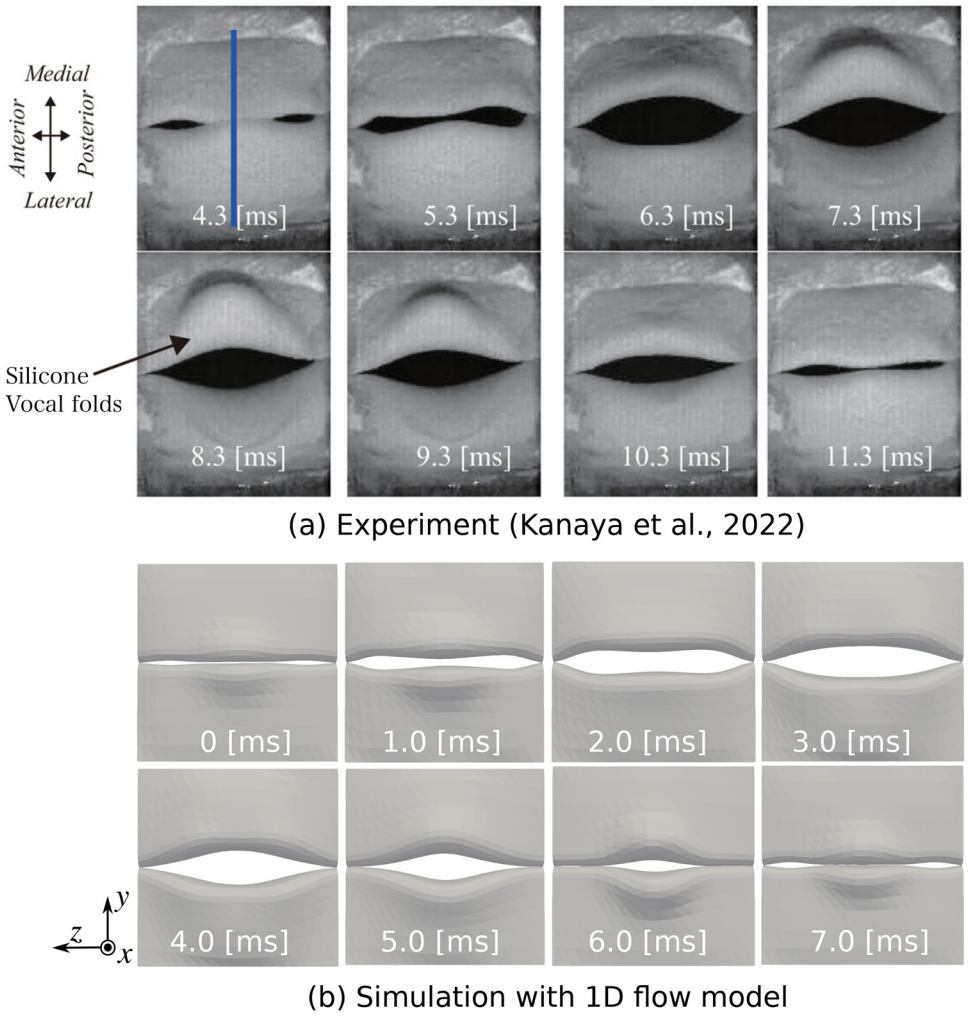

FIG. 4. Superior view of vocal fold vibrations observed in the mechanical experiment and the computational model. (a) The experiment of high-speed imaging was conducted by Kanaya *et al.*[32] (b) The computational vocal fold model was calculated by using the 1D flow model without the vocal tract. Kanaya *et al.*, JASA Express Lett. **2**, 111201 (2022); licensed under a Creative Commons Attribution (CC BY) license.

Although the large vertical motion of the vocal fold medial edge observed at 8.3 ms in the experiment was not reproduced in the computational model, the shape of the glottal opening exhibited similar trends in all images. This shows that, as in previous studies,[18] the eigenmode-based vocal fold model of this study was able to model vocal fold vibrations with sufficient accuracy. The discrepancies between the computation and experiment, such as the shorter glottal closure duration and the absence of large vertical edge motions, might be reduced by refining the grid resolution, as shown in Appendix B. Furthermore, accounting for the effect of geometric nonlinearity associated with large vibration amplitude, as well as nonlinear modal interactions, could also improve the results.

It should be noted that the current 1D flow model was mainly validated for normal, symmetric phonation. In clinical and pathological conditions involving asymmetric or irregular vocal fold vibrations, three-dimensional effects such as complex flow separation, nonuniform tissue deformation, and interactions between the vocal folds may become more significant.

B. 3D flow and pressure fields

In the following, we examine the flow and pressure fields within one vibration cycle. The 3D flow field in the central plane of the vocal folds ($z = 0$) at a lung pressure of 1800 Pa is shown in Fig. 5. The figure presents the flow field in different phases, from complete glottal closure at $t/T = 0$, through the opening phases at $t/T = 0.25$ and 0.5, to the closing phase at $t/T = 0.75$.

At $t/T = 0$, the jet flow from the vocal folds ceased, and turbulent air remained downstream of the glottis. As the vocal folds began to open at $t/T = 0.25$, the turbulent airflow was developed and directed straight toward the downstream region of the vocal tract. By the time $t/T = 0.75$ in the closing phase, the maximum flow velocity reached 70 m/s. In this case, the Reynolds number calculated from the mean flow rate and the glottal length L_g was $Re = U_g/(L_g\nu) = 1848$. The flow structures produced in the 3D model were similar to those observed in previous particle image velocimetry measurements.[2] In that study, a laminar core region extended about 5 mm downstream from the vocal folds, followed by a transitional region of approximately 10 mm, and the jet became fully turbulent around 15 mm downstream. In addition, vortices were formed in the jet shear layer, which were consistent with our observations.

Figure 6 compares the surface pressure distribution obtained from the 3D model at each time step with the 1D model. The horizontal axis of the surface pressure distribution in the 3D flow [Fig. 6(b)] corresponds to the points P0–P21 indicated in Fig. 6(a). Figure 6(c) compares the surface pressure distribution observed in the 1D model with that obtained at three cross sections in the 3D model.

At time $t/T = 0$, although the vocal folds were almost closed, a small flow rate was observed in both the 1D and 3D models [see Fig. 9(b)]. As a result, in the 1D model, the pressure remained nearly constant from the subglottal region to the glottal exit, whereas in the 3D model, there was a noticeable difference between the subglottal and supraglottal pressures. The discrepancy at the downstream surface (>P18) indicates that the 1D vocal tract model has difficulty in predicting the downstream pressure when the vocal folds were almost closed.

As the glottis began to open, the surface pressure on the vocal folds gradually decreased toward the downstream direction. Regarding this pressure drop at $t/T = 0.25$, the 3D and 1D flow models showed

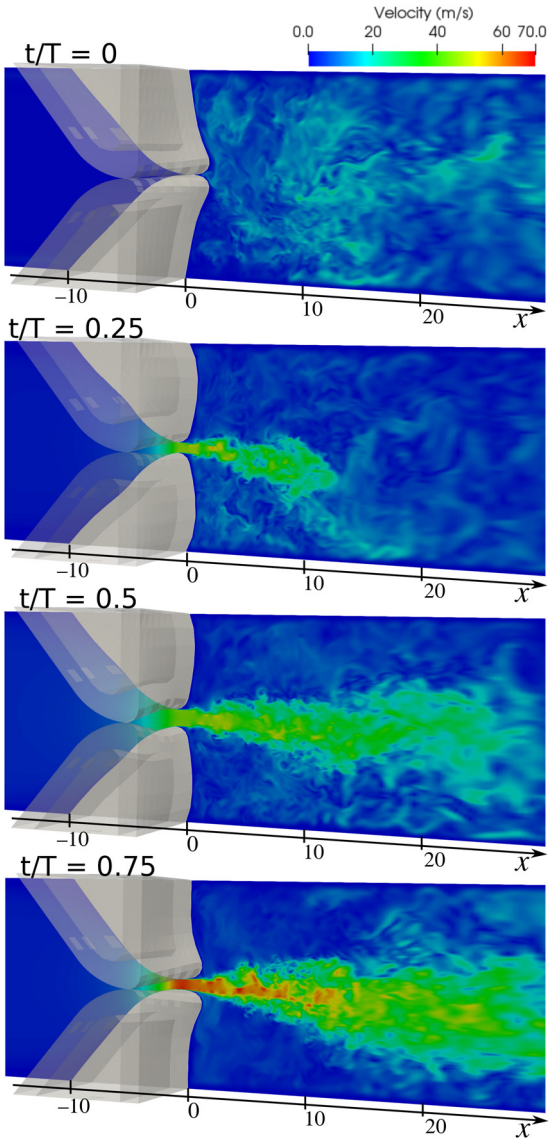

FIG. 5. Instantaneous flow field of the 3D flow model on the central plane ($z = 0$) at closed phase ($t/T = 0$), opening phase ($t/T = 0.25$ and 0.5), and closing phase ($t/T = 0.75$). The lung pressure was set to $p_{lung} = 1800$ Pa.

good agreement. Although the 3D model exhibited some differences in pressure along the anterior–posterior direction, these variations were minor (<200 Pa) compared to the overall pressure drop from the subglottal region (>1000 Pa).

At $t/T = 0.5$, when the glottis was fully open, the subglottal pressure were similar between the 1D and 3D flow models. However, the 1D flow model underestimated the pressure downstream of the glottis due to the overestimated flow rate (see Fig. 9). During the closing phase at $t/T = 0.75$, the three-dimensional model showed considerable pressure differences between the center and the anterior or posterior ends of the glottis. Although the 1D flow model fell within the range of these pressures, the 1D flow pressure could not capture such

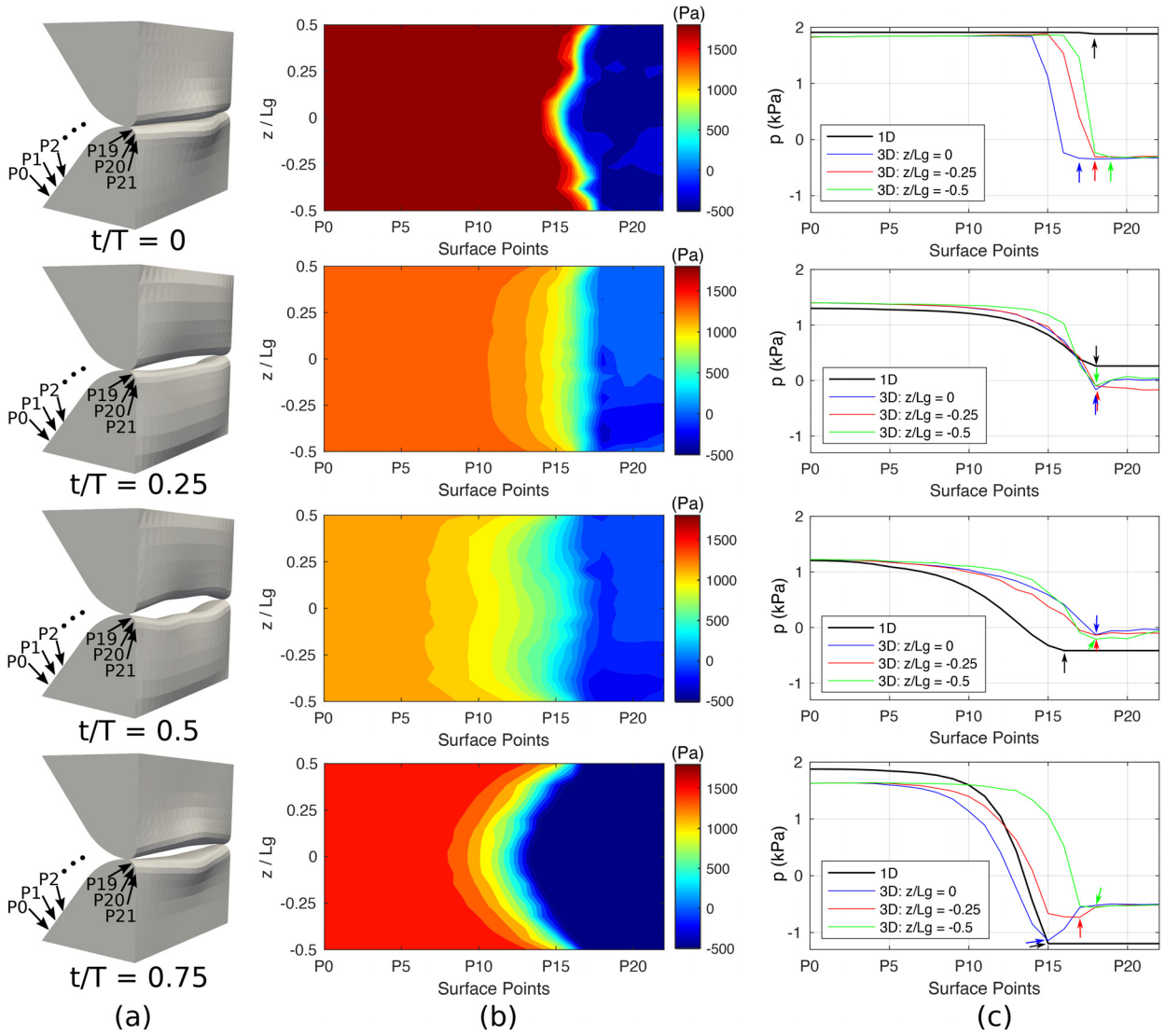

FIG. 6. Surface pressure distributions at four different phases within one oscillation cycle. The horizontal axis of the surface pressure distribution in the 3D flow (b) corresponds to the points P0–P21 along the flow direction indicated in (a). The surface pressure distribution predicted by the 1D model was compared in (c) with that observed in the 3D flow model at three cross sections along the anterior–posterior direction. The flow separation point was indicated by arrows for each model. The lung pressure was set to $p_{lung} = 1800$ Pa.

variations, which likely contributed to differences in displacement of the vocal folds.

C. Comparison of 1D and 3D flow models

Figure 7 plots the displacement of the vocal fold center ($x = y = z = 0$) under $p_{lung} = 1000$ and 1800 Pa. The time is normalized by the vibration period, with $t = 0$ corresponding to the complete closure at the middle of oscillation cycles. The fundamental frequency of the 3D flow model was 120.8 Hz at $p_{lung} = 1000$ Pa and 125.4 Hz at $p_{lung} = 1800$ Pa, while in the 1D flow model, f_0 was 122.6 Hz at $p_{lung} = 1000$ Pa and 127.7 Hz at $p_{lung} = 1800$ Pa. In both flow models, the waveform varied slightly from cycle to cycle, resulting in quasi-periodic vibrations. For $p_{lung} = 1000$ Pa, the mean difference was 0.17 mm in the x-direction and 0.07 mm in the y-direction, and for $p_{lung} = 1800$ Pa, the mean difference was 0.46 mm in the x-direction and 0.10 mm in the y-direction.

In terms of the overall waveform, the displacement of the 1D flow model was slightly smaller than that of the 3D model for $p_{lung} = 1000$ Pa, whereas the displacement of the 1D flow model was slightly larger for $p_{lung} = 1800$ Pa. Compared to our previous study using a two-mass model,[23] the present study with 3D vocal fold vibration exhibited a smaller discrepancy in the medial–lateral (y-axis) displacement (~ 0.10 mm vs ~ 0.42 mm). In the 3D flow model at the higher subglottal pressure, the y-axis displacement reached zero in every cycle, indicating complete glottal closure, whereas in the 1D model, the

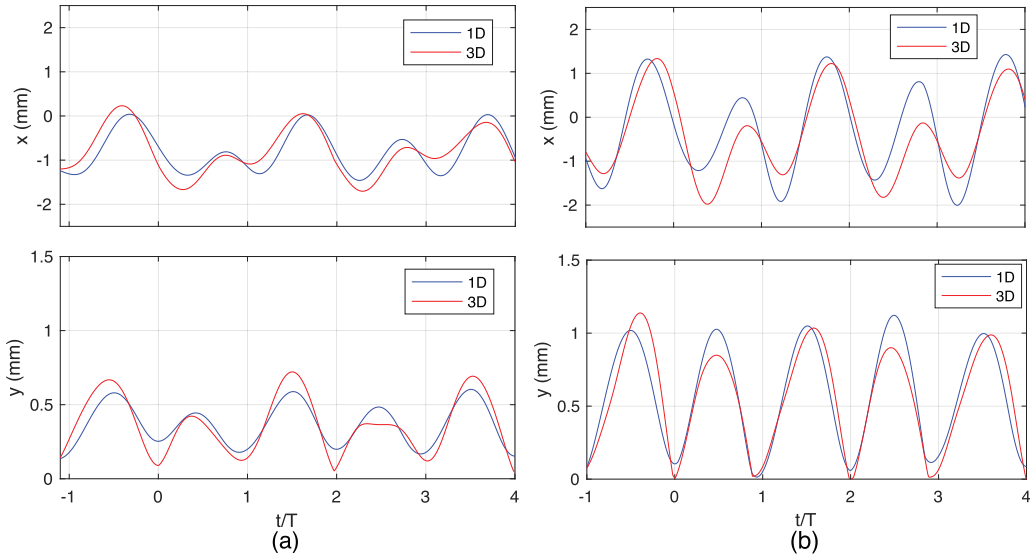

FIG. 7. Displacement waveforms at the center of vocal folds in the 1D and 3D flow models. The lung pressure was set to $p_{lung} = 1000$ Pa (a) and $p_{lung} = 1800$ Pa (b). The time t is normalized by the each time period T of the vibration.

displacement reached zero only intermittently. The flow separation point predicted by the 3D flow model was located slightly downstream of the minimum glottal area, whereas the 1D flow model assumed separation at the minimum glottal area [as shown in Fig. 6(c)]. In addition, the pressure recovery downstream from the separation point slightly increased the supraglottal pressure values in the 3D flow, which was not included in the 1D flow model. These discrepancies might have contributed to the difference in the vibration amplitude.

Figure 8 displays the subglottal pressure waveforms measured just below the glottis (at $x = -10$ mm in the 3D model). For both lung pressures of $p_{lung} = 1000$ Pa and $p_{lung} = 1800$ Pa, the pressure fluctuation was slightly larger in the 1D flow model. The mean pressure differences were 32 and 160 Pa for $p_{lung} = 1000$ Pa and $p_{lung} = 1800$ Pa, respectively. The discrepancy between the 1D and 3D models might be attributed to variations in the glottal opening waveform, as shown in Fig. 7.

At $p_{lung} = 1800$ Pa, small ripples appeared in both models, similar to those observed in our previous study,[23] due to subglottal and/or vocal tract acoustics. The mean subglottal pressures were 859.5 Pa at $p_{lung} = 1000$ Pa and 1527 Pa at $p_{lung} = 1800$ Pa in the 1D flow model, while in the 3D flow model, the mean subglottal pressures were 847.2 Pa at $p_{lung} = 1000$ Pa and 1428 Pa at $p_{lung} = 1800$ Pa, respectively. This difference may also have contributed to discrepancies in vocal fold vibration amplitudes.

Figure 9 presents the waveform of the airflow passing through the glottis and the time derivative of the flow rate, dU_g/dt, which represents the intensity of monopole sound sources. According to Lighthill's acoustic analogy,[43] the time derivative of the mass flow rate can be regarded as a monopole sound source, and the magnitude of this derivative is often used as an indicator of the source strength.[1] To examine the frequency characteristics of the source strength, the spectra of dU_g/dt are also plotted at the bottom.

At $p_{lung} = 1000$ Pa, the maximum flow rate was approximately 500 cm^3/s, and the 1D flow slightly overestimated the flow rate by approximately 70 cm^3/s. At $p_{lung} = 1800$ Pa, the maximum flow rate of the 1D flow model was 1330 cm^3/s, which was larger than that of the 3D model (970 cm^3/s). The mean flow rate difference was 72 cm^3/s for $p_{lung} = 1000$ Pa and 185 cm^3/s for $p_{lung} = 1800$ Pa. The

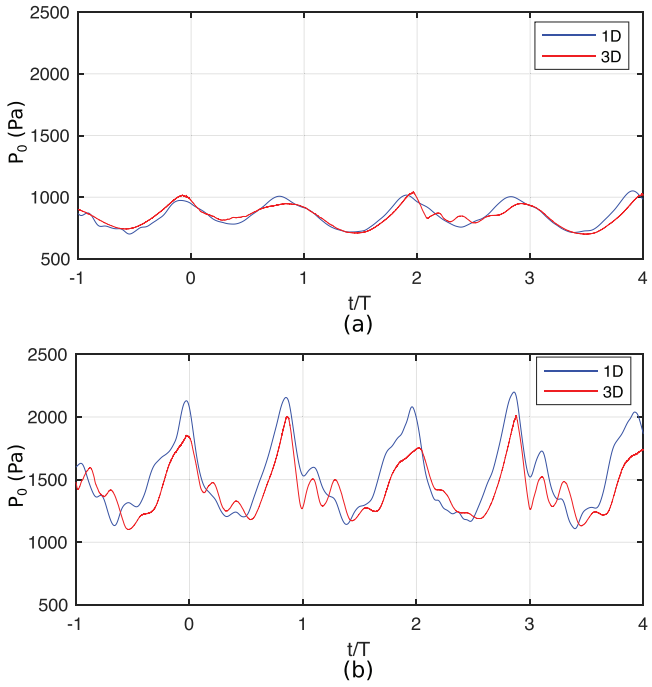

FIG. 8. Pressure waveforms measured just below the glottis for the lung pressure of $p_{lung} = 1000$ Pa (a) and $p_{lung} = 1800$ Pa (b). The pressure was sampled at $x = -10$ mm in the 3D flow model, while the pressure for the 1D model was sampled from the equivalent circuit, which was used as p_0 used in Eq. (9).

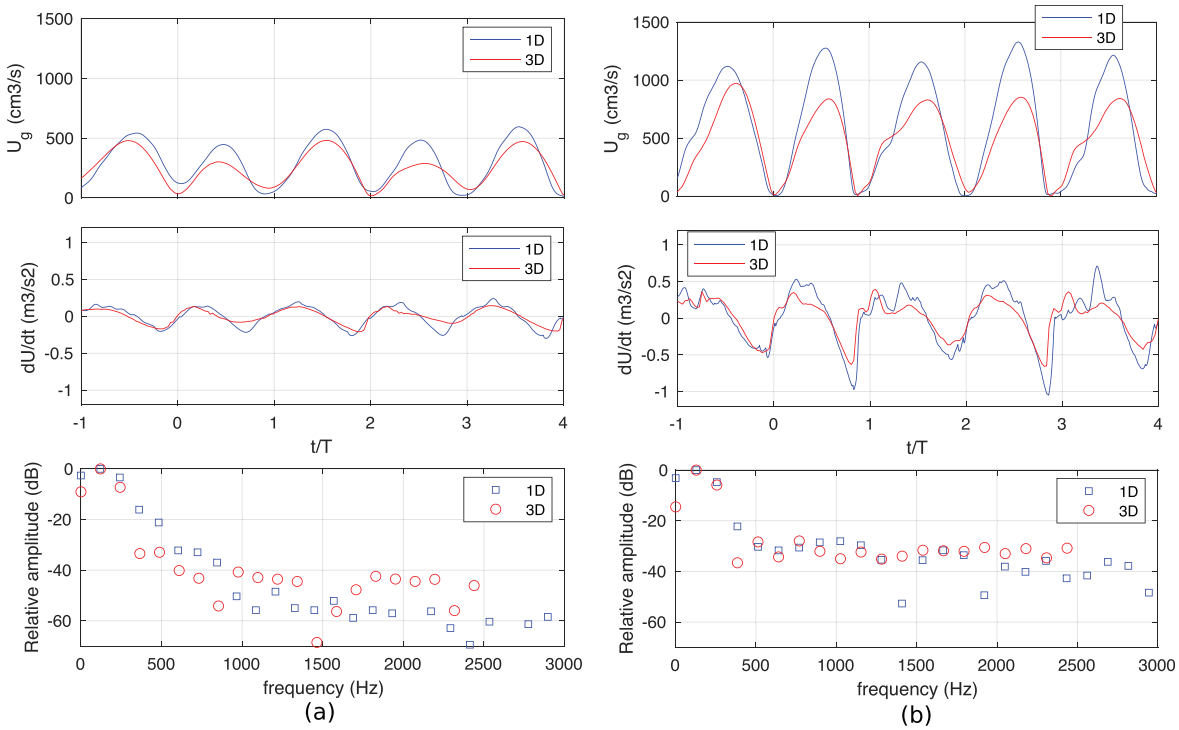

FIG. 9. Glottal flow rate and its time derivative dU_g/dt. The 1D and 3D flow models were compared under the lung pressure of $p_{lung} = 1000$ Pa (a) and $p_{lung} = 1800$ Pa (b). The spectra of dU_g/dt were calculated and are presented in the bottom panel. The amplitudes were normalized relative to the maximum value.

overestimation of the flow rate in the 1D flow model originates from the overestimation of the upstream pressure shown in Fig. 8.

The overall waveforms of dU_g/dt were similar between the 1 and 3D flow models. At $p_{lung} = 1800$ Pa, the small ripples observed in the 3D model were partially reproduced by the 1D model, while the 1D flow model overestimated the source strength by 0.5 m^3/s^2 compared to the 3D flow model. In contrast, at $p_{lung} = 1000$ Pa, the 1D flow model produced source strengths similar to those of the 3D model, with the maximum difference being approximately 0.21 m^3/s^2. These results suggest that, under lung pressures typical of normal phonation, the 1D flow model can reproduce the results of the 3D flow model with good accuracy.

The spectrum of dU_g/dt was calculated using the discretized Fourier transform with a Hann window. The spectrum was normalized by its maximum value to focus on the comparison in the spectral shapes between the 1D and 3D models. Because the waveforms of dU_g/dt were similar between the 1D and 3D flow models, the overall spectral shapes were also similar. Compared with the 3D flow model, the 1D flow model exhibited slightly higher spectral amplitudes at frequencies from 360 to 840 Hz at $p_{lung} = 1000$ Pa, and at 390 Hz at $p_{lung} = 1800$ Pa. In contrast, at $p_{lung} = 1000$ Pa, the 1D flow model underestimated the source amplitude above 1600 Hz.

The spectra of pressure at the mouth opening ($x = 175$ mm) are shown in Fig. 10. The spectral amplitudes were normalized so that the maximum amplitude was set to 0 dB. The first three acoustic resonances (formant frequencies) of the vocal tract were approximately 480, 1420, and 2380 Hz, based on the vocal tract length (175 mm) and a speed of sound of 340 m/s. For both $p_{lung} = 1000$ and 1800 Pa, the 3D and 1D flow models showed spectral peaks at these resonance frequencies. For $p_{lung} = 1000$ Pa, the amplitudes at frequencies from 500 to 700 Hz in the 1D flow were approximately 5 dB larger than those in the 3D flow, and these frequencies matched the overestimated peaks observed in the source spectra in Fig. 9. In contrast, at $p_{lung} = 1000$ Pa, the 1D flow model underestimated the source components above 1600 Hz, leading to weaker higher-frequency harmonics at the lips. The relatively stronger white noise band observed in the 3D flow model suggests that the 1D model cannot capture the turbulence-induced noise generated by incomplete glottal closure in the 3D simulation.

IV. CONCLUSIONS

This study evaluated the accuracy of a 1D flow model in predicting 3D vocal fold vibrations by comparison with silicone model experiments and a 3D flow simulation. The 1D flow model reproduced the vibratory frequency (131 Hz vs 133 Hz) and overall glottal shape observed in the experiment with reasonable accuracy. The 1D flow model also captured the general trends of vocal fold displacement and subglottal pressure of the 3D flow model, although it slightly overestimated vocal fold displacement by approximately 0.5 mm at $p_{lung} = 1800$ Pa. This discrepancy might arise from the difference in the subglottal pressure between the 1D and 3D models, as well as from the 1D flow model's inability to capture localized surface pressure variations along the anterior–posterior direction and to accurately predict the flow separation point, which needs to be investigated in future studies.

As a result, the 1D flow model exhibited larger glottal flow fluctuations and stronger source amplitudes between 360 and 840 Hz,

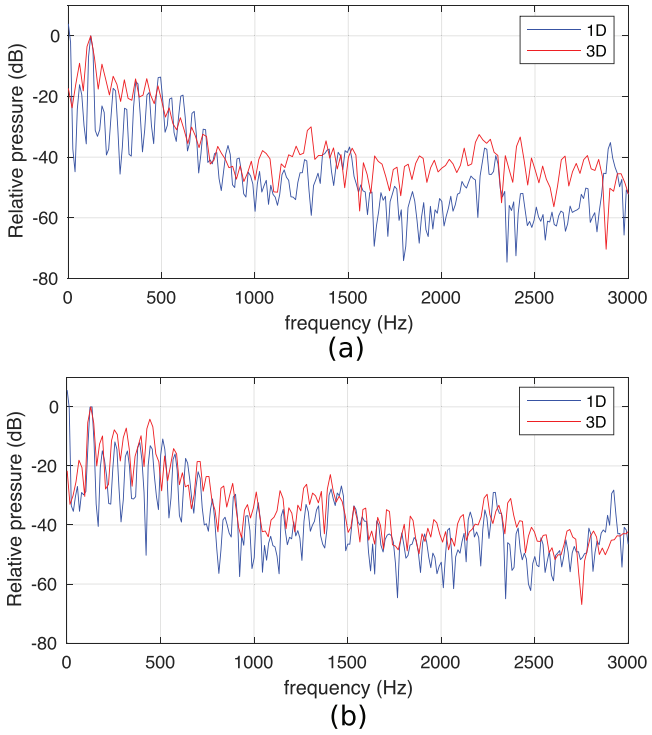

FIG. 10. Spectra of pressure fluctuations sampled at the mouth outlet ($x = 175$ mm). The 1D and 3D flow models were compared under the lung pressure of $p_{lung} = 1000$ Pa (a) and $p_{lung} = 1800$ Pa (b). The spectra were normalized by their respective maximum value.

mainly due to small ripples unique to the 1D flow model's dU_g/dt. The 1D flow model reproduced the mouth pressure up to 3000 Hz, and the formants were reasonably captured. However, the noise band of turbulence-induced sound in the 3D flow model was underestimated in the 1D flow model.

Overall, the results of this study showed that the reduced-order 1D flow model was able to reproduce major vibratory and aerodynamic features of voice production with reasonable accuracy under typical phonatory conditions. However, its limitations become apparent for detailed surface pressures and glottal flow waveforms. Future refinements should incorporate essential three-dimensional effects, which will be important for modeling pathological conditions such as vocal fold paralysis or irregular vibration patterns.

ACKNOWLEDGMENTS

This study was supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (JSPS KAKENHI) (Grant Nos. JP23K17195 and JP22KK0238), Research Grant No. R01DC020240 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, and an Institute of Digital Research and Education (IDRE) Fellowship, University of California, Los Angeles. This work used the computational resources of the Fugaku supercomputer provided by the RIKEN Center for Computational Science through the HPCI System Research Project (Project ID: hp250040). We acknowledge Aya Kawahara at Osaka University for her assistance in constructing the 3D vocal fold model.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Tsukasa Yoshinaga: Formal analysis (equal); Funding acquisition (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal). **Zhaoyan Zhang:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. The source code for the 1D glottal flow model are publicly available in GitHub at https://github.com/yossie333/modeFold_v1.3.

APPENDIX A: EFFECTS OF MODE TRUNCATION

To examine the effect of mode truncation, we conducted numerical simulations of the 1D flow model considering 20 and 100 modes. The comparisons of displacement and flow rate waveforms are shown in Figs. 11 and 12. Increasing the number of modes from 20 to 100 slightly decreased the overall displacement amplitude by

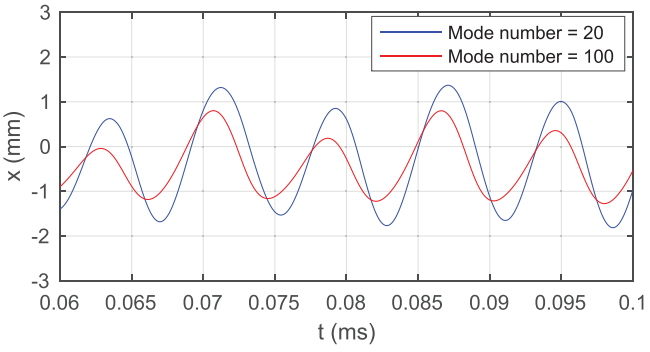

FIG. 11. Displacement of the vocal fold center predicted with 20 and 100 modes included in the 1D flow model.

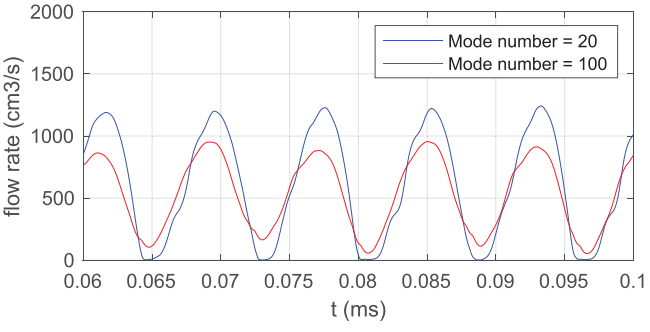

FIG. 12. Glottal flow rate predicted with 20 and 100 modes included in the 1D flow model.

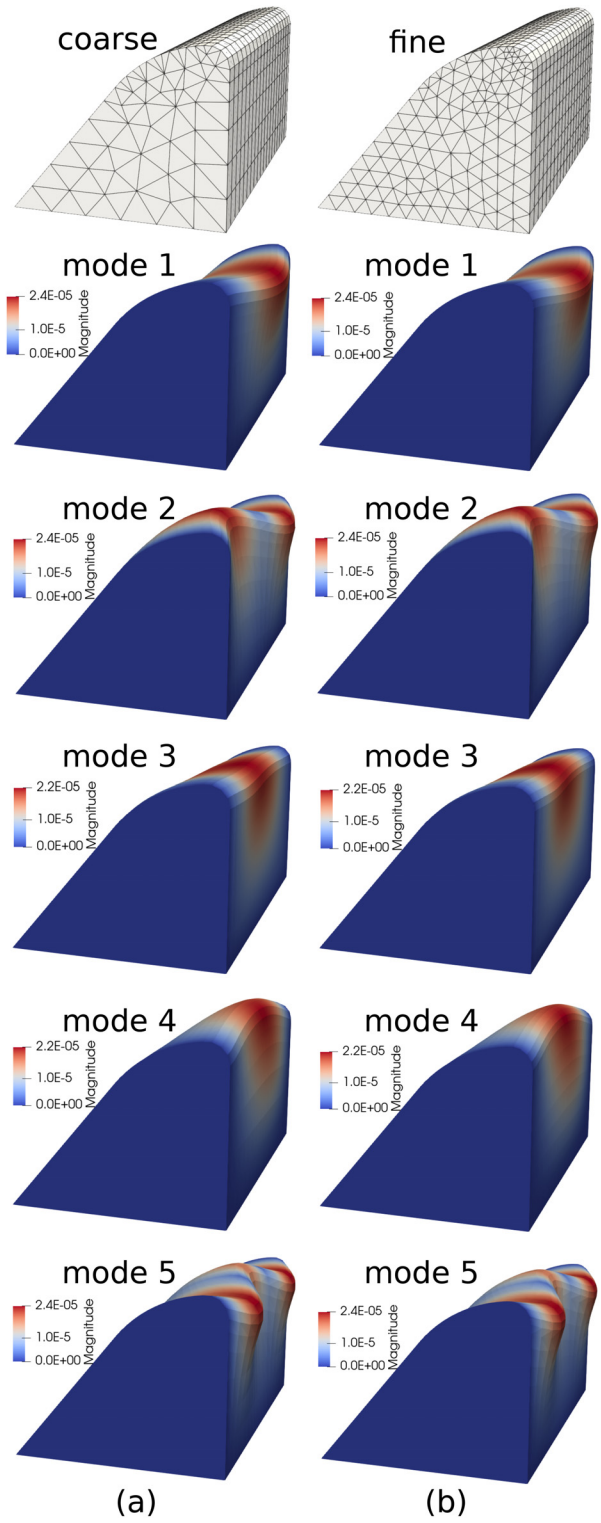

FIG. 13. The computational grids and eigenmode displacements up to fifth modes for (a) coarse grid and (b) fine grid.

about 0.5 mm, while the oscillation frequency and overall waveform shape remained unchanged. The maximum flow rate also decreased by approximately 250 cm^3/s with the higher number of modes. These results indicate that simulations with 20 modes were able to capture the general trend of vocal fold vibration in this case. However, future studies should be conducted with an increased number of modes to improve computational fidelity.

APPENDIX B: GRID RESOLUTION IN THE 3D VOCAL FOLD MODEL

To investigate whether the mesh resolution influences the vibration dynamics in the 3D vocal fold model, we compared two computational grids: a coarse grid consisting of 5650 elements and a fine grid consisting of 11 780 elements. Figure 13 presents the constructed grids and the eigenmode shapes up to the fifth mode obtained from eigenvalue analysis. Table II lists the eigenfrequencies corresponding to each grid. The eigenmodes were nearly identical between the two grids, not only for mode 1 and mode 4, which exhibit large displacements in the superior–inferior and lateral directions, but also for mode 5, which involves more complex displacement patterns. Regarding the natural frequencies, the differences were less than 0.4 Hz even at the 20th mode.

The vibration analyses were performed using the 1D flow model for each grid, and the results are compared in Fig. 14. While the displacements calculated with the fine grid were slightly larger than those with the coarse grid, the vibration frequencies and waveforms exhibited almost the same trends. These results indicate that

TABLE II. Eigenfrequencies of the vocal fold model obtained with coarse and fine grids.

	Eigenfrequency (Hz)	
Mode number	Coarse grids	Fine grids
1	54.2	54.1
2	79.1	79.0
3	94.7	94.6
4	101.9	101.8
5	111.2	111.0
6	117.3	117.1
7	118.8	118.7
8	136.0	135.7
9	145.1	145.0
10	146.1	145.9
11	155.6	155.4
12	156.1	156.0
13	165.5	165.2
14	179.8	179.6
15	181.7	181.5
16	182.7	182.4
17	184.6	184.0
18	188.2	187.9
19	189.1	188.7
20	198.7	198.4

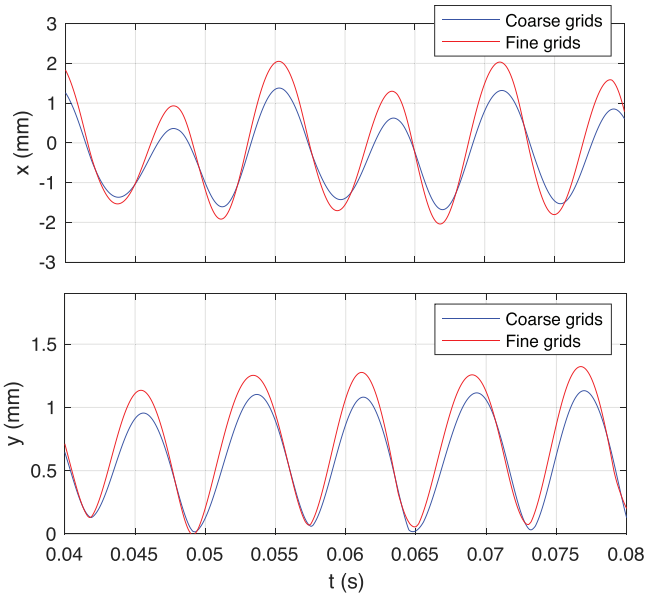

FIG. 14. Displacement waveforms at the center of vocal folds for coarse and fine grids.

although the coarse mesh slightly underestimated the displacement amplitude, the proposed model was still able to capture the essential features of the self-sustained oscillatory behavior driven by fluid–structure interactions.

REFERENCES

[1]Z. Zhang, "Mechanics of human voice production and control," J. Acoust. Soc. Am. **140**, 2614–2635 (2016).
[2]J. Neubauer, Z. Zhang, R. Miraghaie, and D. A. Berry, "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118 (2007).
[3]K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268 (1972).
[4]I. R. Titze and D. T. Talkin, "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation," J. Acoust. Soc. Am. **66**, 60–74 (1979).
[5]I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," J. Acoust. Soc. Am. **97**, 1874–1884 (1995).
[6]M. Zanartu, L. Mongeau, and G. R. Wodicka, "Influence of acoustic loading on an effective single mass model of the vocal folds," J. Acoust. Soc. Am. **121**, 1119–1129 (2007).
[7]X. Pelorson, A. Hirschberg, R. Van Hassel, A. Wijnands, and Y. Auregan, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431 (1994).
[8]T. Kaburagi and Y. Tanabe, "Low-dimensional models of the glottal flow incorporating viscous-inviscid interaction," J. Acoust. Soc. Am. **125**, 391–404 (2009).
[9]Z. Zhang, "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model," J. Acoust. Soc. Am. **139**, 1493–1507 (2016).
[10]S. Chang, C. K. Novaleski, T. Kojima, M. Mizuta, H. Luo, and B. Rousseau, "Subject-specific computational modeling of evoked rabbit phonation," J. Biomech. Eng. **138**, 011005 (2016).
[11]K. Migimatsu and I. T. Tokuda, "Experimental study on nonlinear source–filter interaction using synthetic vocal fold models," J. Acoust. Soc. Am. **146**, 983–997 (2019).
[12]Z. Zhang, "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 (2023).
[13]W. Jiang, B. Geng, X. Zheng, and Q. Xue, "A computational study of the influence of thyroarytenoid and cricothyroid muscle interaction on vocal fold dynamics in an MRI-based human laryngeal model," Biomech. Model. Mechanobiol. **23**, 1801–1813 (2024).
[14]O. Guasch, M. Freixes, M. Arnela, and A. Van Hirtum, "Controlling chaotic vocal fold oscillations in the numerical production of vowel sounds," Chaos, Solitons Fractals **182**, 114740 (2024).
[15]M. De Vries, H. Schutte, A. Veldman, and G. Verkerke, "Glottal flow through a two-mass model: Comparison of Navier–Stokes solutions with simplified models," J. Acoust. Soc. Am. **111**, 1847–1853 (2002).
[16]G. Z. Decker and S. L. Thomson, "Computational simulations of vocal fold vibration: Bernoulli versus Navier–Stokes," J. Voice **21**, 273–284 (2007).
[17]N. Ruty, X. Pelorson, A. Van Hirtum, I. Lopez-Arteaga, and A. Hirschberg, "An in vitro setup to test the relevance and the accuracy of low-order vocal folds models," J. Acoust. Soc. Am. **121**, 479–490 (2007).
[18]M. H. Farahani and Z. Zhang, "Experimental validation of a three-dimensional reduced-order continuum model of phonation," J. Acoust. Soc. Am. **140**, EL172–EL177 (2016).
[19]W. Jiang, X. Zheng, and Q. Xue, "Computational modeling of fluid–structure–acoustics interaction during voice production," Front. Bioeng. Biotechnol. **5**, 7 (2017).
[20]S. Falk, S. Kniesburges, S. Schoder, B. Jakubaß, P. Maurerlehner, M. Echternach, M. Kaltenbacher, and M. Döllinger, "3D-FV-FE aeroacoustic larynx model for investigation of functional based voice disorders," Front. Physiol. **12**, 616985 (2021).
[21]D. Bodaghi, Q. Xue, X. Zheng, and S. Thomson, "Effect of subglottic stenosis on vocal fold vibration and voice production using fluid–structure–acoustics interaction simulation," Appl. Sci. **11**, 1221 (2021).
[22]I. McCollum, A. Throop, D. Badr, and R. Zakerzadeh, "Gender in human phonation: Fluid–structure interaction and vocal fold morphology," Phys. Fluids **35**, 041907 (2023).
[23]T. Yoshinaga, Z. Zhang, and A. Iida, "Comparison of one-dimensional and three-dimensional glottal flow models in left-right asymmetric vocal fold conditions," J. Acoust. Soc. Am. **152**, 2557–2569 (2022).
[24]T. Yoshinaga, Z. Zhang, and A. Iida, "Restraining vocal fold vertical motion reduces source-filter interaction in a two-mass model," JASA Express Lett. **4**, 035201 (2024).
[25]T. Yoshinaga and Z. Zhang, "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 (2025).
[26]R. C. Scherer, S. Torkaman, B. R. Kucinschi, and A. A. Afjeh, "Intraglottal pressures in a three-dimensional model with a non-rectangular glottal shape," J. Acoust. Soc. Am. **128**, 828–838 (2010).
[27]I. R. Titze, L. Maxfield, B. Manternach, A. Palaparthi, R. Scherer, X. Wang, X. Zheng, and Q. Xue, "Pressure distributions in glottal geometries with multichannel airflows," J. Voice (published online 2024).
[28]Z. Zhang, "The influence of material anisotropy on vibration at onset in a three-dimensional vocal fold model," J. Acoust. Soc. Am. **135**, 1480–1490 (2014).
[29]Z. Zhang, "Toward real-time physically-based voice simulation: An eigenmode-based approach," in *Proceedings of Meetings on Acoustics*, Vol. 30 (AIP Publishing, 2017).
[30]K. N. Stevens, *Acoustic Phonetics* (MIT Press, 2000), Vol. 30.
[31]R. C. Scherer, D. Shinwari, K. J. De Witt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh, "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630 (2001).
[32]M. Kanaya, T. Matsumoto, T. Uemura, R. Kawabata, T. Nishimura, and I. T. Tokuda, "Physical modeling of the vocal membranes and their influence on animal voice production," JASA Express Lett. **2**, 111201 (2022).
[33]A. Bouvet, I. Tokuda, X. Pelorson, and A. Van Hirtum, "Influence of level difference due to vocal folds angular asymmetry on auto-oscillating replicas," J. Acoust. Soc. Am. **147**, 1136–1145 (2020).
[34]Z. Zhang, "Vocal fold contact pressure in a three-dimensional body-cover phonation model," J. Acoust. Soc. Am. **146**, 256–265 (2019).

[35]Q. Liu and O. V. Vasilyev, "A brinkman penalization method for compressible flows in complex geometries," J. Comput. Phys. **227**, 946–966 (2007).
[36]H. Yokoyama, A. Miki, H. Onitsuka, and A. Iida, "Direct numerical simulation of fluid–acoustic interactions in a recorder with tone holes," J. Acoust. Soc. Am. **138**, 858–873 (2015).
[37]Y. Tanaka, H. Yokoyama, and A. Iida, "Forced-oscillation control of sound radiated from the flow around a cascade of flat plates," J. Sound Vib. **431**, 248–264 (2018).
[38]D. V. Gaitonde and M. R. Visbal, "PADE-PLUSMN-type higher-order boundary filters for the Navier-Stokes equations," AIAA J. **38**, 2103–2112 (2000).
[39]T. Yoshinaga, K. Nozaki, and A. Iida, "Hysteresis of aeroacoustic sound generation in the articulation of [s]," Phys. Fluids **32**, 105114 (2020).
[40]Z. Zhang, "Laryngeal strategies to minimize vocal fold contact pressure and their effect on voice production," J. Acoust. Soc. Am. **148**, 1039–1050 (2020).
[41]J. B. Freund, "Proposed inflow/outflow boundary condition for direct computation of aerodynamic sound," AIAA J. **35**, 740–742 (1997).
[42]Z. Zhang, "Vocal fold vertical thickness in human voice production and control: A review," J. Voice **39**(5), 1183–1191 (2025).
[43]M. J. Lighthill, "On sound generated aerodynamically. I. General theory," Proc. R. Soc. London Ser. A: Math. Phys. Sci. **211**, 564–587 (1952).