

Overview on state-of-the-art numerical modeling of the phonation process

Michael Döllinger^{1,*} , Zhaoyan Zhang² , Stefan Schoder³ , Petr Šidlof^{4,5} , Bogac Tur¹, and Stefan Kniesburges¹ 

¹Division of Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology Head & Neck Surgery, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Waldstrasse 1, 91054 Erlangen, Germany

²Department of Head and Neck Surgery, University of California, Los Angeles, California 90095, USA

³Aeroacoustics and Vibroacoustics Group, Institute of Fundamentals and Theory in Electrical Engineering, Graz University of Technology, Inffeldgasse 18, 8010 Graz, Austria

⁴Institute of Thermomechanics of the Czech Academy of Sciences, 182 00 Praha 8, Czech Republic

⁵Technical University of Liberec, 461 17 Liberec 1, Czech Republic

Received 16 January 2023, Accepted 7 April 2023

Abstract – Numerical modeling of the human phonatory process has become more and more in focus during the last two decades. The increase in computational power and the use of high-performance computation (HPC) yielded more complex models being closer to the actual fluid-structure-acoustic interaction (FSAI) within the human phonatory process. However, several different simulation approaches with varying mathematical complexity and focus on certain parts of the phonatory process exist. Currently, models are suggested based on ordinary differential equations (reduced order models) but also on partial differential equations based on continuum mechanics as e.g. the Navier–Stokes equations for the flow discretized by Finite-Volume or Finite-Element-Methods. This review will illuminate current trends and recent progress within the area. In summary, the ultimate simulation model satisfying all physiological needs and scientific opinions still has to be developed.

Keywords: Numerical modeling, Computational fluid dynamics, Fluid-structure-interaction, Fluid-structure-acoustic-interaction, Machine learning

1 Introduction

Speech and the underlying phonation and voice production are fundamental prerequisites for a functioning social life and societies. However, many underlying biomechanical aspects are not fully understood, especially the fluid-structure-acoustic-interaction (FSAI) of the phonation aspects. In phonation, the FSAI is composed of the viscous airflow (fluid), the myoelastic motion of the vocal fold tissue (structure), and the production of the acoustic voice signal (acoustics), see [Figure 1](#). The most natural examination approach would be human in-vivo studies, however due to the restricted access to the human deep respiratory tract, the interrelations between airflow, vocal fold dynamics and resulting acoustics in the larynx cannot be investigated in humans. Hence, numerical models have been suggested going back to the 1960s. Since then, increased biomechanical knowledge on the phonatory process and computational power enabled to develop numerical models, simulating the phonatory process, which become more and more realistic.

For simulating the phonation process, several approaches are suggested providing complementary infor-

mation. In general, these models assemble of simple reduced order models and highly resolved continuum based models:

- **Reduced order models:** Models with high computational efficiency due to the simplification of the glottal flow and/or vocal fold mechanics. These models allow large-scale, parametric investigations of the phonatory physics which are otherwise impractical in fully-resolved phonation models.
- **Computational Fluid Dynamics (CFD):** Concentration on the occurring flow dynamics based on continuum mechanics. These models mainly focus on the airflow characteristics for different glottal and vocal tract configurations and use static or prescribed vocal fold oscillations. Commonly, these models are discretized by numerical methods as the Finite-Volume-Method (FVM), in rare cases also by Finite-Difference-Method (FDM) or the Finite-Element-Method (FEM).
- **Fluid-Structure Interaction (FSI):** High-end computational models that consider the interaction between vocal folds biomechanics commonly simulated by FEM and airflow.
- **FSAI:** High-end computational models that additionally allow to compute the acoustic source terms and the sound field. Hereby, FEM or FVM are the

*Corresponding author: michael.doellinger@uk-erlangen.de

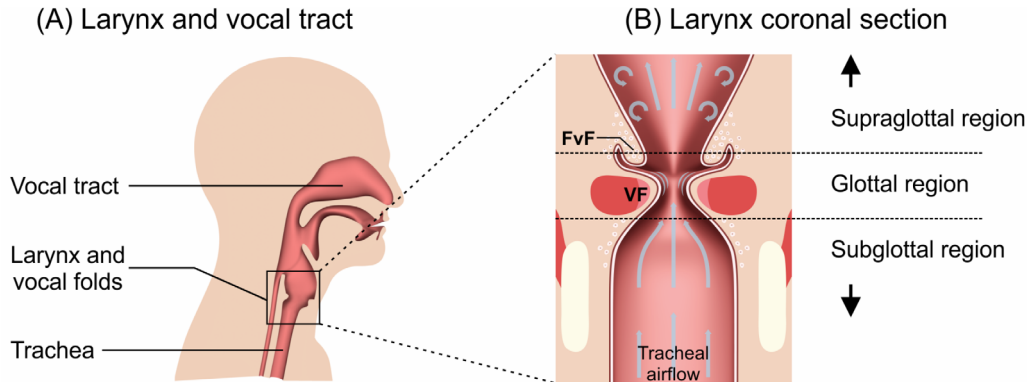


Figure 1. (A) Head with location of larynx. (B) Illustration of fluid-structure-acoustic interaction during phonation.

common methods to simulate the sound generation and propagation.

- Models combined with machine learning methods: To increase computational efficiency of highly spatially and temporally resolved models to estimate underlying biomechanical model parameters.

However, specific characteristics of current continuum based models as the direct vocal fold contact or an individualized geometry are still topics of basic research. As a consequence, those models are often highly computationally expensive (i.e., 3D-FEM or FVM methods). In 2011, a comprehensive review of numerical models, based on methodology, of human phonation was published by Alipour et al. [1]. Hence, this work will focus and provide an overview on computational models published between 2011 and 2022. However, since then, other reviews have been published that focused more on the voice physiology components but not on the mathematical approaches of the simulations [2, 3], experimental and numerical modeling [4], and mechanical characterization of vocal fold tissue, being highly important for input in numerical simulations [5].

2 Pure CFD models without FSI

Historically, the starting point for the numerical modeling of the phonation process was either the pure simulation of the laryngeal fluid dynamics or the pure simulation of the vocal folds vibration based on spring-mass elements with the simplest aerodynamic excitation functions, i.e. step functions of subglottal pressure. As the mass-spring model also constitutes the simplest strategy to model fluid-structure interaction, those models are discussed in the according section further below. In this section, the focus lies on models that purely simulate the fluid dynamics in the larynx. Thereby, those models do not only include static geometric boundary conditions but also models with driven or imposed vocal fold motion. Especially models with imposed vocal fold motion are of high potential with regard to the cost-benefit efficiency. A summary of studies discussed here is displayed in Table 1 for static vocal fold models and in

Table 2 for models with vibrating vocal folds with imposed or driven motion pattern.

2.1 Fluid flow modeling

In all but the very simplified, reduced-order models of phonation, the fluid flow is described by Navier–Stokes equations (NSE) for viscous Newtonian fluid. Some studies [6–10] use the full NSE for the compressible fluid, whose continuous form captures both fluid dynamics and acoustic perturbations. Since the Mach number in glottal flow does not exceed $M = 0.3$, incompressible NSE used e.g. in most studies represent a reasonable choice, too, and result in lower model complexity and computational costs. The incompressible model cannot capture acoustic waves. However, decoupled aeroacoustic simulations can be still run on top of incompressible simulation results.

The question whether compressible flow effects have a large impact on the simulations results has not been clarified yet. De Luzan et al. [11] found only negligible differences between the compressible and incompressible flow field in their 3D static vocal folds model. In contrast, Hájek et al. [12] compared both fluid properties by taking into account the complete fluid-structure interaction and found reasonable differences in the vocal folds motion pattern and therewith in the flow field. However, they used a 2D model which implicitly means to have constant pressure conditions in longitudinal direction of the vocal folds at every time step. Therefore, it is questionable, whether the model overestimates the impact of acoustics on flow and vocal fold dynamics.

The airflow in the trachea is usually laminar, but is dominated by complex turbulent structures in the supraglottal region. Numerical simulation of the highly unsteady turbulent airflow is a challenging issue. Laminar models (i.e. no turbulence modeling) introduce inaccuracy, since they neglect turbulent momentum transfer. Reynolds-Averaged (RANS) models face difficulties for massively separated flows, and are inappropriate for aeroacoustic simulations because they provide only mean flow solution with turbulent fluctuations averaged out. Since Direct Numerical Simulations (DNS) are unfeasible due to prohibitive computational cost, the most promising approach is Large Eddy

Table 1. Overview of the recent CFD models with static geometry since 2010. Abbreviations: VF = vocal fold, M5/M6 = Scherer’s M5/M6 parametric vocal fold model [13], custom = customized, MRI = Magnetic Resonance Imaging, FvF = False Vocal Folds, full/hemi = full larynx/hemilarynx, VT = vocal tract, ic = incompressible, c = compressible, $P_s/P_s(t)$ = constant/unsteady subglottal pressure at inlet, $P_{out}/P_{out}(t)$ = constant/unsteady pressure at outlet, $Q/Q(t)$ = constant/unsteady volume flow rate, FDM = Finite Difference Method, FVM = Finite Volume Method, FEM = Finite Element Method, LES = Large-Eddy-Simulation, WALE = Wall-adaptive Local Eddy (LES) subgrid scale model, RANS = Reynolds Averaged Navier Stokes equations, $k-\varepsilon$ = RANS model, na = not available (not provided in the paper). *Only the first author is mentioned.

Study*	Year	ID	Geometry VF & Model	FvF	Load	Num. scheme/fluid model scheme/turb. model
Mihaescu	2010	[31]	3D M5 similar & hemi		$Q = 27.5$ L/min	FVM/ic (Smargorinsky-Lilly)
Scherer	2010	[14]	3D M6 & full		$P_s = 0.29\text{--}2.45$ kPa (6 constant levels)	FVM/ic, laminar
Schwarze	2011	[116]	3D custom & full		experimental $U(t)$	FVM/ic, LES (no model)
Chisari	2011	[32]	2D half cylinders & full	x	Re = 1000	FDM/ic, laminar
Li	2012	[15]	2D M5 & hemi		$U = 30.198$ m/s	FVM/ic, laminar
Farahani	2013	[33]	2D half cylinders & full	x	Re = 900	FDM/ic, laminar
de Luzan	2015	[11]	3D M5 similar & hemi	x	$P_s = 687$ Pa Re = 3600	FVM/c & ic LES (WALE)
Li	2018	[16]	2D M5 & hemi		$P_s = ?$	FVM/ic laminar
Schickhofer	2019	[9]	3D MRI & full + VT	x	$Q(t)$ model	FVM/c LES (WALE)
Schickhofer	2020	[10]	3D MRI & full + VT	x	$P_s = 558.4$ Pa	FVM/c LES (WALE)
Li	2020	[17]	2D M5 & hemi		$P_s = ?$	FVM/ic laminar
X. Zhang	2020	[18]	3D M5 & full		$P_s = 500$ Pa	FVM/ic RANS ($k-\varepsilon$)
Li	2021	[19]	2D M5 & hemi		$P_s = ?$	FVM/ic laminar
de Luzan	2021	[20]	3D M5 & full	x	Experimental $P_s(t)$	FVM/ic LES (WALE)

Simulations (LES), where large turbulent scales are resolved and small scales modeled.

2.2 Geometries

2.2.1 Vocal fold models

The geometry of the computational models is manifold depending on the focus of the respective study. The variations incorporate different shapes of the vocal folds, including or omitting the ventricular folds (also called false vocal folds) and the vocal tract. Regarding the vocal fold shape, most studies used the so-called “M5 model geometry” shown in Figure 2 which was first introduced as experimental static model by Scherer et al. [13]. Although not officially defined as benchmark model, the M5 model became widely accepted for experimental and computational studies including static [14–20] as well as vibrating vocal folds of all types, i.e. driven dynamics [21–30]. Its big advantage is possibility of the parametric variation of the model contour for representing different shapes of the glottal duct. Thus, several computational studies investigated the flow through static vocal folds at different instances during the oscillation cycle represented by different convergent or divergent transglottal shapes of the coronal glottal duct [14–17, 19, 20]. Most of these models are quasi-2D with a

uniform rectangular glottis and without ventricular folds and vocal tract except the model by de Luzan et al. [20] who performed simulations with 3D M5 models with uniform or divergent glottal duct. Beside the M5 model, other customized and/or physiologically based model types [11, 31] as well as those with a much simpler geometry of the vocal folds as e.g. half-cylinders [32, 33] have been applied.

For models with predefined vocal fold motion, the M5 model geometry also exhibited a large spreading. Except for Zheng et al. [21], all M5 based models are 3D models with rectangular [22, 24] or semi-elliptical glottis shape [25–28]. Other studies used also simplified vocal fold shapes [34] or more realistic but still simplified geometry as in Zörner et al. [35] being realized in a 2D model.

2.2.2 Imposed vibration

There are two groups with imposed vibration of the vocal folds, those with purely medial-lateral oscillations [21, 22, 25] and those with additional rotation motion of the medial surface to reproduce the convergent-to-divergent change of the glottal duct [24–30, 34–36] during an oscillation cycle which is characteristic for the mucosal wave-like motion of human vocal folds. The numerical realization of the vocal folds motion and the handling of the numerical

Table 2. Overview of the recent CFD models with externally imposed geometry motion. Abbreviations: M5 = Scherer’s M5 parametric vocal fold model [13], custom = customized, MRI = Magnetic Resonance Imaging, FvF = False Vocal Folds, full / hemi = full larynx/hemilarynx, VT = vocal tract, ic = incompressible, c = compressible, $P_s/P_s(t)$ = constant/unsteady subglottal pressure at inlet, $P_{out}/P_{out}(t)$ = constant/unsteady pressure at outlet, $Q/Q(t)$ = constant/unsteady volume flow rate, FDM = Finite Difference Method, FVM = Finite Volume Method, FEM = Finite Element Method, LES = Large-Eddy-Simulation, WALE = Wall-adaptive Local Eddy (LES) subgrid-scale model, AMD = Anisotropic minimum dissipation subgrid-scale model, 1DOF = 1-degree-of-freedom motion, 2DOF = 2-degree-of-freedom motion, 6DOF = 6-degree-of-freedom motion, ALE = Arbitrary Lagrangian Eulerian approach with moving meshes, IBM = Immersed Boundary Method, OVM = Overset Mesh Method, na = not available (not provided in the paper). *Only the first author is mentioned.

Study*	Year	ID	Geometry VF	Model	FvF	Load	Num./fluid turb. model	VF motion Scheme/DOF
Mattheus	2011	[34]	2D-/3D custom	full		$Q(t)$ exp.	FVM/ic laminar	ALE
Zheng	2011	[21]	2D M5	full		$P_s = 1$ kPa	FDM/ic laminar	IBM/1DOF
Šidlof	2013	[22]	3D M5	full		$P_s = 2.44$ kPa	FVM/ic laminar	ALE/1DOF
Zörner	2013	[35]	2D custom	full		$P_s = 1000$ & 800 Pa	FEM/ic laminar	ALE/3DOF
Šidlof	2015	[23]	3D M5	full	x	$P_s = 360$ Pa	FVM/ic laminar	ALE/2DOF
Zörner	2016	[24]	3D M5	full	x	$P_s = 300$ Pa	FVM/ic laminar	ALE/2DOF
Jo	2016	[36]	3D exp	hemi		$P_s = 1.45 - 3.17$ kPa	FDM/ic na	na/6DOF exp.
Sadeghi	2019	[25]	3D M5	full	x	$Q(t), P_s, P_s(t)$	FVM/ic LES (WALE)	OVM/ 1&2DOF
Sadeghi (a)	2019	[26]	3D M5	full		$P_s = 3.3$ kPa $P_{out} = 0$ Pa	FVM/ic LES (WALE)	OVM/2DOF
Sadeghi (b)	2019	[27]	3D M5	full	x	$P_s = 775$ Pa & 1500 & 3000 Pa	FVM/ic LES (WALE)	OVM/2DOF
Falk	2021	[28]	3D M5	full + VT	x	$P_s = 775$ Pa	FVM/ic LES (WALE)	OVM/2DOF
Lasota	2021	[29]	3D M5	full + VT	x	$P_s = 350$ Pa	FVM/ic LES (WALE)	ALE/2DOF
Lasota	2023	[30]	3D M5	full + VT	x	$P_s = 350$ Pa	FVM/ic LES (AMD)	ALE/2DOF

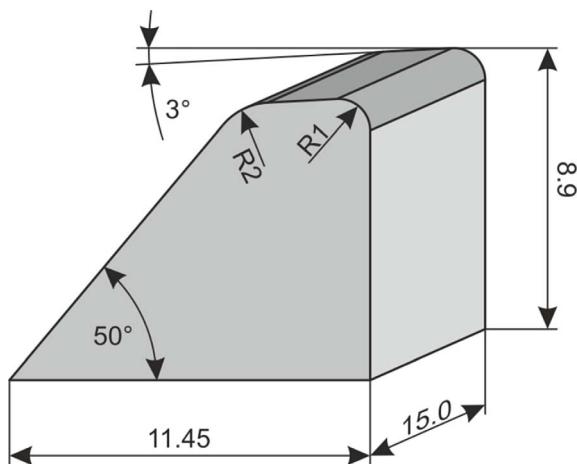


Figure 2. Geometrical dimensions of the M5 model by Scherer et al. [13].

mesh is only rarely described, e.g. for Zheng et al. [21] being the Immersed Boundary Method (IBM) and for Sadeghi and Falk with colleagues [25–28] being the Overset Mesh Method (OVM) implemented in the commercial CFD solver STAR-CCM⁺ (Siemens AG) [37]. For the other studies, the applied numerical methods to realize the vocal fold motion within the numerical mesh are not explicitly described.

2.2.3 Supraglottal geometry: Ventricular folds and vocal tract

Besides the geometry of the immediate glottal region with the two vocal folds, the upstream and especially downstream regions of the computational models are essential parts to study the laryngeal aerodynamics. Especially the ventricular folds immediately downstream of the vocal folds have been of special interest in computational studies [20, 27, 33]. In contrast to the vocal fold geometry, the shape of the ventricular folds is less standardized sometimes with half-circular shape [32, 33], some circular with divergent expansion [23, 24] and some transferred from the M5 model geometry [25, 27, 28]. The important geometric parameters in this context and thus focus of the related studies [11, 20, 32, 33] are the length of the ventricles being the space between vocal and ventricular folds and the gap between the ventricular folds. Also the pure aerodynamic influence of the ventricular folds were in the focus of computational studies [24, 27].

Including the remaining vocal tract up the mouth is of particular interest with regard to the sound generation. Thus, only few larynx models involved a complete vocal tract model in simplified straight form [9, 10, 28–30].

2.3 Numerical methods and boundary conditions

All pure CFD models were simulated based on the established numerical schemes as Finite-Difference (FDM)

Table 3. Overview of the recent CFD models with static geometry with regard to grid study, experimental validation and study topic. Abbreviation: FvF = false vocal folds. *Only the first author is mentioned.

Study*	Year	ID	Grid study	Experimental validation	Study topic
Mihaescu	2010	[31]	(x)		Aerodynamics and comparison with experiments
Scherer	2010	[14]	x	x	Aerodynamics at different subglottal pressure levels
Schwarze	2011	[116]	x	x	Aerodynamics and comparison with experiments
Chisari	2011	[32]		x	Geometry variation FvFs
Li	2012	[15]		x	Glottal duct shape variations
Farahani	2013	[33]	x	x	Geometry variation FvFs
de Luzan	2015	[11]			Geometry variation FvFs
Li	2018	[16]			Geometry variation glottal duct
Schickhofer	2019	[9]	x	x	Aerodynamics & acoustics in compressible flow
Schickhofer	2020	[10]		x	Geometry variation glottal duct in compressible flow
Li	2020	[17]			Geometry variation glottal duct: male vs. female
X. Zhang	2020	[18]		x	Aerodynamics in abnormal glottal duct geometries
Li	2021	[19]			Geometry variation glottal duct
de Luzan	2021	[20]			Geometry variation FvFs gap

[21, 32, 33], Finite-Volume (FVM) [14, 15, 20, 23, 24, 27, 28, 31, 34] and Finite-Element method (FEM) [35].

The impermeable walls in all presented larynx models were characterized as no-slip boundary conditions at which the flow velocity is either zero or equal to the velocity of the respective wall being the case at the surfaces of the vocal folds. Thereby, pure medial-lateral [21, 22] or a superimposed medial-lateral and rotatory vibration pattern [24, 25, 34] was used to move the vocal folds. Another approach was to apply the motion pattern of the vocal folds transferred from simulation with FSI to evaluate the validity of the flow field [35].

Computationally, the adaption of the numerical mesh around the vocal folds is realized if acknowledged in the studies by IBM [21] or the OVM [25, 28] which is a combination of a fixed background mesh and a small deformable and highly resolved mesh around the vocal folds [38–40].

The flow through the glottis was mostly forced by a constant pressure gradient between the inlet and the outlet boundary with absolute values between 300 Pa and 2.45 kPa as shown in Tables 1 and 2. Other models used constant flow velocity or flow rate inlet in combination with constant pressure outlet boundary conditions [31–33]. In one case, de Luzan et al. [20] applied an unsteady pressure gradient with static vocal folds to generate a flow field representing different instances during an oscillation cycle of the vocal folds.

2.4 Model verification

In general, there are two steps that have to be performed to show the validity of computational models based on FDM, FVM and FEM.

1. Grid independence study: Different numerical meshes of the simulation case with different spatial mesh resolutions are applied and selected flow parameters of the simulated physical field (flow, structure or acoustics) are computed. The goal of the grid study is to find the grid with the lowest resolution (number of

grid points or volumes) that produces similar results compared to grids with higher resolution. This ensures physically correct flow field solutions with the smallest simulation wall-time.

2. Experimental validation: Physical parameters extracted from simulations and equivalent experiments at the same location within the physical field are compared in temporal and/or spectral representation to ensure that the simulation case was modeled correctly.

The verification information in the discussed studies are summarized in Tables 3 and 4. Whereas the grid study is mandatory for every simulation and explicitly described in the summarized studies, the experimental validation is difficult in case of the human phonation. The reason is the inaccessibility of the respiratory tract for experimental sensors in living persons without anesthesia. Thus, the computational models summarized here are based on simplified geometrical models, mostly M5 or similar simplified shapes. Furthermore, some studies explicitly include accompanying experimental data to validate the computational results [14, 26].

2.5 Topics of pure CFD models

The static models are based on the quasi-steady approximation [41, 42] that describes the pulsatile flow field as a series of steady and fully developed flow fields. Thus the main topic of the studies using static vocal fold models is the characteristics of the laryngeal flow field for different glottal duct shapes representing different instances during an oscillation cycle of the vocal folds. The second topic is the influence of the ventricular folds on the flow field with regard to interaction the shear layer vertices from the glottal jet with the ventricular folds and the glottal flow resistance [9, 10, 20, 32, 33].

All studies with imposed vocal folds motion focus on the unsteady flow field development downstream of the vocal folds. More specific topics are the differences of results pro-

Table 4. Overview of the recent CFD models with externally imposed geometry motion with regard to grid study, experimental validation and study topic.

Study (1st auth.)	Year	ID	Grid study	Experiment validation	Study topic
Mattheus	2011	[34]			2D vs. 3D flow field, different glottis shapes
Zheng	2011	[21]	x		Geometry variation glottal duct
Šidlof	2013	[22]	x		Geometry variation glottal duct
Zörner	2013	[35]	x		Aerodynamics FSI vs. driven vibration
Šidlof	2015	[23]			Analysis of hybrid aeroacoustic approach in phonation
Zörner	2016	[24]	x		Geometry variation glottis shape
Jo	2016	[36]		x	Analysis of aerodynamics and aerosacoustics
Sadeghi	2019	[25]	x		Variation in motion pattern
Sadeghi (a)	2019	[26]		x	Validation with experimental model
Sadeghi (b)	2019	[27]			Influence of FVFs
Falk	2021	[28]	x		Analysis of functional based disorders
Lasota	2021	[29]			Influence of turbulence model on laryngeal aerodynamics
Lasota	2023	[30]			Implementation and tests of a new SGS model

Table 5. An overview of the used VT geometries and dimensions. Abbreviation: VT = vocal tract, CA = aeroacoustic model, LH = Lighthill’s theory, APE = acoustic perturbation equations, PCWE = perturbed convective wave equation.

ID	VT	CA	Cluster
[24]	/u/	LH, APE	[22, 23, 35]
[28]	/a/	PCWE	[43–46]
[9]	MRI	Direct	–
[10]	MRI	LH	–
[29]	/u/,/i/	PCWE	[30]

duced by 2D and 3D models [34], variations of the motion characteristics (pure lateral vs. lateral-rotatory motion of the medial surface) [25], the differences in glottal jet evolution for different glottis shapes [24] and again the influence of the ventricular folds on the phonation process [27]. Additionally, differences of results of FSI and imposed oscillation models with identical motion patterns have been analyzed [35] as well as the computational accuracy and effort of those models for a potential clinical application [25].

2.6 Acoustics

From the 26 studies included in Tables 3 and 4 with pure CFD or CFD with predefined vocal fold oscillation, 13 studies additionally simulated and analyzed the acoustic outcome (listed in Tab. 5). These flow-induced voice generation models are discussed with special attention to their reproducibility and validity. Furthermore, we cluster studies that explicitly build upon each other (see Tab. 5). In doing so, the studies of [43–46] are closely related to each other and are e.g. discussed by the study of [28]. Schoder et al. [43] validated the aeroacoustic model and [46] discussed the filtering of flow-induced sound sources. In [44], the perturbed convective wave equation (PCWE) source term is evaluated in detail. The studies [22, 23, 35] identify a second cluster and are discussed by considering the study [24]. Both study clusters used the FEM solver openCFS to obtain the acoustic solution [47].

2.6.1 Geometries

The analyzed studies (see Tab. 5) either used vocal tract geometry tuned to the formants of a vowel or an MRI-based (magnetic resonance imaging) geometry [10]. All studies omitted the vocal fold motion during the acoustic propagation simulation. To conclude, the studies considered aimed for a realistic representation of the geometrical details in 3D or the formants of the respective vowels investigated.

2.6.2 Aeroacoustic models

The studies used different aeroacoustic models (see Tab. 5), with a strong tendency to use a viscous acoustic splitting technique. In [24] and the related studies, Lighthill’s equation [48] and the acoustic perturbation equations (APE-2) [49] were applied and compared using the finite element method. The source terms of both theories were visualized to evaluate the differences and to get more insight. In [9], the acoustic radiation for different Finnish vowels ([a], [e], [i], [o], and [u]) was computed directly by considering a compressible fluid flow in the CFD simulation using the compressible form of the Navier–Stokes equations solved by the finite volume method. When resolving compressible flow directly with a CFD simulation, all interactions of the acoustic mode and the fluid dynamics directly captured. In contrast to a hybrid aeroacoustic workflow, the compressible CFD simulation requires more computational processing resources. Regarding Powell’s manipulation of Lighthill’s source term, [10] visualized the Lamb vector inside the upper airways. Finally, [28–30] computed the acoustic field based on the PCWE theory, which allows distinguishing between flow and acoustics inside the flow field (according to the finding in [50]). The PCWE model is an exact reformulation of the APE-2 [49] and was discretized by the finite element method. Since the original PCWE needs a definition of the mean flow field from a given simulation, the PCWE equation cannot be coupled in parallel to a flow simulation. Both the wave operator and the source term require the flow field during the whole simulation time of the acoustic simulation in order to pre-

Table 6. An overview of the acoustic boundary conditions and the flow temperature. Abbreviation: ABC = absorbing boundary condition, PML = perfectly matched layer.

ID	Lungs	Free-field	$T/^\circ\text{C}$
[24]	PML	PML	
[28]	ABC/PML	PML	20
[9]		ABC	27
[10]			27
[29, 30]	PML	PML	20

compute the source term and the mean flow. Recently, the PCWE theory was extended in [51] to account for moving geometries conveniently and depend solely on the instantaneous incompressible flow field. Similar to the reformulation of the PCWE, the LPCE was reformulated into a convective wave equation depending on instantaneous incompressible flow field in [52]. Another benefit of this new model is that it can easily be coupled to surface motion. In contrast, to the recently applied aeroacoustic model using the finite element method, integral methods and the boundary element method have been used to model the far-field propagation as has been summarized in [1].

2.6.3 Material and boundary conditions

The material and boundary condition data is summarized in Table 6. All studies used acoustic sound-hard walls as boundary condition for the tissue, a reasonable assumption regarding the huge impedance change from the air to the tissue. For both, the boundary to the lungs and to the free-field in front of the mouth non-reflecting boundary conditions are used (e.g., perfectly matched layer or absorbing boundary condition). An acoustic analogy was used in [10], which incorporates the free-field condition using Green’s function. Furthermore, the acoustic boundary condition towards the lower airways is not given explicitly in [9]. The temperature affecting the speed of sound was typically chosen in the ambient range, significantly lower than the body temperature. This can be improved with wave equations accounting for a variable temperature.

2.6.4 Model verification

The authors analyzed the examined human phonation models in detail. They compared the results to literature findings and a grid study was conducted for the numerical methods. In addition to that, the flow and acoustic models of [28] were validated by a synthetic experimental setup with satisfactory agreement (see Fig. 3). The few validated cases show a need for accessible experimental data in human voice production (Tab. 7).

2.6.5 Variation of model parameters

The studies related to [24] provided much numerical groundwork for hybrid aeroacoustic workflow, being later used in [28]. Zörner et al. [24] studied different vocal fold oscillation types, which were recently investigated systematically in [28]. The work in [29, 30] evaluated the influence

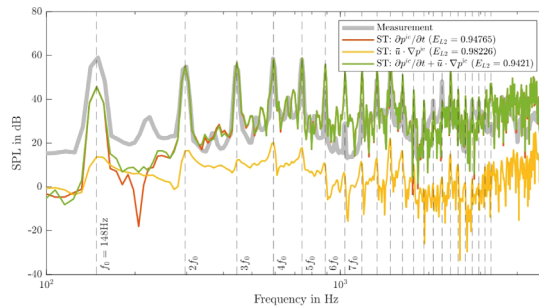


Figure 3. Sound pressure level of simulations using the time derivative of the incompressible pressure, the convective part, and the full PCWE source term as acoustic source in comparison with experimental measurement data at a microphone positioned 1 m from the end of the vocal tract. Figure adapted from [44].

Table 7. An overview of the model verification.

ID	Literature	Grid study	Exp. validation
[24]	✓	✓	—
[28]	✓	✓	✓
[9]	✓	✓	—
[10]	✓	✓	—
[29, 30]	✓	✓	—

of different LES subgrid models on the acoustic radiation. Both studies [9, 10] showed the applicability towards realistic vocal tract geometries and the changes when considering different vowels ([a], [e], [i], [o], and [u]). In [44], the aeroacoustic source term of the PCWE was analyzed. A compensation effect of the individual source terms was found and later explained in [53].

3 Fluid-Structure interaction (FSI) modeling without acoustics

3.1 FSI in phonation

The interaction between the airflow and the elastic structure of the vocal folds lies at the core of the phonation process. Historically, the first computational models of phonation tried to consider the coupling between these two physical domains, disregarding both the interaction between the acoustics and structure, and the delicate interrelation between the fluid dynamics and acoustics. The computational studies of this type published up to 2011 are well summarized in the review of Alipour et al. [1].

In the recent ten years, it has become more and more evident that the assumption of decoupled acoustics seriously impairs the accuracy and applicability of computational models of phonation. Both experimental and numerical evidence suggests that during phonation, the interaction with the acoustic field significantly affects vocal fold oscillation and should not be neglected. A notable example is the subglottal and supraglottal acoustic resonance, which can considerably influence vocal fold self-oscillation. Consequently,

in the last decade the studies focusing purely on fluid-structure interaction (FSI) are becoming rather rare. Yet, in human phonation the coupling between the fluid and structural fields is by far the strongest interaction and its modeling deserves significant attention.

When the vocal folds are adducted to phonation position and exposed to expiratory airflow, the elastic structure is subject to unsteady aerodynamic forces, which can lead to flow-induced vocal fold oscillations. As the vocal folds move, they change the geometry of the channel and influence back the airflow. In the case of modal phonation, this effect is very strong, since the vocal folds collide and completely close the channel for a considerable portion of the oscillation period. A further important characteristic is the energy transfer: for stable oscillations, there must be positive net energy transfer from the airflow to the structure over one oscillation cycle.

3.2 Structural modeling

Structural models of the vocal folds usually include up to four different tissue layers. The simplest models, e.g. [7, 54], use a single isotropic layer. Models trying to capture at least the basic physiology [55, 56] include two isotropic or transverse isotropic layers (body and cover), more complex models [6, 57] even four layers: body, ligament, superficial lamina propria and epithelium.

The constitutive relation between the stress σ and strain ϵ is either linear elastic $\sigma = E\epsilon$, hyperelastic (calculated from the strain energy density function), or viscoelastic $\sigma = E\epsilon + \eta\dot{\epsilon}$, where E is the Young modulus and η viscosity. The linear elastic and hyperelastic models can include optional Rayleigh damping $\mathbb{B} = c_1\mathbb{M} + c_2\mathbb{K}$ introduced on the level of mass and stiffness matrices \mathbb{M} and \mathbb{K} , with constants c_1 and c_2 tuned from experiments.

3.3 FSI modeling

From the modeling perspective, FSI poses serious challenges. First, the information between the fluid and solid domains (aerodynamic forces and new geometry of the domain boundary) has to be exchanged in each time-step of the simulation. Theoretically, in a monolithic approach the FSI problem could be discretized into a single matrix. However, this method is very rarely used, since it results in huge ill-conditioned linear systems. Moreover, the fluid and structural domains are often discretized using different numerical methods – the structural dynamics are almost exclusively solved by the Finite Element Method (FEM), while the primary choice for the fluid dynamics is the Finite Volume Method (FVM). Thus, the FSI in phonation is almost always solved by a coupling algorithm, where each domain is solved separately and the boundary conditions on the common interface are exchanged iteratively. When the time-step of the simulation is sufficiently low, the coupled approach offers acceptable accuracy, lower model complexity and computational costs.

The second challenge is the fact that both the structural and fluid computational domains are time-dependent, since

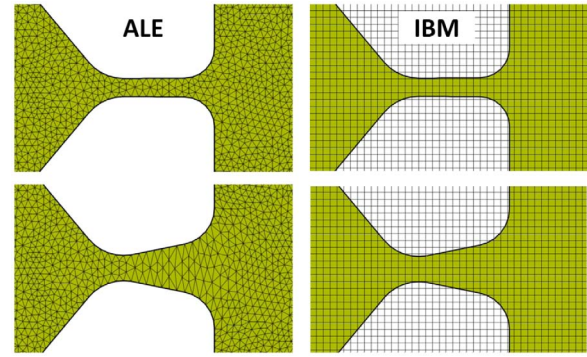


Figure 4. Difference between the body-fitted moving meshes in the ALE approach (left) and static nonconforming IBM grids (right).

they deform due to vocal fold motion. The first possible approach is the usage of moving computational meshes, conforming to the interface (vocal fold surface). In this case, the equations are usually rewritten in the Arbitrary Lagrangian–Eulerian (ALE) approach, which uses a mapping between the reference (fixed) and current (moving) domains [38, 58]. ALE equations are commonly implemented both in commercial and open-source codes, and preserve the order of accuracy and convergence properties of the numerical methods on static meshes. However, in the case of phonation modeling, the ALE approach faces problems in the glottal region. As the vocal folds approach and glottis gets narrow (or even closes completely), the ALE mapping often produces highly distorted or even inverted mesh elements and causes serious convergence issues.

The second approach, which was pioneered by Mittal, Luo, Seo et al. [59–61], is the Immersed Boundary Method (IBM). Instead of using complex moving body-fitted meshes, the simulations are run on fixed and simple Cartesian grids, which do not conform to the vocal fold surface (see Fig. 4). This greatly simplifies the task of grid generation and avoids problems with mesh deformation. On the other hand, the difficult task in IBM is tracking of the solid interface and treatment of the boundary conditions at the vocal fold surface. Imposing these boundary conditions is not straightforward, can decrease the order of accuracy and cause parallelization and performance issues.

The third possible method on how to deal with moving objects or channel walls is the OVM approach, as described in Section 2.3. None of the FSI studies listed in Table 8 used this method. Only studies with predefined vocal fold dynamics applied OVM recently, e.g. [28, 44] as described in Section 2.

3.4 Overview of the FSI models

The studies published between 2011 and 2022 focusing on fluid-structure interaction in human phonation are listed in Table 8. Note that only models with full two-way FSI, which do not solve for acoustics, are included. The one-way FSI models (with forced vocal fold motion) are commented in Section 2, and the FSI studies including the acoustic interactions are described in Section 4.

Table 8. Overview of the recent FSI models. Abbreviations: F = fluid, S = solid; M5 = Scherer’s M5 parametric vocal fold model [13], MRI = Magnetic Resonance Imaging; ic = incompressible, c = compressible, 2L = 2-layer, L-EL = linear elastic, H-EL = hyperelastic, V-EL = viscoelastic, IS = isotropic, TR-IS = transverse isotropic, RD = Rayleigh damping; ALE = Arbitrary Lagrangian Eulerian approach with moving meshes, IBM = Immersed Boundary Method; FEM = Finite Element Method, DGFEM = Discontinuous Galerkin FEM, FVM = Finite Volume Method, 2DOF = 2-degree-of-freedom (reduced order model), ODE = ordinary differential equations; NA = not available (not provided in the paper).

Study [ID]	Year	VF geom.	F/S model	F/S mesh	FSI	Numerics (F/S)	Software
Smith [57]	2012	M5	ic/4L H-EL IS RD	32k/50k	ALE	FEM/FEM	ADINA
Shurtz [6]	2013	M5	c/4L H-EL IS RD	32k/60k	ALE	FEM/FEM	ADINA
Feistauer et al. [7]	2013	Sinusoidal	c/1L L-EL IS RD	5k/2k	ALE	FEM/DGFEM	In-house
Yang [8]	2017	M5	c/2L V-EL IS	24k/6k	IBM	FEM/FEM	NA
Sváček and Horáček [54]	2018	Parabolic	ic/2DOF	NA/none	ALE	FEM/ODE	In-house
Jiang [56]	2022	MRI-based	ic/2L V-EL TR-IS	40k/2M	IBM	FEM/FVM	In-house

Table 9. An overview of the recent FSAI models. Abbreviations: VF = vocal folds, FvF = false vocal folds, VT = vocal tract, D = spatial dimensions of the model, c = compressible flow model, ic+a = hybrid approach using a coupling between an incompressible flow model and an acoustic model, L = layer, M = mass model. *Only the first author is mentioned.

Study* [ID]	Year	FSAI	VF	fVF	VT	D
Daily [64]	2013	c	2L	–	–	2
Jiang [67]	2017	ic+a	3L	✓	✓	3
Jiang [68]	2018	ic+a	3L	✓	✓	3
Bodaghi [70]	2021	ic+a	3M	✓	✓	2
Bodaghi [69]	2021	ic+a	3L	✓	✓	3
Hájek [12]	2021	ic/c	4L	–	✓	2

Surprisingly, with the exception of [56], all the other models of this class published between 2011 and 2022 are only two-dimensional and the geometry of the larynx is highly simplified – there is no attempt to model the ventricular folds and the vocal tract is a straight channel. Also, none of the authors tries to model the turbulence in the supraglottal region, all simulations are laminar. The fluid meshes are relatively coarse, even for 2D simulations. Clearly, these studies do not aim to model the phonation in the most accurate way. Instead, they focus on the modeling of the fluid-structure interaction itself, and try to investigate the effect of the model parameters on the flow-induced vibration.

The most important boundary condition for the FSI problem is the condition for the fluid flow at inlet of the flow domain. Most of the studies use constant pressure, ranging between 600 Pa [6, 57] and 1.8 kPa [56]. Feistauer et al. [7] selected a less physiologically relevant condition of constant velocity at inlet. In [54], both prescribed velocity and constant pressure are tested, and the influence of the boundary condition on flow-induced oscillation is investigated.

About half of the studies, i.e. [6, 8, 56], report on grid dependence tests, the others do not. Verification and validation in phonation modeling is a tough issue, since benchmark problems are nonexistent and comparison with in-vivo experiments is highly complicated by inter-subject variability and poor repeatability of measurements. There are some experimental data of synthetic larynx models, though, but these models are 3D models which makes a quantitative comparison of the simulated data difficult. Thus, the only paper which tried to validate the results of numerical simulations with experimental data is the study

of Jiang et al. [56], which is also the only 3D study in the FSI modeling group.

4 Fluid-Structure-Acoustic Interaction (FSAI) modeling

A logical next step in modeling the human phonation process includes an acoustic model in the incompressible fluid-structure-interaction simulation or investigates a compressible fluid-structure-interaction simulation leading to a so-called Fluid-Structure-Acoustic Interaction (FSAI) model. After some contributions from 2005 until the early years of the last decade, a few contributions dealt with FSAI in human phonation. A sample of six publications closely related to each other is discussed in this section. Furthermore, the literature study on FSAI showed that several publications with strongly misleading FSAI-claiming titles occurred in the last ten years. Table 9 shows a summary of recent FSAI models in the field of human voice production and compares the development to previous studies [62, 63]. The study [64] is an extension of the study [63]. The other studies build up on the knowledge of [62–64] and further develop the capabilities of the immersed-boundary FEM for human phonation applications.

4.1 Geometries

The 3D studies jointly aimed for a realistic description of the upper human airways (see Fig. 5). The 2D studies aim to investigate the underlying computational models and limit themselves to a more simplistic geometry.

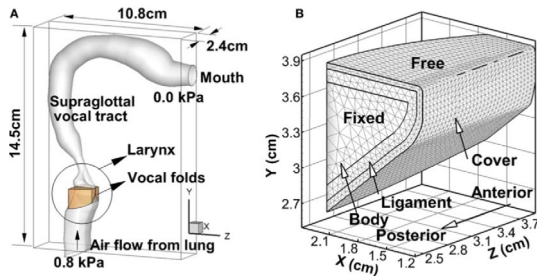


Figure 5. (A) The computational domain and geometry of the vocal folds, larynx, and vocal tract. (B) The inner-layer structure of the vocal fold as well as the boundary conditions applied on vocal fold walls. Figure adapted from [67].

Table 10 reports the geometrical properties of the reported models. Recent studies are directed toward a realistic description of human physiology, including first attempts modeling the contact using a small artificial gap between the closing vocal folds. The 3D studies used a curved vocal tract based on an in vivo-based “neutral” vowel model [65]. It was superimposed onto a realistic airway center line from the in vivo MRI measurement [66].

4.2 Models and material properties

The investigated studies modeled the solid mechanics, fluid dynamics, and acoustics with slightly different numerical techniques and material parameters; both are discussed in this section. For the numerical discretization, FEM is used and the studies aiming for realistic 3D geometry combined it with IBM. The vocal folds are modeled by a multi-layered structure with the material parameters given for the layers in Table 11. In [64], the tissue is modeled by a hyperelastic Ogden model. The density, Poisson’s ratio, and bulk modulus values of both layers were $\rho = 1070 \text{ kg/m}^3$, $\nu_p = 0.49$, and 100 kPa, respectively. Damping was simulated using a Rayleigh damping scheme with coefficients $\alpha = 24.1915$ and $\beta = 0.000127$. The vocal fold tissue was modeled as viscoelastic, transversely isotropic material in the other four studies [67–69]. A constant density $\rho = 1043 \text{ kg/m}^3$, in-plane transversal and longitudinal Poisson ratio ($\nu_p = 0.9$, $\nu_{pz} = 0$) inside the layers is used. Experimental estimates for these simulation parameters can be found in [1]. A four layer model was considered [12]. In addition to the body, cover and ligament and epithelium layer with a layer thickness of 0.05 mm was used. The stiffness parameters of the epithelium are the following $E_p = 25 \text{ kPa}$ and $\nu_p = 0.49$. Study [70] is not covered in Table 11 since the vocal fold material was modeled by a three-mass model.

The fluid parameters are defined by the air temperature given in Table 12. In the studies of [67], the viscosity is artificially reduced by 1/4 to reduce the computational costs of the turbulent structures arising. The flow is typically modeled as viscous incompressible flow of a Newtonian fluid, with the material parameters adapted to air. The coupling to the acoustic is performed with a Linearized Perturbed Compressible Equations model (LPCE) in four studies

and with a weakly compressible simulation in one of the studies. As recently shown [51], a single scalar wave equation could replace the four equations LPCE model without further restrictions. Furthermore, no study used a direct numerical simulation or LES to describe the turbulent scales realistically. Finally, all studies aim to investigate the relevance of the acoustic feedback concerning the generated voice signal.

4.3 Applied loads

As depicted in Table 12, the vocal fold motion was driven by the inlet pressure throughout the selected studies. The range varied between 250 Pa and 1000 Pa.

4.4 Boundary conditions

The boundary conditions of the solid mechanic equations are a fixation (zero displacements) of the vocal folds at the cartilage and a full coupling condition of the velocity and the stresses at the interface to the fluid. The contact is modeled artificially by a contact plane and a residual gap. A pressure inlet and outlet condition drive the flow, and a no-penetration and a no-slip condition are applied to the walls. For the acoustics equation in four studies [67–70], sound-hard wall boundary conditions are applied at the tissue walls. A total reflecting boundary condition models the free field radiation. Towards the lower airways, a Dirichlet boundary condition was described in [67, 68] or an anechoic termination was used in [69, 70].

4.5 Model verification

An overview about the model verification of FSAI models is shown in Table 13. The study [12, 64] compared the results to results from literature as the other investigation analyzed here. Furthermore, the model verification was performed by refining the grid and time step sizes. The computational models used in [67–69] build upon established FSI models from previous years and the experience gained there. In [70], a grid independence study was carried out. Additionally, the results of the applied methods were benchmarked with a proprietary solver.

4.6 Variation of model parameters

In [64], the impact of a varying length of the subglottal tract and the variations induced by modeling the fluid as slightly incompressible are investigated. The results of this study are promising, and the applicability to 3D and realistic geometries should be investigated in the future. In [12], an incompressible and compressible flow simulation are applied to model the human phonation process. The study [68] investigated a varying longitudinal cover and ligament layers thickness. It was found that the varied thickness resulted in up to 24% stiffness reduction in the middle and up to 47% stiffness increase near the anterior and posterior ends of the vocal fold. However, the average stiffness is not affected. A longitudinal layer thickness variation result in a more energy-efficient vibration than the vibra-

Table 10. An overview of the recent FSAI models, denoted by the ID. Abbreviations: VF = vocal folds, VT = vocal tract, L = length of the model, cs = cross-section.

ID	VF	VT	L /mm	Comment
[64]	[117]	–	40–300	–
[67]	[91]	Curved [65]	307	Cylindrical cs
[68]	[91]	Curved [65]	307	Cylindrical cs
[70]	[118]	Pipe	200	20 mm width
[69]	[91]	Curved [65]	340	Cylindrical cs
[12]	[13]	/i/	250	–

Table 11. An overview of the mechanical parameters of the VF in FSAI models. Abbreviations: E_p and E_{pz} = transversal and longitudinal Young's modulus, respectively, G_{pz} = longitudinal shear modulus, η = damping ratio.

ID	Cover	Ligament	Body
[64]	d /mm, E_p /kPa 2, 5	–	–, 15
[67, 68]	d /mm, E_p /kPa, E_{pz} /kPa, G_{pz} /kPa, η /Pa s	1.1, 3.31, 66, 40, 0.75	–, 3.99, 80, 20, 1.25
[69]	0.5, 2.01, 40, 10, 0.15	1.1, 3.31, 66, 40, 0.225	–, 3.99, 80, 20, 0.375
[12]	d /mm, E_p /kPa, v_p 1, 2, 0.49	–, 8, 0.49	–, 65, 0.4

Table 12. An overview of the fluid parameters of the FSAI model.

ID	T	Flow	Load	Acoustics
[64]	18 °C	ic/c viscous	900 Pa	Weakly compressible
[67]	35 °C	ic, viscous	800 Pa	LPCE
[68]	37 °C	ic, viscous	1000 Pa	LPCE
[69, 70]	37 °C	viscous	800 Pa	LPCE
[12]	36 °C	ic/c viscous	250 Pa	Compressible

Table 13. An overview of the model verification procedure of the FSAI model.

ID	Literature	Grid study	Exp. validation
[64]	✓	✓	–
[67]	✓	(✓)	–
[68]	✓	(✓)	–
[70]	✓	–	–
[69]	✓	–	–
[12]	✓	–	–

tions with a constant layer thickness. In [69], subglottic stenosis's effect on glottal flow dynamics, vocal fold vibration, and acoustics during voice production was investigated. The results show that subglottic stenosis affects voice production only when its severity is beyond a threshold. For the glottal flow rate and acoustics, this threshold is 75%, and at 90% for the vocal fold vibrations.

5 Reduced-order models of phonation

Fully resolving the complex physics involved in human voice production in three-dimensions is computationally challenging and expensive, as shown in the sections above.

The high computational costs make it impractical to use these computational models for extensive parametric investigation of the physics of voice production or practical applications. Thus, an important goal of voice research is to develop reduced-order models of phonation that are computationally efficient enough to allow large-scale parametric voice simulations for e.g. parameter estimation [71, 72] or predicting clinical intervention outcomes. Since communication is the main interest of voice production, reduced-order models in principle can be developed such that perceptually relevant physiologic features are sufficiently represented and features of minimum perceptual relevance are simplified [2], thus improving computational efficiency. Such simplification can be applied to both the fluid and structure sides.

On the fluid side, while recent experimental studies showed that the glottal flow is highly three-dimensional and exhibits many complex phenomena such as flow separation, intraglottal vortex generation, shear layer instabilities, and transition to turbulence [60, 73–75], the acoustic and perceptual relevance of these three-dimensional flow features, particularly to the harmonic component of voiced speech, remains unclear. In an experimental study, Zhang and Neubauer [76] quantified the acoustic relevance of supraglottal vortical structures by experimentally disturbing the supraglottal flow field and observing its impact

on the produced voice. They showed that alterations in the supraglottal flow field produced no significant changes in the produced sound, suggesting that the three-dimensional supraglottal flow may be simplified in phonation models. The relevance of intraglottal vortices to voice production was also questioned and found to be small [77–79], especially for subglottal pressures typical of normal-intensity phonation. While more studies will be needed to establish the relevance of three-dimensional flow features to voice production, these studies suggest that many three-dimensional flow features may be simplified to some degree to improve computational efficiency of phonation models.

In early models of phonation such as the two-mass and three-mass models, these three-dimensional flow features are often neglected and the glottal flow is simplified to be one-dimensional (1D). Many studies have aimed to quantify the accuracy of one-dimensional flow models by comparing to either simulations solving the Navier-Stokes equations or experiment. Decker and Thomson [63] compared 1D flow models to a 2D Navier–Stokes flow model, and their results showed that the 1D flow models were able to predict the flow rate, intraglottal pressure, and glottal width with reasonable accuracy, particularly at conditions of small glottal width. At large glottal width, 1D models overestimated the flow rate. In a series of studies, Luo and colleagues [80–82] proposed improvements to the 1D flow model, and showed that when coupled to a 3D vocal fold model, these 1D flow models were able to predict phonation frequency, vibration amplitude, and vertical phase difference in vocal fold vibration with reasonable agreement with those from 3D flow models. Yoshinaga et al. [83] evaluated the accuracy of 1D flow models in left-right asymmetric vocal fold conditions, and showed that the 1D flow model was able to predict vocal fold vibratory patterns and selected voice outcome measures with reasonable agreement with the predictions from the three-dimensional Navier–Stokes-based flow model (Fig. 6). An interesting finding of the study by Yoshinaga et al. [83] is that such agreement was reached despite differences in the predicted flow pressure on vocal fold surface, indicating that vocal fold properties play a larger role than the glottal flow in determining the overall pattern of vocal fold vibration and the produced voice.

Studies also showed that predictions from phonation models using a low-dimensional flow model compared well with experiments, despite the many complex flow phenomena being neglected in these models [84–86].

On the structural side, the vocal folds are known to exhibit nonlinear, anisotropic, viscoelastic mechanical behavior. Although there have been recent efforts toward developing structurally-based constitutive models [5, 87–90], implementation of these constitutive models is computationally expensive, especially in three-dimensional continuum models. As a result, they have yet to be used in phonation models. In phonation models, the vocal folds have been modeled as hyperelastic, often isotropic, materials (e.g., [6, 86]). The computational cost is generally high, particularly when a large-deformation, large-strain formulation is used.

One simplification often made to improve computational speed is to simplify the vocal folds as linear elastic

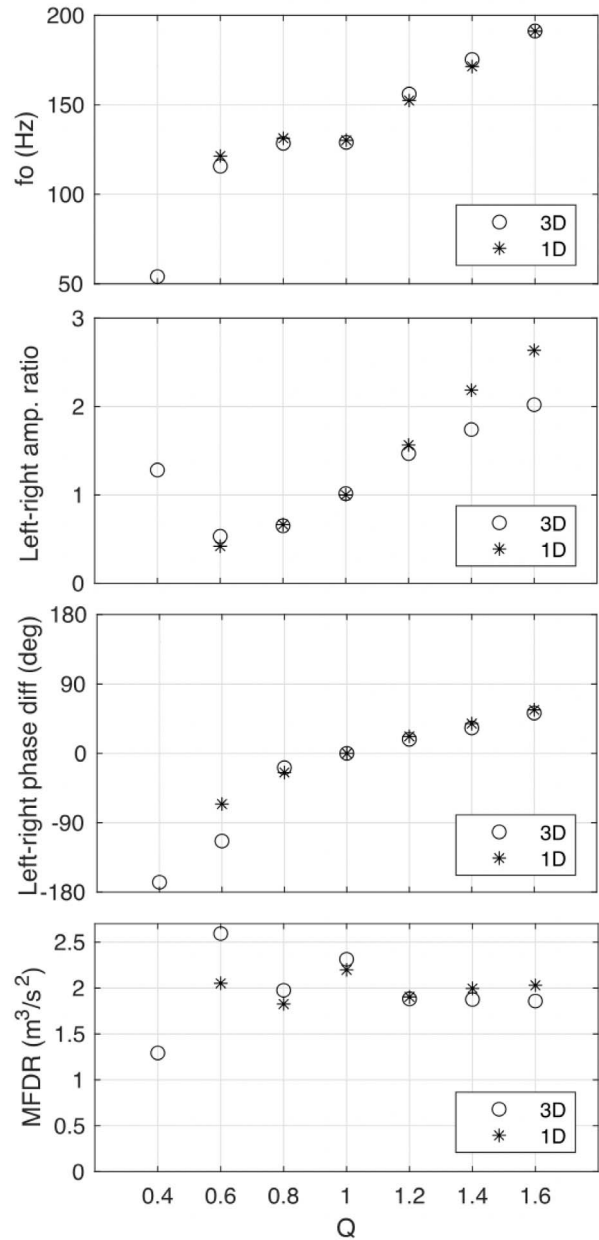


Figure 6. Voice outcome measures predicted from 1D flow-based reduced-order models agree reasonably well with those from 3D Navier-Stokes-based flow models in left-right symmetric and asymmetric vocal fold conditions (as quantified by the left-right stiffness asymmetry ratio Q in the abscissa). The measures include from the top to bottom the fundamental frequency f_0 , left-right vocal fold vibration amplitude ratio, left-right phase difference in vocal fold vibration, and maximum flow declination rate (MFDR). Figure adapted from [83].

materials. It is generally understood that the elastic moduli in these models should be interpreted as the tangent elastic moduli around a specific deformation state of the vocal fold, which would vary with vocal fold posturing. Thus, the effect of material nonlinearity, often due to changes in vocal fold posturing, can be investigated by parametric variations in the elastic moduli and vocal fold geometry in these

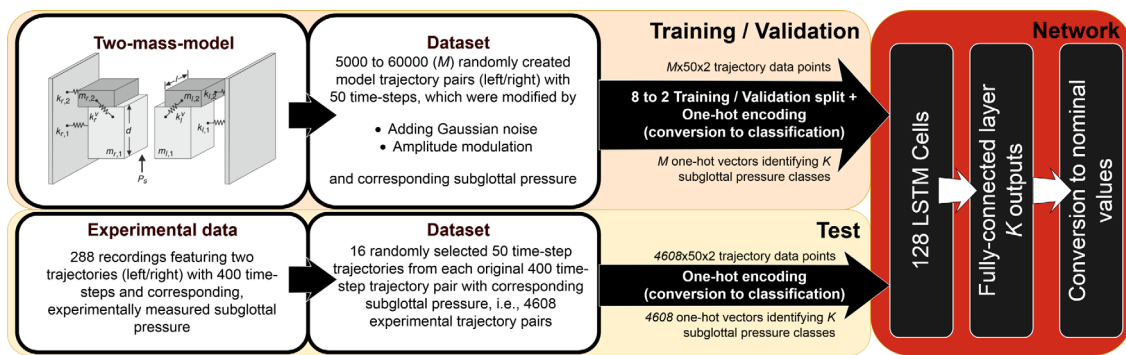


Figure 7. Schematic of biomechanical parameter estimation by adapting a two-mass-model towards vocal fold vibrations recorded by high-speed imaging by using a recurrent neural network. Figure adopted from [106].

models. The linear elasticity simplification has been widely used in phonation models (e.g., [62, 91–93]).

Another simplification often made in vocal fold models is to assume small-strain deformation of the vocal folds, which neglects geometric nonlinearity in the vocal folds. However, Chang et al. [94] showed that while this simplification was able to predict qualitative trends of phonation, neglecting geometric nonlinearity led to significant differences in the glottal gap width, flow rate, and impact stress, as errors due to the small-strain simplification was amplified by vocal fold contact and fluid-structure interaction.

Further improvement in computational efficiency on the structural side can be achieved by reducing the order of the system governing equations of the vocal folds. Zhang [95, 96] proposed an eigenmode-based reduced-order model, in which the system governing equations of the vocal fold are projected onto the space spanned by the in vacuo eigenmodes of the vocal folds. He showed that the model was able to predict phonation frequency, sound pressure level, closed quotient of vocal fold vibration, and the amplitude difference between the first and second harmonics (H1–H2) in the output voice spectrum with reasonable accuracy with the use of the first 100 vocal fold in vacuo eigenmodes [96]. This number is significantly smaller than the degrees of freedom in typical finite element models of phonation which is often in the order of tens of thousands, thus significantly improving computational efficiency. With the improved computational efficiency, this reduced-order model has been used in a series of large-scale, three-dimensional, parametric studies, using either simplified [97–99] or MRI-based realistic vocal fold geometry [93, 100]. The large number of conditions investigated (about 200,000 vocal fold conditions) allowed for the first time a systematic investigation of the global cause-effect relationship between vocal fold physiology (vocal fold geometry, stiffness, and subglottal pressure) and voice outcomes (vocal fold vibration, glottal flow, and voice acoustics) in a large range of vocal conditions.

Due to challenges in experimentally measuring material properties and geometry of the vocal folds in human or animal models, computational models of phonation are often qualitatively validated by their ability to produce voice measures (e.g., fundamental frequency, vocal intensity, and various flow-based measures) that are within typical human range; e.g., [56, 81, 82, 86, 101, 102]. However, the

comparison was often qualitative or limited to a small number of voice conditions. Due to the experimental challenges, fully-resolved fluid-structure-acoustics methods as discussed in earlier sections will likely play an important role in the development and validation of reduced-order models of phonation, by systematically understanding the acoustic and perceptual relevance of different physical components and gradually integrating them into reduced-order models.

6 Numerical modeling and machine learning

As described above, developing and applying numerical simulation models enables the analysis of fundamental biomechanical behaviour of the phonatory process and the occurring fundamental FSAI processes. However, for reproducing a specific phonatory process or specific vocal fold vibrations as e.g. recorded with high-speed imaging [103], corresponding model parameters have to be adapted. Automated estimation of biomechanical model parameters from numerical simulations of the phonatory process has been performed for many years [104]. Applied probabilistic approaches [105] and optimization approaches [55] belong also in the broader area of artificial intelligence. However, in contrast to machine learning methods that need training data to develop the corresponding mathematical model behind, these approaches directly compute automatically the desired model parameters.

Combining numerical phonatory simulation models with machine learning methods is a relatively new approach, but is becoming more and more popular. Since 2019 several studies have been performed. The first study reporting parameter estimation of a lumped mass model (two-mass model – 2MM) using a recurrent neural network was Gomez et al. 2019 [106]. They trained their model with trajectories generated by the 2MM and used 288 ex-vivo vocal fold dynamics to test their model, see Figure 7. The suggested work showed that the subglottal pressure estimation is in the same accuracy range as classical optimization techniques [102] by reducing the online computational costs to a minimum.

Computing the glottal midline in HSV data during phonation is highly important for judging the dynamic vocal fold left-right symmetry. So far, the needed midline

detection was a semi-automatic task needing user interaction and was hence, to a certain degree, user dependent [107]. To overcome this shortcoming, Kist et al. 2020 [108] applied deep neural networks (DNN) on a six-mass-model (6MM) to automatically estimate the glottal midline detection. They showed that DNNs outperformed semi-automatic approaches. However, all methods, semi- or fully automatic, achieved sufficient accurate results. Finally, they suggested the so called “GlottisNet” that enables the simultaneous estimation of the glottis midline and the segmentation of the glottal area.

Li et al. [109] suggest a new numerical 1D flow model that contains an analytical computation of the entrance effect and the pressure loss in the glottis. This was derived by training the 1D flow model with flow data based on a complex 3D fluid-structure interaction model. For training, they used a regression approach.

Zhang et al. [110] introduced a so called generalized reduced-order model (ROM) where they estimate glottal flow and glottal pressure distribution along the vocal fold based on subglottal pressure and the glottal shape provided by universal kinematics equations (UKE). Within the ROM, a given glottal shape is optimized (i.e. matched) in each time-step by the UKE using a genetic optimization (GA) algorithm. The DNN, trained with a Navier–Stokes fluid solver model, estimates the glottal flow and glottal pressure distribution. Based on this, a FEM solid solver computes the glottal shape in the next time-step. In a subsequent study, Zhang et al. [111] improved the accuracy of the same parameter estimation by using a long short-term memory (LSTM) network and also reduced the computational time by discarding the GA optimization.

Zhang [112–114] developed a simulation-based machine learning model to solve the inverse problem of voice production. The goal was to estimate vocal fold physiology (geometry, position, and mechanical properties) and subglottal pressure from the produced voice outcome. With the improved computational efficiency of the reduced-order model developed in his earlier work [95, 97, 98], he was able to perform voice production simulations for a large number of voicing conditions. In Zhang [114], a total of 221,400 vocal fold conditions were simulated. Results from these simulations were then used to train a neural network to estimate vocal fold physiology and subglottal pressure from the produced voice. Unlike other machine learning studies that used time-series data as input, Zhang [112] used perceptually-important voice features extracted from the simulated voice outcomes as input to the neural network, which significantly reduced the amount of input data and training efficiency. The voice features included measures of voice acoustics (fundamental frequency, vocal intensity, etc.), aerodynamics (mean and peak-to-peak glottal flow, maximum flow declination rate, etc.), and vocal fold vibration [113]. The output of the neural network included vocal fold length, medial-lateral depth, vertical thickness, glottal angle (a measure of vocal fold approximation), vocal fold stiffness along the transverse and longitudinal direction, and subglottal pressure. While the neural network was trained using simulation data, comparison to excised

human larynx experiment showed that the neural network was about to predict the subglottal pressure and vocal fold geometry with reasonable accuracy [112]. Zhang [114] further applied this neural network to data from human subjects. The results showed reasonable accuracy in estimating the subglottal pressure, and the neural network was able to qualitatively differentiate soft and loud voice production regarding differences in the subglottal pressure and degree of vocal fold adduction.

7 Conclusion

This overview on numerical approaches for simulating the human phonation process illustrates the diversity of current research. The various approaches concentrate on different components of the phonatory process but also demonstrate the differences of research groups on how the phonatory process should be simulated and analysed. There is still an intensive discussion on what components of the phonatory process are more important and should be considered in simulations and what can be neglected. A promising way to determine what components of the phonatory process are essential may be answered by FSAI simulations in the future. Most importantly, FSAI models that account for the contact of the vocal folds are expected to shed new light on the phonatory process. However, to deliver this expectation, the model validation of the accuracy based on experimental data is essential although being very challenging for FSAI models. This validation step could potentially path the way for driven vocal folds models based on clinical imaging procedures and their usage in patient-tailored applications. As a final conclusion, there is no doubt that in future more and more models will combine simulations with machine learning techniques [115].

Conflict of interest

Authors declare no conflict of interest.

Acknowledgments

S. Schoder acknowledges the support from the ÖAW research grant “Understanding voice disorders”, received from “Dr. Anton Oelzelt-Newin’sche Stiftung”. M. Döllinger acknowledges the support from Deutsche Forschungsgesellschaft (DFG), grant no. DO1247/21-1.

References

1. F. Alipour, C. Brucker, D. Cook, A. Gommel, M. Kaltenbacher, W. Mattheus, L. Mongeau, E. Nauman, R. Schwarze, I. Tokuda, S. Zorner: Mathematical models and numerical schemes for the simulation of human phonation, *Current Bioinformatics* 6, 3 (2011) 323–343.
2. Z. Zhang: Mechanics of human voice production and control. *Journal of the Acoustical Society of America* 140, 4 (2016) 2614–2635.

3. C. Calvache, L. Solaque, A. Velasco, L. Penuela: Biomechanical models to represent vocal physiology: a systematic review. *Journal of Voice* 37 (2023) 465.e1–465.e18.
4. M. Döllinger, S. Kniesburges, M. Kaltenbacher, M. Echternach: Current methods for modelling voice production. *HNO* 64, 2 (2016) 82–90.
5. A.K. Miri: Mechanical characterization of vocal fold tissue: a review study. *Journal of Voice* 28, 6 (2014) 657–667.
6. T.E. Shurtz, S.L. Thomson: Influence of numerical model decisions on the flow-induced vibration of a computational vocal fold model. *Computers & Structures* 122 (2013) 44–54. *Computational Fluid and Solid Mechanics* 2013.
7. M. Feistauer, J. Hasnedlová-Prokopová, J. Horáček, A. Kosík, V. Kučera: DGFEM for dynamical systems describing interaction of compressible fluid and structures. *Journal of Computational and Applied Mathematics* 254 (2013) 17–30.
8. J. Yang, X. Wang, M. Krane, L.T. Zhang: Fully-coupled aeroelastic simulation with fluid compressibility – For application to vocal fold vibration. *Computer Methods in Applied Mechanics and Engineering* 315 (2017) 584–606.
9. L. Schickhofer, J. Malinen, M. Mihaescu: Compressible flow simulations of voiced speech using rigid vocal tract geometries acquired by MRI. *Journal of the Acoustical Society of America* 145, 4 (2019) 2049–2061.
10. L. Schickhofer, M. Mihaescu: Analysis of the aerodynamic sound of speech through static vocal tract models of various glottal shapes. *Journal of Biomechanics* 99 (2020) 109484.
11. C.F. de Luzan, J. Chen, M. Mihaescu, S.M. Khosla, E. Gutmark: Computational study of false vocal folds effects on unsteady airflows through static models of the human larynx. *Journal of Biomechanics* 48, 7 (2015) 1248–1257.
12. P. Hájek, P. Švancara, J. Horáček, J.G. Švec: Finite-element modeling of vocal fold self-oscillations in interaction with vocal tract: Comparison of incompressible and compressible flow model. *Journal of Applied and Computational Mechanics* 15, 12 (2021) 133–152.
13. R.C. Scherer, D. Shinwari, K.J. De Witt, C. Zhang, B.R. Kucinski, A.A. Afjeh: Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *Journal of the Acoustical Society of America* 109, 4 (2001) 1616–1630.
14. R.C. Scherer, S. Torkaman, B.R. Kucinski, A.A. Afjeh: Intraglottal pressures in a three-dimensional model with a non-rectangular glottal shape. *Journal of the Acoustical Society of America* 128, 2 (2010) 828–838.
15. S. Li, R.C. Scherer, M. Wan, S. Wang: The effect of entrance radii on intraglottal pressure distributions in the divergent glottis. *Journal of the Acoustical Society of America* 131, 2 (2012) 1371–1377.
16. S. Li, R.C. Scherer, L.P. Fulcher, X. Wang, L. Qiu, M. Wan, S. Wang: Effects of vertical glottal Duct Length on intraglottal pressures and phonation threshold pressure in the uniform glottis. *Journal of Voice* 32, 1 (2018) 8–22.
17. S. Li, R.C. Scherer, M. Wan, S. Wang, B. Song: Intraglottal pressure: a comparison between male and female larynxes. *Journal of Voice* 34, 6 (2020) 813–822.
18. X. Zhang, Y. Wang, W. Zhao, W. Wei, Z. Tao, H. Zhao: Vocal cord abnormal voice flow field study by modeling a bionic vocal system. *Advanced Robotics* 34, 1 (2020) 28–36.
19. S. Li, R.C. Scherer, M. Wan: Effects of vertical glottal duct length on intraglottal pressures in the convergent glottis. *Applied Sciences* 11, 10 (2021) 4535.
20. C.F. de Luzan, L. Oren, E. Gutmark, S.M. Khosla: Quantification of the intraglottal pressure induced by flow separation vortices using large eddy simulation. *Journal of Voice* 35, 6 (2021) 822–831.
21. X. Zheng, R. Mittal, S. Bielamowicz: A computational study of asymmetric glottal jet deflection during phonation. *Journal of the Acoustical Society of America* 129, 4 (2011) 2133–2143.
22. P. Šidlof, J. Horáček, V. Řidký: Parallel CFD simulation of flow in a 3D model of vibrating human vocal folds. *Computers & Fluids* 80 (2013) 290–300.
23. P. Šidlof, S. Zörner, A. Hüppe: A hybrid approach to the computational aeroacoustics of human voice production. *Biomechanics and Modeling in Mechanobiology* 14 (2015) 473–488.
24. S. Zörner, P. Šidlof, A. Hüppe, M. Kaltenbacher: Flow and acoustic effects in the larynx for varying geometries. *Acta Acustica united with Acustica* 102, 2 (2016) 257–267.
25. H. Sadeghi, S. Kniesburges, M. Kaltenbacher, A. Schützenberger, M. Döllinger: Computational models of laryngeal aerodynamics: potentials and numerical costs. *Journal of Voice* 33, 4 (2019) 385–400.
26. H. Sadeghi, S. Kniesburges, S. Falk, M. Kaltenbacher, A. Schützenberger, M. Döllinger: Towards a clinically applicable computational larynx model. *Applied Sciences* 9, 11 (2019) 2288.
27. H. Sadeghi, M. Döllinger, M. Kaltenbacher, S. Kniesburges: Aerodynamic impact of the ventricular folds in computational larynx models. *Journal of the Acoustical Society of America* 145, 4 (2019) 2376–2387.
28. S. Falk, S. Kniesburges, S. Schoder, B. Jakubaš, P. Maurerlehner, M. Echternach, M. Kaltenbacher, M. Döllinger: 3D-FV-FE aeroacoustic larynx model for investigation of functional based voice disorders. *Frontiers in Physiology* 12 (2021) 616985.
29. M. Lasota, P. Šidlof, M. Kaltenbacher, S. Schoder: Impact of the subgrid-scale turbulence model in aeroacoustic simulation of human voice. *Applied Sciences* 11, 4 (2021) 1970.
30. M. Lasota, P. Šidlof, P. Maurerlehner, M. Kaltenbacher, S. Schoder: Anisotropic minimum dissipation subgrid-scale model in hybrid aeroacoustic simulations of human phonation. *The Journal of the Acoustical Society of America* 153, 2 (2023) 1052–1063.
31. M. Mihaescu, S.M. Khosla, S. Murugappan, E.J. Gutmark: Vortex dipolar structures in a rigid model of the larynx at flow onset. *Journal of the Acoustical Society of America* 127, 1 (2010) 435–444.
32. N.E. Chisari, G. Artana, D. Sciamarella: Vortex dipolar structures in a rigid model of the larynx at flow onset. *Experiments in Fluids* 50 (2011) 397–406.
33. M.H. Farahani, J. Mousel, F. Alipour, S. Vigmostad: A numerical and experimental investigation of the effect of false vocal fold geometry on glottal flow. *Journal of Biomechanical Engineering* 135, 12 (2013) 1210061.
34. W. Mattheus, C. Brücker: Asymmetric glottal jet deflection: Differences of two- and three-dimensional models. *JASA-EL* 130 (6) (2011) EL373–EL379.
35. S. Zörner, M. Kaltenbacher, M. Döllinger: Investigation of prescribed movement in fluid-structure interaction simulation for the human phonation process. *Computers & Fluids* 86 (2013) 133–140.
36. Y. Jo, H. Ra, Y.J. Moon, M. Döllinger: Three-dimensional computation of flow and sound for human hemilarynx. *Computers & Fluids* 134–135 (2016) 41–50.
37. A.G. Siemens: Simcenter STAR-CCM+, 2023. <https://www.plm.automation.siemens.com/global/en/products/simcenter/STAR-CCM.html>.
38. J. Donea, A. Huerta, J-Ph Ponthot, A. Rodríguez-Ferran: Arbitrary Lagrangian-Eulerian methods, in: *The Encyclopedia of Computational Mechanics*, Vol. 1, John Wiley & Sons Ltd, 2004, pp. 414–437.

39. H. Hadzic: Development and application of finite volume method for the computation of flows around moving bodies on unstructured, overlapping grids. PhD thesis, Hamburg University of Technology, 2006.
40. J.L. Steger, F.C. Dougherty, J.A. Benek: A chimera grid scheme. [multiple overset body-conforming mesh system for finite difference adaptation to complex aircraft configurations]. *Advances in Grid Generation* (1983) 59–69.
41. K. Ishizaka: Fluid mechanical considerations of vocal cord vibration. Speech Communication Research Lab, Santa Barbara, 1972.
42. X. Pelorson, A. Hirschberg, R.R. Van Hassel, A.P.J. Wijnands, Y. Auregan: Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *Journal of the Acoustical Society of America* 96, 6 (1994) 3416–3431.
43. S. Schoder, M. Weitz, P. Maurerlehner, A. Hauser, S. Falk, S. Kniesburges, M. Döllinger, M. Kaltenbacher: Hybrid aeroacoustic approach for the efficient numerical simulation of human phonation. *Journal of the Acoustical Society of America* 147, 2 (2020) 1179–1194.
44. S. Schoder, P. Maurerlehner, A. Wurzinger, A. Hauser, S. Falk, S. Kniesburges, M. Döllinger, M. Kaltenbacher: Aeroacoustic sound source characterization of the human voice production-perturbed convective wave equation. *Applied Sciences* 11, 6 (2021) 2614.
45. P. Maurerlehner, S. Schoder, C. Freidhager, A. Wurzinger, A. Hauser, F. Kraxberger, S. Falk, S. Kniesburges, M. Echternach, M. Döllinger, M. Kaltenbacher: Efficient numerical simulation of the human voice. *e & i Elektrotechnik und Informationstechnik* 138, 3 (2021) 219–228.
46. S. Schoder, F. Kraxberger, S. Falk, A. Wurzinger, K. Roppert, S. Kniesburges, M. Döllinger, M. Kaltenbacher: Error detection and filtering of incompressible flow simulations for aeroacoustic predictions of human voice. *Journal of the Acoustical Society of America* 152, 3 (2022) 1425–1436.
47. S. Schoder, K. Roppert: openCFS: Open source finite element software for coupled field simulation – part acoustics, 2022. Arxiv preprint [arXiv:2207.04443](https://arxiv.org/abs/2207.04443).
48. M.J. Lighthill: On sound generated aerodynamically I. General theory. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 211, 1107 (1952) 564–587.
49. R. Ewert, W. Schröder: Acoustic perturbation equations based on flow decomposition via source filtering. *Journal of Computational Physics* 188, 2 (2003) 365–398.
50. P. Maurerlehner, S. Schoder, J. Tieber, C. Freidhager, H. Steiner, G. Brenn, K.H. Schäfer, A. Ennemoser, M. Kaltenbacher: Aeroacoustic formulations for confined flows based on incompressible flow data. *Acta Acustica* 6 (2022) 45.
51. S. Schoder: PCWE for FSAI–Derivation of scalar wave equations for fluid-structure-acoustics interaction of low Mach number flows. 2023. arXiv preprint [arXiv:2211.07490](https://arxiv.org/abs/2211.07490).
52. J. Piepiorka, O. von Estorff: Numerical investigation of hydrodynamic/acoustic splitting methods in finite volumes including rotating domains, in ICA 2019 23rd international congress on acoustics, Universitätsbibliothek der RWTH Aachen, 2019.
53. S. Schoder, M. Kaltenbacher, É. Spieser, H. Vincent, C. Bogey, C. Bailly: Aeroacoustic wave equation based on Pierce’s operator applied to the sound generated by a mixing layer, in 28th AIAA/CEAS Aeroacoustics 2022 Conference. 2022, 2896.
54. P. Sváček, J. Horáček: Finite element approximation of flow induced vibrations of human vocal folds model: Effects of inflow boundary conditions and the length of subglottal and supraglottal channel on phonation onset. *Applied Mathematics and Computation* 319 (2018) 178–194.
55. A. Yang, M. Stingl, D.A. Berry, J. Lohscheller, D. Voigt, U. Eysholdt, U. Döllinger, M. Döllinger: Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model. *Journal of the Acoustical Society of America* 130, 2 (2011) 948–964.
56. W. Jiang, C. Farbos de Luzan, X. Wang, L. Oren, S.M. Khosla, Q. Xue, X. Zheng: Computational modeling of voice production using excised canine larynx. *Journal of Biomechanical Engineering* 144, 2 (2022) 021003.
57. S.L. Smith, S.L. Thomson: Effect of inferior surface angle on the self-oscillation of a computational vocal fold model. *Journal of the Acoustical Society of America* 131, 5 (2012) 4062–4075.
58. M. Feistauer, P. Sváček, J. Horáček: Numerical simulation of fluid-structure interaction problems with applications to flow in vocal folds, in: T. Bodnár, G. Galdi, Š. Nečasová (Eds.), *Fluid-structure interaction and biomedical applications*, *Advances in Mathematical Fluid Mechanics*, Springer, Basel, 2014, pp. 321–393. https://doi.org/10.1007/978-3-0348-0822-4_5.
59. R. Mittal, G. Iaccarino: Immersed boundary methods. *Annual Review of Fluid Mechanics* 37, 1 (2005) 239–261.
60. H. Luo, R. Mittal, S.A. Bielamowicz: Analysis of flow-structure interaction in the larynx during phonation using an immersed-boundary method. *Journal of the Acoustical Society of America* 126, 2 (2009) 816–824.
61. J.H. Seo, R. Mittal: A high-order immersed boundary method for acoustic wave scattering and low-Mach number flow-induced sound in complex geometries. *Journal of Computational Physics* 230, 4 (2011) 1000–1019.
62. G. Link, M. Kaltenbacher, M. Breuer, M. Döllinger: A 2D finite-element scheme for fluid–solid–acoustic interactions and its application to human phonation. *Computer Methods in Applied Mechanics and Engineering* 198, 41–44 (2009) 3321–3334.
63. G.Z. Decker, S.L. Thomson: Computational simulations of vocal fold vibration: Bernoulli versus Navier-Stokes. *Journal of Voice* 21, 3 (2007) 273–284.
64. D.J. Daily, S.L. Thomson: Acoustically-coupled flow-induced vibration of a computational vocal fold model. *Computers & Structures* 116 (2013) 50–58.
65. B.H. Story: A parametric model of the vocal tract area function for vowel and consonant simulation. *Journal of the Acoustical Society of America* 117, 5 (2005) 3231–3254.
66. B.H. Story, I.R. Titze, E.A. Hoffman: Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America* 100, 1 (1996) 537–554.
67. W. Jiang, X. Zheng, Q. Xue: Computational modeling of fluid–structure–acoustics interaction during voice production. *Frontiers in Bioengineering and Biotechnology* 5 (2017) 7.
68. W. Jiang, Q. Xue, X. Zheng: Effect of longitudinal variation of vocal fold inner layer thickness on fluid-structure interaction during voice production. *Journal of Biomechanical Engineering* 140, 12 (2018) 1210081–1210089.
69. D. Bodaghi, Q. Xue, X. Zheng, S.L. Thomson: Effect of subglottic stenosis on vocal fold vibration and voice production using fluid–structure–acoustics interaction simulation. *Applied Sciences* 11, 3 (2021) 1221.
70. D. Bodaghi, W. Jiang, Q. Xue, X. Zheng: Effect of supraglottal acoustics on fluid-structure interaction during human voice production. *Journal of Biomechanical Engineering* 143, 4 (2021) 041010.
71. A. Yang, D.A. Berry, M. Kaltenbacher, M. Döllinger: Three-dimensional biomechanical properties of human vocal

- folds: parameter optimization of a numerical model to match in vitro dynamics. *Journal of the Acoustical Society of America* 131, 2 (2012) 1378–1390.
72. M. Döllinger, P. Gomez, R.R. Patel, C. Alexiou, C. Bohr, A. Schützenberger: Biomechanical simulation of vocal fold dynamics in adults based on laryngeal high-speed videendoscopy. *PloS One* 12, 11 (2017) e0187486.
 73. J. Neubauer, Z. Zhang, R. Miraghaie, D.A. Berry: Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds. *Journal of the Acoustical Society of America* 121, 2 (2007) 1102–1118.
 74. M. Triep, C. Brücker: Three-dimensional nature of the glottal jet. *Journal of the Acoustical Society of America* 127, 3 (2010) 1537–1547.
 75. L. Oren, S. Khosla, E. Gutmark: Intraglottal geometry and velocity measurements in canine larynges. *Journal of the Acoustical Society of America* 135, 1 (2014) 380–388.
 76. Z. Zhang, J. Neubauer: On the acoustical relevance of supraglottal flow structures to low-frequency voice production. *Journal of the Acoustical Society of America* 128, 6 (2010) EL378–EL383.
 77. M.H. Farahani, Z. Zhang: A computational study of the effect of intraglottal vortex-induced negative pressure on vocal fold vibration. *Journal of the Acoustical Society of America* 136, 5 (2014) EL369–EL375.
 78. A. Pirnia, E.A. Browning, S.D. Peterson, B.D. Erath: Discrete and periodic vortex loading on a flexible plate; application to energy harvesting and voiced speech production. *Journal of Sound and Vibration* 433 (2018) 476–492.
 79. B.Q. Kettlewell: The influence of intraglottal vortices upon the dynamics of the vocal folds. Master's thesis, University of Waterloo, 2015.
 80. Z. Li, Y. Chen, S. Chang, H. Luo: A reduced-order flow model for fluid–structure interaction simulation of vocal fold vibration. *Journal of Biomechanical Engineering* 142, 2 (2020) 0210051–02100510.
 81. Y. Chen, Z. Li, S. Chang, B. Rousseau, H. Luo: A reduced-order flow model for vocal fold vibration: From idealized to subject-specific models. *Journal of Fluids and Structures* 94 (2020) 102940.
 82. Z. Li, Y. Chen, S. Chang, B. Rousseau, H. Luo: A one-dimensional flow model enhanced by machine learning for simulation of vocal fold vibration. *Journal of the Acoustical Society of America* 149, 3 (2021) 1712–1723.
 83. T. Yoshinaga, Z. Zhang, A. Iida: Comparison of one-dimensional and three-dimensional glottal flow models in left-right asymmetric vocal fold conditions. *Journal of the Acoustical Society of America* 152, 5 (2022) 2557–2569.
 84. N. Ruty, X. Pelorson, A. Van Hirtum, I. Lopez-Arteaga, A. Hirschberg: An in vitro setup to test the relevance and the accuracy of low-order vocal folds models. *Journal of the Acoustical Society of America* 121, 1 (2007) 479–490.
 85. T. Kaburagi, Y. Tanabe: Low-dimensional models of the glottal flow incorporating viscous-inviscid interaction. *Journal of the Acoustical Society of America* 125, 1 (2009) 391–404.
 86. M.H. Farahani, Z. Zhang: Experimental validation of a three-dimensional reduced-order continuum model of phonation. *Journal of the Acoustical Society of America* 140, 2 (2016) EL172–EL177.
 87. J.E. Kelleher, T. Siegmund, M. Du, E. Naseri, R.W. Chan: The anisotropic hyperelastic biomechanical response of the vocal ligament and implications for frequency regulation: A case study. *Journal of the Acoustical Society of America* 133, 3 (2013) 1625–1636.
 88. A.K. Miri, H.K. Heris, U. Tripathy, P.W. Wiseman, L. Mongeau: Microstructural characterization of vocal folds toward a strain-energy model of collagen remodeling. *Acta Biomaterialia* 9, 8 (2013) 7957–7967.
 89. Z. Zhang: Structural constitutive modeling of the anisotropic mechanical properties of human vocal fold lamina propria. *Journal of the Acoustical Society of America* 145, 6 (2019) EL476–EL482.
 90. A. Terzolo, L. Bailly, L. Orgéas, T. Cochereau, N. Henrich-Bernardoni: A micro-mechanical model for the fibrous tissues of vocal folds. *Journal of the Mechanical Behavior of Biomedical Materials* 128 (2022) 105118.
 91. I.R. Titze, D.T. Talkin: A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *Journal of the Acoustical Society of America* 66, 1 (1979) 60–74.
 92. Q. Xue, X. Zheng, R. Mittal, S. Bielamowicz: Subject-specific computational modeling of human phonation. *Journal of the Acoustical Society of America* 135, 3 (2014) 1445–1456.
 93. L. Wu, Z. Zhang: Voice production in a MRI-based subject-specific vocal fold model with parametrically controlled medial surface shape. *The Journal of the Acoustical Society of America* 146, 6 (2019) 4190–4198.
 94. S. Chang, F.B. Tian, H. Luo, J.F. Doyle, B. Rousseau: The role of finite displacements in vocal fold modeling. *Journal of Biomechanical Engineering* 135, 11 (2013) 111008.
 95. Z. Zhang: Regulation of glottal closure and airflow in a three-dimensional phonation model: Implications for vocal intensity control. *Journal of the Acoustical Society of America* 137, 2 (2015) 898–910.
 96. Z. Zhang: Toward real-time physically-based voice simulation: An eigenmode-based approach, in *Proceedings of Meetings on Acoustics 173EAA(1)*. Acoustical Society of America (2017) 060002.
 97. Z. Zhang: Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *Journal of the Acoustical Society of America* 139, 4 (2016) 1493–1507.
 98. Z. Zhang: Effect of vocal fold stiffness on voice production in a three-dimensional body-cover phonation model. *Journal of the Acoustical Society of America* 142, 4 (2017) 2311–2321.
 99. Z. Zhang: Contribution of laryngeal size to differences between male and female voice production. *Journal of the Acoustical Society of America* 150, 6 (2021) 4511–4521.
 100. L. Wu, Z. Zhang: Impact of the paraglottic space on voice production in an MRI-based vocal fold model. *Journal of Voice* (2021) <https://doi.org/10.1016/j.jvoice.2021.02.021>.
 101. Z. Zhang, T. Hieu Luu: Asymmetric vibration in a two-layer vocal fold model with left-right stiffness asymmetry: Experiment and simulation. *Journal of the Acoustical Society of America* 132, 3 (2012) 1626–1635.
 102. P. Gómez, A. Schützenberger, S. Kniesburges, C. Bohr, M. Döllinger: Physical parameter estimation from porcine ex vivo vocal fold dynamics in an inverse problem framework. *Biomechanics and Modeling in Mechanobiology* 17, 3 (2018) 777–792.
 103. M. Kunduk, M. Döllinger, A. McWhorter, J. Lohscheller: Assessment of the variability of vocal fold dynamics with and between recordings with high-speed imaging and by Phonovibrogram. *Laryngoscope* 120, 5 (2010) 981–987.
 104. M. Döllinger, T. Braunschweig, J. Lohscheller, U. Eysholdt, U. Hoppe: Normal voice production: computation of driving parameters from endoscopic digital high speed images. *Methods of Information in Medicine* 42 (2003) 271–276.
 105. P.J. Hadwin, S.D. Peterson: An extended Kalman filter approach to non-stationary Bayesian estimation of reduced-order vocal fold model parameters. *Journal of the Acoustical Society of America* 141, 4 (2017) 2909–2920.

106. P. Gomez, M. Semmler, C. Bohr, M. Döllinger: Laryngeal pressure estimation with a recurrent neural network. *IEEE Journal of Translational Engineering in Health and Medicine* 7 (2019) 8590726.
107. R.R. Patel, D. Dubrovskiy, M. Döllinger: Characterizing vibratory kinematics in children and adults with high-speed digital imaging. *Journal of Speech, Language, and Hearing Research* 57, 2 (2014) 674–686.
108. A.M. Kist, J. Zilker, P. Gómez, A. Schützenberger, M. Döllinger: Rethinking glottal midline detection. *Scientific Reports* 10, 1 (2020) 20723.
109. Z. Li, S. Chang, B. Rousseau, H. Luo: A one-dimensional flow model enhanced by machine learning for simulation of vocal fold vibration. *Journal of the Acoustical Society of America* 149, 3 (2021) 1712–1723.
110. Y. Zhang, W. Jiang, L. Sun, J. Wang, X. Zheng, Q. Xue: A deep-learning based generalized empirical flow model of glottal flow during normal phonation. *Journal of Biomechanical Engineering* 144, 9 (2022) 091001.
111. Y. Zhang, T. Pu, C. Zhou, H. Cai: An improved glottal flow model based on Seq2Seq LSTM for simulation of vocal fold vibration. *Journal of Voice* (2022) S0892–1997(22).
112. Z. Zhang: Estimation of vocal fold physiology from voice acoustics using machine learning. *Journal of the Acoustical Society of America* 147(3), (2020)EL264–EL270.
113. Z. Zhang: Voice feature selection to improve performance of machine learning models for voice production inversion. *Journal of Voice* (2021). <https://doi.org/10.1016/j.jvoice.2021.03.004>.
114. Z. Zhang: Estimating subglottal pressure and vocal fold adduction from the produced voice in a single-subject study (L). *Journal of the Acoustical Society of America* 151, 2 (2022) 1337–1340.
115. I.R. Titze, J.C. Lucero: Voice simulation: the next generation. *Applied Sciences* 12, 22 (2022) 11720.
116. R. Schwarze, W. Mattheus, J. Klostermann, C. Brücker: Starting jet flows in a three-dimensional channel with larynx-shaped constriction. *Computers & Fluids* 48 (2011) 68–83.
117. S.L. Thomson, L. Mongeau, S.H. Frankel: Aerodynamic transfer of energy to the vocal folds. *Journal of the Acoustical Society of America* 118, 3 (2005) 1689–1700.
118. B.H. Story, I.R. Titze: Voice simulation with a body-cover model of the vocal folds. *Journal of the Acoustical Society of America* 97, 2 (1995) 1249–1260.

Cite this article as: Döllinger M. Zhang Z. Schoder S. Šidlof P. Tur B, et al. 2023. Overview on state-of-the-art numerical modeling of the phonation process. *Acta Acustica*, 7, 25.