DERIVING VOCAL FOLD OSCILLATION INFORMATION FROM RECORDED VOICE SIGNALS USING MODELS OF PHONATION

Wayne Zhao, Rita Singh

Department of Electrical and Computer Engineering and School of Computer Science
Carnegie Mellon University, Pittsburgh, USA

ABSTRACT

The self-sustained motion of vocal folds during phonation is resultant from an intricate balance of bio-mechanical and aerodynamic forces across the glottis, predicated on the physical properties of the vocal folds. Estimating the bio-parameters of the vocal folds, and characterizing their motion is extremely important in diagnostic settings. This is traditionally done through actual measurements of the vocal fold properties, and videostroboscopic observations of their motion in clinical settings. Over the past decades, several mathematical models of phonation have been proposed that comprise dynamical systems with parameters that correspond to the physical properties of the vocal folds, and solutions that emulate their self-sustained motion. However, these models cannot reproduce the glottal airflow dynamics of individual speakers, unless their parameters are physically set to match those of the speakers, which again must be currently measured in clinical settings. We propose a methodology to deduce the parameters of two such models for individual speakers using recorded speech. This allows for their solutions to closely approximate each speaker's vocal fold motions in a customized manner. Further machine-learning based analysis of these solutions can reveal patterns that are discriminative for the underlying influences on the speaker's vocal folds, allowing for the construction of effective computational diagnostic aids from voice. We demonstrate the viability of our proposed methodology by using it for the deduction of various vocal pathologies from voice signals.

Index Terms— Vocal fold oscillation models, phonation models, parameter estimation, voice pathologies, voice profiling

1. INTRODUCTION

Phonation is a complex bio-mechanical process wherein the glottal airflow, mediated by the muscles in the larynx and driven by an intricate balance of aerodynamic and mechanical forces across the glottis, maintains the vocal folds in a state of self-sustained vibrations [1, 2]. During this process, the eigenmodes of vibration of both folds synchronize, or strive to do so, depending on the state of the vocal folds. Minor perturbations in their bio-physical or bio-mechanical characteristics,

which may be caused by myriad influences, alter this motion. In applications such as voice-based diagnostic aids, estimating these alterations and deducing the nature of underlying perturbations or bio-parameters is important. However, this is difficult to do on an individual basis using traditional clinical means that must measure the properties of vocal folds and videostroboscopically observe their motion.

The primary focus of this paper is to help overcome this problem by solving it through computational means, using physical models of phonation and recorded voice signals, thus alleviating the need for taking physical measurements of vocal fold motion. To explain the premises of the proposed solution, we first briefly review the process of phonation, and the general approaches to phonation modeling.

1.1. The bio-mechanical process of phonation

By the myoelastic-aerodynamic theory of phonation, the forces in the laryngeal region that initiate and maintain it relate to (a) pressure balances and airflow dynamics within the supra-glottal and sub-glottal regions and (b) muscular control within the glottis and the larynx. The balance of forces necessary to cause self-sustained vibrations during phonation is created by two physical phenomena: the Bernoulli effect and the Coandă effect. Figure 1 illustrates the interaction between these effects that drives the oscillations of the vocal folds.

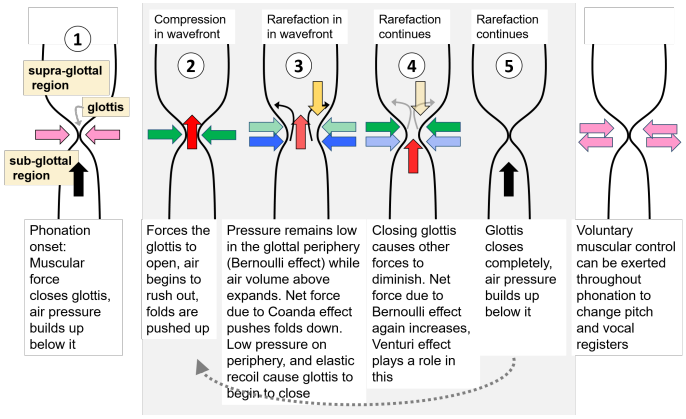


Fig. 1. Schematic of the balance of forces through one cycle of the self-sustained vibrations of the vocal folds. The color codes for the arrows depict net forces due to the following: Pink–muscular; Green–Bernoulli effect; Yellow–Coandă effect; Blue–vocal fold elasticity and other factors; Black and Red–air pressure. Lighter shades of each color depict smaller forces. Figure from [3] with permission.

The process of phonation begins with the closing of the glottis. This closure is voluntary and facilitated by the laryngeal muscles. Once closed, the muscles do not actively play a role in sustaining the vibrations. Glottal closure is followed by a contraction of the lungs which pushes out air and causes an increase in pressure just below the glottis. When this subglottal pressure crosses a threshold, the vocal folds are pushed apart, and air rushes out of the narrow glottal opening

into the much wider supra-glottal region, creating negative intra-glottal pressure (with reference to atmospheric air pressure) [3].

From the airflow perspective, the glottis thus forms a flow separation plane. The air expansion in this region and the low pressure created in the vicinity of the glottis through the Coandă effect induced entrainment cause a lowering of pressure close to the glottis and a net downward force on the glottis. At the same time, lowered pressure in the glottal region due to the Bernoulli effect that ensues from the high-velocity air volume flow through the glottis exerts a negative force on the glottis. The negative Bernoulli pressure causes elastic recoil, causing it to begin to close again. The closing reduces the volume flow through the glottis, diminishing the downward forces acting on it. Increased pressure buildup in the sub-glottal region causes the glottis to open again. This chain of oscillations continues in a self-sustained fashion throughout phonation until voluntary muscle control intervenes to alter or stop it or as the respiratory volume of air in the lungs is exhausted. The exact physics of the airflow through the glottis during phonation is well studied, e.g., [4, 5, 2, 6, 7, 8].

1.2. General approaches to phonation modeling

Physical models of phonation, e.g. [5, 9, 10, 11, 12, 13, 14, 3], attempt to explain this complex physical process using relations derived from actual physics, especially aerodynamics and the physics of mechanical structures.

For modeling purposes, we note that phonation is not the only source of excitation of the the vocal tract in producing *speech* sounds, which comprise both voiced and unvoiced sounds. However, phonation is indeed the primary source of excitation of the vocal tract in the production of *voiced* sounds, wherein the oscillation of the vocal folds modulates the pressure of the airflow to produce a (quasi-) periodic glottal flow wave at a fundamental frequency (the pitch), which in turn results in the occurrence of higher order harmonics. The resultant glottal flow further excites the vocal tract, which comprises the laryngeal cavity, the pharynx, and the oral and nasal cavities, to produce individual sounds. The vocal tract serves as a resonance chamber that produces formants. The identities of the different sounds produced within it are derived from these resonances, which in turn are largely dependent on the configurations of the vocal tract specified by their time-varying cross-sectional area.

From this perspective, phonation modeling has typically involved the modeling of two sub-processes: the self-sustained vibration of the vocal folds, and the propagation of the resultant pressure wave through the vocal tract [15]. Each sub-process model has associated parameters that determine the model output, given an input.

Depending on the level of approximations made and following the division of the process from the perspective mentioned above, models of phonation are of two kinds: *vocal folds models* (or *vocal folds oscillation models*, or *oscillation models*), and *vocal tract models*. The **vocal folds models** describe the vibration of vocal folds and their aerodynamic interaction with airflow. Such models are of four broad types: one-mass models e.g. [2, 16, 17, 10, 18], two-mass models e.g. [5, 9], multi-mass models [12], and finite element models [11]. Each of these has proven to be useful

in different contexts. On the other hand, the **vocal tract models** describe the interaction of the glottal pressure wave with the vocal tract, which is turn has been described in the literature by varied models, such as statistical models [19], geometric models [20], biomechanical models [21], etc. In addition, in order to describe the aero-acoustic interaction of the glottal airflow and the vocal tract, different models are applied – such as reflection type line analog models, transmission line circuit analog models [22], hybrid time-frequency domain models [23], finite-element models [24], etc.

1.3. The problem of parameter estimation

Each of the models includes a set of parameters that determine its *state*, and *output*, given an input. For instance, given the parameters for an oscillation model, the glottal flow waveform can be determined; given the glottal flow waveform as input, and the parameters for vocal tract model, the acoustic signal can be determined.

Such determinations have great practical use. For example, with speaker-appropriate parameter setting, the output of these models can be used as a proxy for the actual vocal fold motion, and vocal fold properties of individual speakers. This opens the doors to machine-learning based analysis of the model solutions, which can be used to automatically deduce underlying pathologies and other bio-parameters on a speaker-specific basis. As an example of the latter, the parameters of vocal fold oscillation models (such as the asymmetry parameter) can be used to discriminate between different types of voice disorders, since these are largely attributed to the asymmetry of vocal folds vibration [25]. Each choice of model parameters leads to a unique characterization of the state space of the dynamic system that comprises the model, within which the model behavior can be observed and described. Model behaviors may include ordered or chaotic behaviors, whose stability can be specified through entities such as Lyapunov exponents. Model behaviors can be matched to, or used to explain the signals observed in various voice disorders, and thus can be used to characterize them. Thus there are strong advantages of being able to estimate the parameters of these models in an individualized manner using simpler and non-physical means.

The estimation of model parameters from the model output is commonly termed as the *inverse problem*. Thus, while the models themselves are extremely useful in understanding the complex dynamics of the phonation process, and allow us to analyze various phonation-related phenomena that are observed during speech production, their use is limited to this. The inverse problem of the estimation of parameters of these models, though potentially highly useful in other aspects as discussed above, is quite difficult to solve. For example, in order to estimate the parameters of a vocal tract model, one must take into account the vocal tract coupling, the effect of the lossy medium that comprises the walls of the vocal tract, lip radiation, etc. Without serious approximations, the inverse problem in this case is eventually intractable. To get around these requirements, the ideal way is to obtain physical measurements of the vocal fold oscillations, and the glottal flow using high-speed videostroboscopy and other techniques such as physical or computer simulation. These are not always feasible. Other approaches simplify the solution by discretizing the vocal tract as a sequence of consecutive tubes of varying cross-sectional area, or with a mesh-grid. However, these

approximations invariably increase the estimation error. Both of these conventional workarounds have shortcomings – the former approach is restricted when direct measurements are unavailable, and the latter is subject to serious approximation errors.

This paper addresses this problem and provides a methodology for solving it through purely computational means. In the sections that follow, we will briefly review our previously proposed Adjoint Least Squares Estimation (ADLES) algorithm [26] to estimate the parameters of a vocal oscillation model from voice recordings. In this approach, the parameters are estimated by an iterative process that minimizes the error between an estimated glottal flow waveform and one generated through the physical model.

We then describe our proposed algorithm to estimate the parameters of a vocal tract model (or body-cover model). In this algorithm, the estimation of the glottal flow waveform is not required, and the voice signal can be directly used as reference. The algorithm proposed iteratively re-estimates the parameters by minimizing the error between the reference voice sample, and the waveform generated by the model. We call this algorithm the VTMPE-FB algorithm, standing for "Vocal Tract Model Parameter Estimation – Forward-Backward" algorithm for the estimation of the parameters of the full body-cover model and thereby, the motion of the vocal folds.

2. VOCAL FOLDS, VOCAL TRACT AND JOINT MODELS

In this section we describe the formulation of the vocal folds oscillation and vocal tract models, which we then combine into a single model. The proposed ADLES-VFT algorithm is based on the joint model.

We begin with a schematic illustration of the phonation process. This is given in Figure 2.

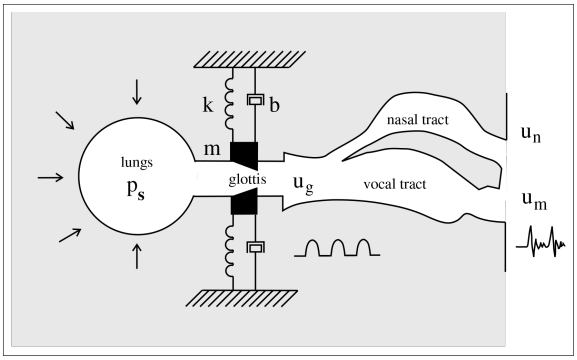


Fig. 2. Illustration of the phonation process. Airflow from the lungs, driven by the subglottal pressure P_s, passes through the glottis, and vocal folds are set into a state of self-sustained vibration, producing the glottal flow u_g which is a quasi-periodic pressure wave. The vibration of vocal folds is analogous to a pair of mass-spring-damper oscillators. Further, the glottal flow resonates in the speaker's vocal tract and nasal tract and produces voiced sound.

2.1. One-mass models

One-mass models describe the vibration of the vocal folds as that of a single mass-damper-spring oscillator:

$$M\ddot{x} + B\dot{x} + Kx = f(x, \dot{x}, t) \tag{1}$$

where x is lateral displacement of a mass M, B and K are damping and stiffness coefficients respectively, f is the driving force, and t is time [2]. The driving force is velocity-dependent and can be estimated by Bernoulli's energy law:

$$P_g = P_s - \frac{1}{2}\rho v^2 \tag{2}$$

where P_g is the mean glottal pressure, P_s is sub-glottal pressure, ρ is air density, and v is the air particle velocity. The kinetic pressure in the supra-glottal region is neglected [2].

2.2. Two-mass models

The two-mass models describe vocal fold motion as two coupled mass-damper-spring oscillators

$$M_1\ddot{x}_1 + B_1\dot{x}_1 + K(x_1 - x_2) + R_1 = F_1$$
$$M_2\ddot{x}_2 + B_2\dot{x}_2 + K(x_2 - x_1) + R_2 = F_2$$

where x_i, M_i, B_i are the i-th oscillator's displacement, mass, and viscous damping coefficient, K is the coupling stiffness between the two masses, F_i is the driving force, and Ri is the elastic restoring force [9]. This model assumes (1) small air inertia and quasi-steady glottal flow, (2) negligible supra-glottal pressure, and (3) that the nonlinearity induced by vocal fold collision is small. These assumptions lead to small-amplitude oscillations and model simplification [9].

2.3. Multi-mass models

Multi-mass models have a greater degrees of freedom and hence can model vocal fold motion with high precision. They are based mass-spring-damper motion dynamics which are widely used in multiple problem settings (e.g. []). For the i-th mass component, the equation of motion is:

$$M_i\ddot{\boldsymbol{x}}_i = \boldsymbol{F}_i^A + \boldsymbol{F}_i^V + \boldsymbol{F}_i^L + \boldsymbol{F}_i^C + \boldsymbol{F}_i^D \tag{3}$$

where $\boldsymbol{x}_i = (x_i, y_i, z_i)$ is the three-dimensional displacement, M_i is the mass, $\boldsymbol{F}_i^A$ is the anchor force associated with the anchor spring and damper, $\boldsymbol{F}_i^V$ and $\boldsymbol{F}_i^L$ are the vertical and longitudinal coupling forces associated with spring and damping, $\boldsymbol{F}_i^C$ is the collision restoring force, and $\boldsymbol{F}_i^D$ is the driving force from glottal pressure [12]. In [12], 50 masses are used.

2.4. Finite element models

Finite element models discretize the vocal fold motion in space and time – the geometry of the vocal fold is discretized into small elements (cells). In each cell, the applicable differential equation governed by the law of physics is solved. These models can handle complex geometries, continuous deformation, and complex driving forces [11].

Consider a cube element with six stress and strain components. By the principles of elasticity in mechanics we have:

$$\boldsymbol{\sigma} = \boldsymbol{S}\boldsymbol{\epsilon} \tag{4}$$

where $\boldsymbol{\sigma}$ is the stress tensor, $\boldsymbol{\epsilon}$ is the strain tensor, and $\boldsymbol{S}$ is the stiffness matrix consisting of Young's modulus, shear modulus, and Poisson's ratio [11]. The relation between stress and displacement is governed by:

$$\sigma_x = C_1\mu\frac{\partial u}{\partial x} + C_2\mu\frac{\partial w}{\partial z} \tag{5}$$

$$\sigma_z = C_2\mu\frac{\partial u}{\partial x} + C_1\mu\frac{\partial w}{\partial z} \tag{6}$$

$$\tau_{xy} = \mu'\frac{\partial u}{\partial y} \tag{7}$$

$$\tau_{yz} = \mu'\frac{\partial w}{\partial y} \tag{8}$$

$$\tau_{zx} = \mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right) \tag{9}$$

where τ is the shear stress, u and w are the lateral and vertical components of the displacement vector, μ and μ' are shear moduli, and C_1 and C_2 are constants [11]. This system of partial differential equations can be efficiently solved by finite element methods.

2.5. Vocal tract models

As mentioned earlier, vocal tract models are of various types.

Statistical models model the vocal tract as statistical factors or components. For instance, factor analysis describes the vocal tract profile as a sum of articulatory components and analyzes the relationship between individual or combination of components and vocal tract parameters [19].

Geometric models attempt to depict the shape and geometric configurations of the vocal tract. They specify articulatory state with vocal tract parameters that define the position and shape of tongue, lips, jaw, larynx, etc [20]. However, such models are not scalable because they do not account for the continuous variations of the anatomy and articulatory state, require clinical measurements such as from magnetic resonance imaging, and are not amendable to coupling with vocal fold models.

Bio-mechanical models simulate the geometry and articulatory movements of the vocal tract using displacement-based finite element methods and take into account the continuous tissue deformation and variation of the physiological, biomechanical, and viscoelastic properties of muscles [21]. They are more scalable and accurate, and allow for the modeling of more fine-grained control over muscular forces, articulator positions, and movements.

To study the interaction between vocal folds and vocal tract, modeling approaches often take analog approaches in the digital circuit regime, and model the propagation of glottal flow in the vocal tract as a transmission line circuit [22]. One can evaluate the system (vocal tract)'s transfer function in time and frequency domain and acquire the system output in response to the input (glottal flow) [23].

In this paper, we take a different approach. We unite the vocal fold and tract models into a single model, calling it a joint vocal fold-tract (JVFT) model. We present this model in the setting of the proposed solution to the inverse problem of estimating its parameters in a later section, for better continuity. In the next section, we explain the key concepts needed to solve the inverse problem of model parameter estimation. Following that, we will describe the solution to the inverse problem in the case of a vocal folds models, and finally move on to describing the JVFT model, and the proposed solution to its inverse problem. This is followed by a section describing experimental results.

3. ESTIMATION OF MODEL PARAMETERS: THE INVERSE PROBLEM

Both vocal folds and vocal tract models attempt to accurately represent the actual dynamics of phonation. Their solutions are therefore flows in phase space that are also (by proxy) governed by various bio-mechanical parameters of the vocal folds such as elasticity, resistance, Young's modulus, viscosity, etc., as well as the configurations of vocal tract such as time-varying cross-sectional area. While these models effectively solve the *forward* problem of accurately emulating vocal fold and vocal tract dynamics during phonation, the *inverse* problem of finding the correct model parameters given a set of observed speech signals had not been addressed until recently [26].

The inverse problem is challenging to solve in real-life settings, especially in the case of the more accurate coupling of vocal folds - vocal tract models. The inverse problem in fact becomes intractable in many cases. For example, to estimate the parameters for the vocal folds oscillation model, one needs to consider the vocal tract coupling, the effect of lossy medium and lip radiation, and many other factors. Two broad categories of approaches are used in these settings: one approach is to isolate and only examine the vocal fold model. For this, however, one must acquire measurements of the vocal fold displacements. This in turn requires either high-speed photography [27] or physical or numerical simulations [11, 28], which are often not easily accessible. Even with the measurements, solving the inverse problem remains hard [29]. It is usually solved via iterative matching procedures [30, 31, 32], stochastic optimization, or heuristic procedures [33, 12]. The second (alternative) category of approaches attempts to discretize the vocal tract with consecutive, cross-sectional area varying tubes or with a mesh-grid [34, 35], simplifying the solution. However, such approximation increases the estimation error.

3.1. Forward and Backward Approaches for Inverse problems

To address the problems inherent in conventional approaches to solving such inverse problems, we propose a solution framework incorporating a backward approach and a forward approach.

3.1.1. The backward approach

The backward approach simply eliminates the need for a vocal tract model by estimating the glottal flow from speech signals via inverse filtering. The model solutions are iteratively compared to the glottal flow waveform, and the model parameters are iteratively optimized to minimize the error between the two. For this, in the next section we describe the adjoint least-squares (ADLES) method [26] to effectively solve an ODE-constrained functional minimization problem in the context of a specific asymmetric vocal folds model, to accurately estimate its parameters.

3.1.2. The forward approach

The forward approach combines the vocal folds oscillation model and the vocal tract propagation model. Our solution is proposed in the context where the vocal folds oscillation model is a one-mass model with asymmetry parameters described by coupled ODEs, and the vocal tract model is an acoustic wave propagation model described by PDEs. However, the framework and approach apply to all models in these categories (model-specific derivations may be needed, though). When combined, these two selected models accurately represent phonation for both normal and disordered voices. The solution we propose is an iterative adjoint method to solve the ODE/PDE constrained inverse problem in this case. It enables the estimation of model parameters *directly* from speech recordings (or speech waveforms, without requiring the estimation of the glottal flow waveform). Since the model is now extended represents the process (and fine-grained nuances) of phonation more accurately.

4. SOLVING THE INVERSE PROBLEM FOR A VOCAL FOLDS MODEL

In this section, we present the backward approach to estimate the parameters of a vocal fold model in detail. While the algorithm itself has been briefly mentioned in earlier literature, we give a more complete description with fuller details below. This is also necessary to build up the solution for the inverse problem in the case of the joint vocal fold-tract model, presented in the next section.

For this, we choose a well-studied model that captures the asymmetric movements of the vocal folds during phonation. Such a model is especially useful in the discrimination of vocal fold pathologies, since they tend to affect the motion of the vocal folds in an asymmetric (idiosyncratic) manner in most cases.

4.1. The asymmetric vocal folds oscillation model

We use the the one-mass asymmetric body-cover model illustrated in Figure 3.

This model incorporates an asymmetry parameter, which can emulate the asymmetry in the vibratory motions of left and right vocal folds, hence is also ideally suited to modeling pathological phonation [36].

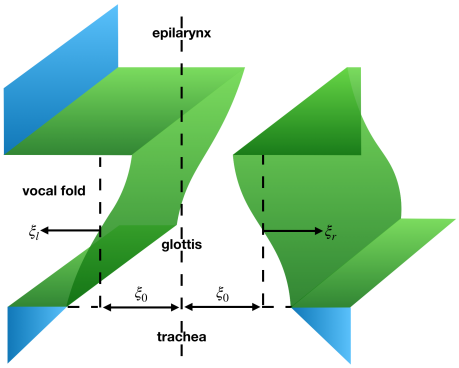


Fig. 3. Diagram of the one-mass body-cover model for vocal folds. The lateral displacements at the midpoint of the left and right vocal folds are denoted as ξ_l and ξ_r, and ξ_0 represents the half glottal width at rest.

The key assumptions made in formulating this model are:

(a) The degree of asymmetry is independent of the oscillation frequency;

(b) The glottal flow is stationary, frictionless, and incompressible;

(c) All subglottal and supraglottal loads are neglected, eliminating the effect of source-vocal tract interaction;

(d) There is no glottal closure and hence no vocal fold collision during the oscillation cycle;

(e) The small-amplitude body-cover assumption is applicable (see definition below).

Assumption 4.1 (Body-cover assumption)**.** The body-cover assumption assumes that a glottal flow-induced mucosal wave travels upwards within the transglottal region, causing a small displacement of the mucosal tissue, which attenuates down within a few millimeters into the tissue as an energy exchange happens between the airstream and the tissue [2].

This assumption allows us to represent the mucosal wave as a one-dimensional surface wave on the mucosal surface (the cover) and treat the remainder of the vocal folds (the body) as a single mass or safely neglect it. As a result, the oscillation model can be linearized, and the oscillatory conditions are much simplified while maintaining the model's accuracy.

We adopt the specific formulation for the one-mass asymmetric model from [18]. Referring again to Figure 3, we denote the center-line of the glottis as the z-axis. At the midpoint ($z = 0$) of

the thickness of the vocal folds, the left and right vocal folds oscillate with lateral displacement ξ_l and ξ_r, resulting in a pair of coupled Van der Pol oscillators:

$$\begin{aligned}\ddot{\xi}_r + \beta(1+\xi_r^2)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r &= \alpha(\dot{\xi}_r + \dot{\xi}_l)\\ \ddot{\xi}_l + \beta(1+\xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l &= \alpha(\dot{\xi}_r + \dot{\xi}_l)\end{aligned} \tag{10}$$

where β is the coefficient incorporating mass, spring and damping coefficients, α is the glottal pressure coupling coefficient, and Δ is the asymmetry coefficient. For a male adult with normal voice, the reference values for the model parameters (from clinical measurements) are usually approximately set to $\alpha = 0.5$, $\beta = 0.32$ and $\Delta = 0$.

The inverse problem of estimating the parameters such models has been approached in some studies via iterative matching procedures [30, 31, 32], stochastic optimization or heuristic procedures [33, 12].

We have proposed an adjoint least-squares (ADLES) method [26] for this. We explain the method in more detail below. Firstly, our objective is formulated as follows:

The vibration of vocal folds oscillates the air particles at the glottal region, producing a pressure wave that propagates through the upper vocal channel into the open air. The acoustic pressure $p_L(t) \coloneqq p(L,t)$, which represents the speech signal measured by a microphone near the mouth, is a result of the pressure wave $p_0(t) \coloneqq p(0,t)$ at the glottis modulated by the upper vocal channel. If we denote the effect of the upper vocal channel as a filter

$$\mathcal{F} : L^2(T) \to L^2(T) \tag{11}$$

$$p_0(t) \mapsto p_L(t) \tag{12}$$

we can deduce $p_0(t)$ from $p_L(t)$ using inverse filtering [37]

$$p_0(t) = \mathcal{F}^{-1}(p_L(t)) \tag{13}$$

Let $A(x)$ be the area function of the vocal channel for $x \in [0, L]$ and $A(0)$ represent the cross-sectional area at the glottis. The corresponding volume velocity $u_0(t)$ through the vocal channel is given by

$$u_0(t) = \frac{A(0)}{\rho c} p_0(t) \tag{14}$$

where c is the speed of sound, and ρ is the ambient air density. As a result, given a measured speech signal $p_m(t)$, we have:

$$u_0^m(t) = \frac{A(0)}{\rho c} \mathcal{F}^{-1}(p_m(t)) \tag{15}$$

Alternatively, we can derive $u_0(t)$ from the displacement of vocal folds as

$$u_0(t) = \tilde{c} d \left(2\xi_0 + \xi_l(t) + \xi_r(t)\right) \tag{16}$$

where ξ_0 is the half glottal width at rest and is set to 0.1 cm, d is the length of vocal fold and is set to 1.75 cm, and $\tilde{c}$ is the air particle velocity at the midpoint of the vocal fold. Our objective is then to minimize the difference:

$$\min \int_0^T \left(u_0(t) - u_0^m(t)\right)^2 dt \Leftrightarrow \tag{17}$$

$$\min \int_0^T \left(\tilde{c}d\left(2\xi_0 + \xi_l(t) + \xi_r(t)\right) - \frac{A(0)}{\rho c}\mathcal{F}^{-1}(p_m(t))\right)^2 dt \tag{18}$$

such that

$$\ddot{\xi}_r + \beta(1+\xi_r^2)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{19}$$

$$\ddot{\xi}_l + \beta(1+\xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{20}$$

$$\xi_r(0) = C_r \tag{21}$$

$$\xi_l(0) = C_l \tag{22}$$

$$\dot{\xi}_r(0) = 0 \tag{23}$$

$$\dot{\xi}_l(0) = 0 \tag{24}$$

where C_r and C_l are constants.

4.2. The Adjoint Least Squares (ADLES) Solution

To solve the functional least squares in (18), we require the gradients of (18) w.r.t. the model parameters α, β and Δ. Subsequently, we can adopt any gradient-based (local or global) method to obtain the solution.

Considering the residual

$$R = \tilde{c}d\left(2\xi_0 + \xi_l(t) + \xi_r(t)\right) - \frac{A(0)}{\rho c}\mathcal{F}^{-1}(p_m(t)) \tag{25}$$

We have

$$f(\xi_l, \xi_r; \vartheta) = R^2 \tag{26}$$

and

$$F(\xi_l, \xi_r; \vartheta) = \int_0^T f(\xi_l, \xi_r; \vartheta) dt \tag{27}$$

where $\vartheta = [\alpha, \beta, \Delta]$. We define the Lagrangian:

$$\begin{aligned}
&\mathcal{L}(\vartheta) = \\
&\int_0^T \Bigg[f + \lambda \left(\ddot{\xi}_r + \beta \left(1 + \xi_r^2\right) \dot{\xi}_r + \xi_r - \frac{\Delta}{2} \xi_r - \alpha \left(\dot{\xi}_r + \dot{\xi}_l \right) \right) \\
&+ \eta \left(\ddot{\xi}_l + \beta \left(1 + \xi_l^2\right) \dot{\xi}_l + \xi_l + \frac{\Delta}{2} \xi_l - \alpha \left(\dot{\xi}_r + \dot{\xi}_l \right) \right) \Bigg] dt \\
&+ \mu_l \left(\xi_l(0) - C_l \right) + \mu_r \left(\xi_r(0) - C_r \right) + \nu_l \dot{\xi}_l(0) + \nu_r \dot{\xi}_r(0)
\end{aligned} \tag{28}$$

where λ, η, μ and ν are Lagrangian multipliers. Taking the derivative of the Lagrangian w.r.t. the model parameter α yields:

$$\begin{aligned}
\mathcal{L}_\alpha = \int_0^T \Bigg[& 2\tilde{c} dR \left(\partial_\alpha \xi_l + \partial_\alpha \xi_r \right) \\
& + \lambda \Bigg(\partial_\alpha \ddot{\xi}_r + 2\beta \dot{\xi}_r \xi_r \partial_\alpha \xi_r + \beta \left(1 + \xi_r^2\right) \partial_\alpha \dot{\xi}_r \\
& + \partial_\alpha \xi_r - \frac{\Delta}{2} \partial_\alpha \xi_r - \alpha \left(\partial_\alpha \dot{\xi}_r + \partial_\alpha \dot{\xi}_l \right) - \left(\dot{\xi}_r + \dot{\xi}_r \right) \Bigg) \\
& + \eta \Bigg(\partial_\alpha \ddot{\xi}_l + 2\beta \dot{\xi}_l \xi_l \partial_\alpha \xi_l + \beta \left(1 + \xi_l^2\right) \partial_\alpha \dot{\xi}_l \\
& + \partial_\alpha \xi_l + \frac{\Delta}{2} \partial_\alpha \xi_l - \alpha \left(\partial_\alpha \dot{\xi}_r + \partial_\alpha \dot{\xi}_l \right) - \left(\dot{\xi}_r + \dot{\xi}_r \right) \Bigg) \Bigg] dt \\
& + \mu_l \partial_\alpha \xi_l(0) + \mu_r \partial_\alpha \xi_r(0) + \nu_l \partial_\alpha \dot{\xi}_l(0) + \nu_r \partial_\alpha \dot{\xi}_r(0)
\end{aligned} \tag{29}$$

Integrating the term $\lambda \partial_\alpha \ddot{\xi}_r$ twice by parts yields:

$$\int_0^T \lambda \partial_\alpha \ddot{\xi}_r dt = \int_0^T \partial_\alpha \xi_r \ddot{\lambda} dt - \partial_\alpha \xi_r \dot{\lambda} \Big|_0^T + \partial_\alpha \dot{\xi}_r \lambda \Big|_0^T \tag{30}$$

Applying the same to $\eta \partial_\alpha \ddot{\xi}_l$, substituting into (29) and simplifying the final expression we

obtain:

$$\begin{aligned}\mathcal{L}_\alpha = \int_0^T \Bigg[& \left(\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2} \right) \lambda + 2\tilde{c}dR \right) \partial_\alpha \xi_r \\ & + \left(\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2} \right) \lambda + 2\tilde{c}dR \right) \partial_\alpha \xi_l \\ & + \left(\beta(1+\xi_r^2)\lambda - \alpha(\lambda+\eta) \right) \partial_\alpha \dot{\xi}_r \\ & + \left((\beta(1+\xi_l^2)\eta - \alpha(\lambda+\eta) \right) \partial_\alpha \dot{\xi}_l \\ & - (\dot{\xi}_r + \dot{\xi}_l)(\lambda+\eta) \Bigg] dt \\ & + \left(\mu_r + \dot{\lambda} \right) \partial_\alpha \xi_r(0) - \dot{\lambda} \partial_\alpha \xi_r(T) \\ & + (\nu_r - \lambda) \, \partial_\alpha \dot{\xi}_r(0) + \lambda \partial_\alpha \dot{\xi}_r(T) \\ & + (\mu_l + \dot{\eta}) \, \partial_\alpha \xi_l(0) - \dot{\eta} \partial_\alpha \xi_l(T) \\ & + (\nu_l - \eta) \, \partial_\alpha \dot{\xi}_l(0) + \eta \partial_\alpha \dot{\xi}_l(T) \end{aligned} \tag{31}$$

Since the partial derivative of the model output ξ w.r.t. the model parameter α is difficult to compute, we eliminate the related terms by setting

$$\text{For } 0 < t < T:$$

$$\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2} \right) \lambda + 2\tilde{c}dR = 0 \tag{32}$$

$$\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2} \right) \eta + 2\tilde{c}dR = 0 \tag{33}$$

$$\beta \left(1 + \xi_r^2 \right) \lambda - \alpha \left(\lambda + \eta \right) = 0 \tag{34}$$

$$\beta \left(1 + \xi_l^2 \right) \eta - \alpha \left(\lambda + \eta \right) = 0 \tag{35}$$

with initial conditions

$$\text{At } t = T:$$

$$\lambda(T) = 0 \tag{36}$$

$$\dot{\lambda}(T) = 0 \tag{37}$$

$$\eta(T) = 0 \tag{38}$$

$$\dot{\eta}(T) = 0 \tag{39}$$

As a result, we obtain the derivative of F w.r.t. α as:

$$F_\alpha = \int_0^T - \left(\dot{\xi}_r + \dot{\xi}_l \right) (\lambda + \eta) dt \tag{40}$$

The derivatives of F w.r.t. β and Δ are similarly obtained as:

$$F_\beta = \int_0^T \left(\left(1 + \xi_r^2\right) \dot{\xi}_r \lambda + \left(1 + \xi_l^2\right) \dot{\xi}_l \eta \right) dt \tag{41}$$

$$F_\Delta = \int_0^T \tfrac{1}{2} \big(\xi_l \eta - \xi_r \lambda \big) dt \tag{42}$$

Having calculated the gradients of F w.r.t. the model parameters, we can now apply gradient-based algorithms to optimize our objective (18).

For instance, applying gradient descent, we have:

$$\begin{aligned} \alpha^{k+1} &= \alpha^k - \tau^\alpha F_\alpha \\ \beta^{k+1} &= \beta^k - \tau^\beta F_\beta \\ \Delta^{k+1} &= \Delta^k - \tau^\Delta F_\Delta \end{aligned} \tag{43}$$

where $\tau^{\cdot}$ is the step-size. The overall algorithm is summarized as follows:

1. Integrate (19) and (20) with initial conditions (21), (22), (23) and (24) from 0 to T, obtaining ξ_r, ξ_l, $\dot{\xi}_r$ and $\dot{\xi}_l$.

2. Integrate (32), (33), (34) and (35) with the initial conditions (36), (37), (38) and (39) from T to 0, obtaining λ, $\dot{\lambda}$, η and $\dot{\eta}$.

3. Update α, β and Δ with (43).

5. SOLVING THE INVERSE PROBLEM FOR A VOCAL TRACT MODEL

We have proposed a backward approach to solve the vocal fold model and the ADLES method for efficiently estimating the model parameters from speech signals. Next, we incorporate the modeling for vocal tract and present a forward-backward paradigm for solving the vocal tract model. We finally extend the ADLES method to solve the inverse problem of estimating the parameters of the integrated vocal fold-tract model. We call this algorithm the ADLES-VFT (standing for ADLES Vocal Fold-Tract) algorithm.

5.1. Modeling wave propagation in the vocal tract

The vocal tract can be viewed as a compact, orientable, differentiable manifold M embedded in $\mathbb{R}^3$. Its boundary ∂M includes the wall of the vocal tract. Consider the tangent bundle TM. Denote the set of all vector fields on TM as $\Gamma(TM)$, which is a $C^\infty(M)$-module [38]. A vector field is a smooth section on TM, $\Gamma(TM) \ni \boldsymbol{X} : M \to TM$. It associate each point $\boldsymbol{p} \in M$ with a

tangent vector $\bar{\boldsymbol{v}}(\boldsymbol{p}) \coloneqq \boldsymbol{X}|_{\boldsymbol{p}} : C^\infty(M) \xrightarrow{\sim} \mathbb{R}$ [38]. Let $\gamma(t) : \mathbb{R} \supseteq I \to M$ be a maximal integral curve [38] through $\boldsymbol{p}$ at t_0, which is a solution to:

$$\begin{aligned}\gamma'(t) &= \boldsymbol{X}(\gamma(t)) \\ \gamma(t_0) &= \boldsymbol{p}\end{aligned}$$

The curve $\gamma(t)$ is a one-parameter group. When acting on the Lie group M, it gives the flow $\Phi : \mathbb{R} \times M \to M$. $\Phi_t(\boldsymbol{p}) = \gamma(t)$. The particle velocity at $\boldsymbol{p}$ is given by $\boldsymbol{v}(\boldsymbol{p}, t) \coloneqq \gamma'(t) = \bar{\boldsymbol{v}}(\boldsymbol{p}) \circ \gamma(t)$.

The planar motion of the pressure wave in the vocal tract is governed by the equations [39]:

$$\frac{1}{\rho c^2}\frac{\partial \hat{p}}{\partial t} + \mathrm{div}\boldsymbol{v} = 0 \tag{44}$$

$$\rho\frac{\partial \boldsymbol{v}}{\partial t} + \mathrm{grad}\hat{p} = 0 \tag{45}$$

where $\hat{p}(\boldsymbol{p}, t)$ is the acoustic pressure, div is the divergence operator, grad is the gradient operator, ρ is the ambient air density, and c is the speed of sound. Equation (44) describes the conservation of mass, and (45) describes the conservation of momentum [39].

For notational convenience, we use cylindrical coordinates $\boldsymbol{p} = (r, \theta, x)$, where x direction aligns with the central axis of vocal tract. Denote the inner surface of vocal tract as Σ, and the shape function of inner surface as $r = R(\theta, x)$. Then the cross-sectional area of the vocal tract is:

$$A(x) = \int_0^{2\pi} d\theta \int_0^{R(\theta,x)} r dr \tag{46}$$

the average acoustic pressure is:

$$p(x, t) = \frac{1}{A(x)} \int_0^{2\pi} d\theta \int_0^{R(\theta,x)} \hat{p} r dr \tag{47}$$

and the volume velocity is:

$$u(x, t) = \int_0^{2\pi} d\theta \int_0^{R(\theta,x)} v_x r dr \tag{48}$$

where v_x is the x component of $\boldsymbol{v}$. Integrating (44) over the volume of vocal tract bounded by cross sections at x_0 and x gives:

$$0 = \int_M \frac{1}{\rho c^2}\frac{\partial \hat{p}}{\partial t} + \mathrm{div}\boldsymbol{v} \tag{49}$$

$$= \int_{x_0}^{x} \left[\int_0^{2\pi} d\theta \int_0^{R} \frac{1}{\rho c^2}\frac{\partial \hat{p}}{\partial t} r dr\right] dx' + \int_M \mathrm{div}\boldsymbol{v} \tag{50}$$

$$= \frac{1}{\rho c^2}\int_{x_0}^{x} A(x')\frac{\partial p(x', t)}{\partial t} dx' + \iint_\Sigma n_{\boldsymbol{v}} d\sigma + u(x, t) - u(x_0, t) \tag{51}$$

where we substitute equations (50) to (51) into (47) and (48), and apply Stokes' theorem [39, 22]; $n_{\boldsymbol{v}}$ is the component of $\boldsymbol{v}$ normal and outward to the inner surface Σ.

The element of area $d\sigma$ is given by [39, 22]:

$$d\sigma = S(\theta, x)d\theta dx \tag{52}$$

where $Sd\theta dx$ is a top 2-form on Σ [38]. Substituting (52) into (51) and differentiating w.r.t. x yields:

$$\frac{A(x)}{\rho c^2}\frac{\partial p}{\partial t} + \frac{\partial u}{\partial x} + \int_0^{2\pi} n_{\boldsymbol{v}}(\theta, x, t)S(\theta, x)d\theta = 0 \tag{53}$$

Following similar steps, integrating the x component of (45) over the cross section at x yields:

$$\rho\frac{\partial u}{\partial t} + A(x)\frac{\partial p}{\partial x} + \int_0^{2\pi} \left(p(x,t) - p_w(\theta, x, t)\right) \frac{\partial}{\partial x}\left(\frac{1}{2}R^2\right) d\theta = 0 \tag{54}$$

where p_w is the pressure acting on the wall of the vocal tract.

5.2. The integrated vocal tract model

To simplify our problem, we combine the wave equations (53) and (54) into a single vocal tract model. Differentiating (53) w.r.t x and (54) w.r.t. t, and canceling out the pressure term gives:

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2} &= c^2\frac{\partial^2 u}{\partial x^2} + \frac{1}{\rho}\frac{\partial A}{\partial x}\frac{\partial p}{\partial t} - \frac{1}{\rho}\partial_t \int_0^{2\pi} \left(p(x,t) - p_w(\theta, x, t)\right) \frac{\partial}{\partial x}\left(\frac{1}{2}R^2\right) d\theta \\ &\quad + c^2\partial_x \int_0^{2\pi} n_{\boldsymbol{v}}(\theta, x, t)S(\theta, x)d\theta \qquad (55) \\ &= c^2\frac{\partial^2 u}{\partial x^2} + f(x,t) \qquad (56)\end{aligned}$$

where the vocal tract profile is absorbed into a single term $f(x,t)$. This represents the characteristics of the vocal tract, i.e., the effect of the nonuniform yielding wall on the acoustic flow dynamics, which needs to be estimated by our algorithm.

5.3. The combined vocal folds-tract model: Formulation of the inverse problem

We now formulate the problem of estimating the parameters of the combined vocal fold-tract model from speech measurements. Let $\Omega \times T$ be the domain of volume velocity u, where Ω is the spatial domain, and T is the time domain. In one-dimensional case, $\Omega = [0, L]$ where L is the length of vocal tract, and $T = [0, t_m]$ where t_m is the maximum of T. Given measured acoustic pressure $p_m(t)$ at the lip, the corresponding volume velocity is given by [40]:

$$u_m(t) = \frac{A(L)}{\rho c}p_m(t) \tag{57}$$

where $A(L)$ is the opening area at the lip, c is the speed of sound, and ρ is the ambient air density. We denote $u_0(t) \coloneqq u(0,t)$, $u_L(t) \coloneqq u(L,t)$. The glottal flow $u_0(t)$ can be derived from the vocal folds displacement model (10) by:

$$u_0(t) = \tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t)) \tag{58}$$

where ξ_0 is the half glottal width at rest, d is the length of the vocal fold, and $\tilde{c}$ is the air particle velocity at the midpoint of the vocal fold (see Figure 3).

Let $\mathcal{H}$ be the nonlinear operator representing acoustic wave propagation from the glottis to the lip. We have the forward propagation process:

$$\begin{aligned} \mathcal{H} : \mathcal{L}^2(\Omega \times T) \times \mathcal{L}^2(\Gamma \times T) &\to \mathcal{L}^2(\Gamma \times T) \\ (f, u_0) &\mapsto u_L \end{aligned} \tag{59}$$

where f is the vocal tract profile in (56), and $\Gamma = \partial\Omega$ is the boundary.

We can split Γ into two parts $\Gamma = \Gamma_0 \bigcup \Gamma_1$, $\Gamma_0 \bigcap \Gamma_L = \emptyset$ corresponding to $x = 0$ and $x = L$. But we neglect the difference to ease our derivation. Note that in one-dimensional case $u(t)$ and $u_L(t)$ are only functions of t. However, more generally, they are functions of both x on the boundary Γ and t. We now define two nonlinear operators as:

$$\begin{aligned} \mathcal{H}_f : \mathcal{L}^2(\Gamma \times T) &\to \mathcal{L}^2(\Gamma \times T) \\ u_0 &\mapsto u_L \end{aligned} \tag{60}$$

and

$$\begin{aligned} \mathcal{F} \coloneqq \mathcal{H}_{u_0} : \mathcal{L}^2(\Omega \times T) &\to \mathcal{L}^2(\Gamma \times T) \\ f &\mapsto u_L \end{aligned} \tag{61}$$

Note that both $\mathcal{H}_f$ and $\mathcal{F}$ are bounded. Our objective is to minimize the difference between the measured volume velocity u_m and the predicted volume velocity u_L near the lip subject to

constraints:

$$\min \int_0^T \left(\mathcal{H}_f(u_0(t)) - u_m(t)\right)^2 dt \tag{62}$$

$$\Leftrightarrow \min \int_0^T \left(\mathcal{H}_f(\tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t))) - \frac{A(L)}{\rho c} p_m(t)\right)^2 dt \tag{63}$$

$$\text{subject to} \quad \ddot{\xi}_r + \beta(1+\xi_r^2)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{64}$$

$$\ddot{\xi}_l + \beta(1+\xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{65}$$

$$\text{(I.C.1)} \quad \xi_r(0) = C_r \tag{66}$$

$$\text{(I.C.2)} \quad \xi_l(0) = C_l \tag{67}$$

$$\text{(I.C.3)} \quad \dot{\xi}_r(0) = 0 \tag{68}$$

$$\text{(I.C.4)} \quad \dot{\xi}_l(0) = 0 \tag{69}$$

$$\tag{70}$$

where (64) and (65) represent the asymmetric vocal folds displacement model (10), I.C. stands for initial condition, and Cs are constants. Next, we derive an efficient strategy to estimate the parameters α, β, and Δ such that (63) (the least squares objective based on the vocal tract model) is minimized.

This essentially represents a combination of the vocal folds-tract models from a computational perspective.

5.4. Solving the inverse problem for the Vocal fold-Tract model: The ADLES-VFT algorithm

We solve the inverse problem for the combined vocal folds-tract model using the via Forward-Backward method proposed below. We call this algorithm the Adjoint Least Squares - Vocal folds-Tract (ADLES-VFT) parameter estimation algorithm.

In order to solve the parameter estimation problem represented by (70), first we need to estimate

the vocal tract profile f in $\mathcal{H}_f$ and (56). Specifically, we need to solve:

$$\frac{\partial^2 u(x,t)}{\partial t^2} = c^2 \frac{\partial^2 u(x,t)}{\partial x^2} + f(x,t) \tag{71}$$

subject to

$$\text{(B.C.1)} \quad u(0,t) = u_g(t) \tag{72}$$

$$\text{(B.C.2)} \quad u(L,t) = u_m(t) \tag{73}$$

$$\text{(B.C.3)} \quad \frac{\partial u}{\partial n_\Gamma} = 0 \tag{74}$$

$$\text{(I.C.1)} \quad u(x,0) = 0 \tag{75}$$

$$\text{(I.C.2)} \quad \frac{\partial u(x,0)}{\partial t} = 0 \tag{76}$$

$$\tag{77}$$

where B.C. stands for boundary condition, u_g, u_m are volume velocity at the glottis and lip, and n_Γ is the outward unit normal to the boundary Γ. We now derive the solution to (77).

In order to estimate $f \in \mathcal{L}^2(\Omega \times T)$, we take an iterative approach, i.e.,

$$f^{k+1} = f^k + \tau \delta f^k \tag{78}$$

where $\delta f^k \in \mathcal{L}^2(\Omega \times T)$ is a small variation, and τ is a step size. A Taylor expansion of $\mathcal{F}$ (61) at f^k gives:

$$\mathcal{F}(f^k + \delta f^k) = \mathcal{F}(f^k) + \mathcal{F}'(f^k)\delta f^k + O\left((\delta f^k)^2\right) \tag{79}$$

where $\mathcal{F}'$ is the Fréchet derivative [41]. Omitting higher order terms, we obtain:

$$\mathcal{F}'(f^k)\delta f^k = \mathcal{F}(f^k + \delta f^k) - \mathcal{F}(f^k) \tag{80}$$

where $\mathcal{F}'(f)$ is a nonlinear operator

$$\begin{aligned} \mathcal{F}'(f) : \mathcal{L}^2(\Omega \times T) &\to \mathcal{L}^2(\Gamma \times T) \\ \delta f &\mapsto \delta u_L \end{aligned} \tag{81}$$

Correspondingly, the adjoint operator [41, 42, 43] is:

$$\begin{aligned} \mathcal{F}'(f)^* : \mathcal{L}^2(\Gamma \times T) &\to \mathcal{L}^2(\Omega \times T) \\ \delta u_L &\mapsto \delta f \end{aligned} \tag{82}$$

We would like $\mathcal{F}(f^k + \delta f^k) = u_L^k + \delta u_L^k \xrightarrow{k\to\infty} u_m$. This is equivalent to solving:

$$\begin{aligned} &\min \ \|\delta f^k\|_2^2 \\ \text{subject to} \quad &\mathcal{F}'(f^k)\delta f^k = u_m - \mathcal{F}(f^k) \end{aligned} \tag{83}$$

It is simple to obtain the solution to (83).

$$\delta f^k = -\mathcal{F}'(f^k)^* \left[\mathcal{F}'(f^k)\mathcal{F}'(f^k)^*\right]^{-1} \left(\mathcal{F}(f^k) - u_m\right) \tag{84}$$

where $\mathcal{F}'(f^k)^*$ is the adjoint operator. It is difficult to compute $\mathcal{F}'(f^k)\mathcal{F}'(f^k)^*$. We use its property of positive-definiteness to approximate it by $\gamma \boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix.

We denote the estimation residual as:

$$r^k := u_m - \mathcal{F}(f^k) \tag{85}$$

We have

$$\delta f^k = \tfrac{1}{\gamma}\mathcal{F}'(f^k)^* r^k \tag{86}$$

Now consider the wave equation (71). Let $u + \delta u$ be a solution with variation $f + \delta f$. Substitution into (71) yields:

$$\frac{\partial^2(u+\delta u)}{\partial t^2} = c^2 \frac{\partial^2(u+\delta u)}{\partial x^2} + f + \delta f \tag{87}$$

Subtracting (71) yields:

$$\frac{\partial^2 \delta u}{\partial t^2} = c^2 \frac{\partial^2 \delta u}{\partial x^2} + \delta f \tag{88}$$

subject to

$$\text{(B.C.1)} \quad \frac{\partial \delta u}{\partial n_\Gamma} = 0 \tag{89}$$

$$\text{(I.C.1)} \quad \delta u(x,0) = 0 \tag{90}$$

$$\text{(I.C.2)} \quad \frac{\partial \delta u(x,0)}{\partial t} = 0 \tag{91}$$

$$\tag{92}$$

Next, we use a time reversal technique [39] and backpropagate the difference (85) into the vocal tract, which gives:

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2} + f(x,t) \tag{93}$$

subject to

$$\text{(B.C.1)} \quad \frac{\partial z}{\partial n_\Gamma} = r \tag{94}$$

$$\text{(I.C.1)} \quad z(x,t_m) = 0 \tag{95}$$

$$\text{(I.C.2)} \quad \frac{\partial z(x,t_m)}{\partial t} = 0 \tag{96}$$

$$\tag{97}$$

where z is the time reversal of u. It follows [44] that:

$$\langle \delta f, z \rangle_{\Omega \times T} = \int_0^{t_m} \int_\Omega \delta f z dx dt \tag{98}$$

$$= \int_0^{t_m} \int_\Omega \left(\frac{\partial^2 \delta u}{\partial t^2} - c^2 \frac{\partial^2 \delta u}{\partial x^2} \right) z dx dt \tag{99}$$

$$= \int_0^{t_m} \int_\Omega \left(\frac{\partial^2 \delta u}{\partial t^2} - c^2 \frac{\partial^2 \delta u}{\partial x^2} \right) z dx dt - \int_0^{t_m} \int_\Omega \left(\frac{\partial^2 z}{\partial t^2} - c^2 \frac{\partial^2 z}{\partial x^2} - f \right) \delta u dx dt \tag{100}$$

$$= \int_0^{t_m} \int_\Omega \left(\frac{\partial^2 \delta u}{\partial t^2} z - \frac{\partial^2 z}{\partial t^2} \delta u \right) dx dt - c^2 \int_0^{t_m} \int_\Omega \left(\frac{\partial^2 \delta u}{\partial x^2} z - \frac{\partial^2 z}{\partial x^2} \delta u \right) dx dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{101}$$

$$= \int_\Omega \left(\frac{\partial \delta u}{\partial t} z - \frac{\partial z}{\partial t} \delta u \right) \Bigg|_0^{t_m} dx dt - c^2 \int_0^{t_m} \int_\Omega \left(\frac{\partial^2 \delta u}{\partial x^2} z - \frac{\partial^2 z}{\partial x^2} \delta u \right) dx dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{102}$$

$$= -c^2 \int_0^{t_m} \int_\Omega \left(\frac{\partial^2 \delta u}{\partial x^2} z - \frac{\partial^2 z}{\partial x^2} \delta u \right) dx dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{103}$$

$$= -c^2 \int_0^{t_m} \int_\Omega \left(z d \frac{\partial \delta u}{\partial x} - \delta u d \frac{\partial z}{\partial x} \right) dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{104}$$

$$= -c^2 \int_0^{t_m} \left(\int_\Gamma z \frac{\partial \delta u}{\partial n_\Gamma} ds - \int_\Omega \frac{\partial \delta u}{\partial x} \frac{\partial z}{\partial x} dx - \int_\Gamma \delta u \frac{\partial z}{\partial n_\Gamma} ds + \int_\Omega \frac{\partial \delta u}{\partial x} \frac{\partial z}{\partial x} dx \right) dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{105}$$

$$= c^2 \int_0^{t_m} \int_\Gamma \delta u \frac{\partial z}{\partial n_\Gamma} ds dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{106}$$

$$= c^2 \int_0^{t_m} \int_\Gamma \delta u r ds dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{107}$$

$$= c^2 \int_0^{t_m} \int_\Gamma \mathcal{F}'(f) \delta f r ds dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{108}$$

$$= c^2 \int_0^{t_m} \int_\Omega \delta f \mathcal{F}'(f)^* r dx dt + \int_0^{t_m} \int_\Omega f \delta u dx dt \tag{109}$$

$$= c^2 \int_0^{t_m} \int_\Omega \delta f \mathcal{F}'(f)^* r dx dt - \int_0^{t_m} \int_\Omega \delta f u dx dt \tag{110}$$

$$= c^2 \int_0^{t_m} \int_\Omega \delta f \left(\mathcal{F}'(f)^* r - u \right) dx dt \tag{111}$$

where we substitute from (98) to (100) into (88) and (93); from (100) to (103) we apply initial conditions (90), (91), (95) and (96); from (103) to (105) we integrate by parts; from (105) to (106) we apply boundary condition (89); from (106) to (107) we use boundary condition (94); from (107) to (108) we use definition (81); from (108) to (109) we use definition (82) and the duality property

$$\langle \mathcal{F}'(f)\delta f, r\rangle_{\Gamma\times T} = \langle \delta f, \mathcal{F}'(f)^* r\rangle_{\Omega\times T} \tag{112}$$

from (109) to (110), we assume that the second-order variation is small, i.e.,

$$\langle f+\delta f, u+\delta u\rangle = \langle f, u\rangle + \langle f, \delta u\rangle + \langle \delta f, u\rangle + \langle \delta f, \delta u\rangle \approx \langle f, u\rangle \tag{113}$$

(or $\delta(fu) = \delta(f)u + f\delta(u) \approx 0$.) By the arbitrariness of δf, it follows that:

$$z = c^2(\mathcal{F}'(f)^* r - u) \tag{114}$$

and hence

$$\mathcal{F}'(f)^* r = \frac{z}{c^2} + u \tag{115}$$

Substitution into (86) and (78) yields:

$$f^{k+1} = f^k + \frac{\tau}{\gamma}\left(\frac{z^k}{c^2} + u^k\right) \tag{116}$$

Hence, we obtain an iterative forward-backward approach for solving the vocal tract profile f.

5.4.1. Estimating model parameters via the adjoint least squares method

Now, we derive solution to the parameter estimation problem (70) using the adjoint least squares method proposed in Section 4.2.

Denote the estimation error as:

$$f(\xi_l, \xi_r; \vartheta) = \left(\mathcal{H}_f(\tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t))) - \frac{A(L)}{\rho c} p_m(t)\right)^2 \tag{117}$$

and

$$F(\xi_l, \xi_r; \vartheta) = \int_0^{t_m} f(\xi_l, \xi_r; \vartheta) dt \tag{118}$$

where $\vartheta = [\alpha, \beta, \Delta]$ are the parameters of the vocal folds model (10). We would like to obtain update rules for the model parameters α, β, and Δ, i.e.,

$$\alpha^{k+1} = \alpha^k - \tau^\alpha F_{\alpha^k} \tag{119}$$

$$\beta^{k+1} = \beta^k - \tau^\beta F_{\beta^k} \tag{120}$$

$$\Delta^{k+1} = \Delta^k - \tau^\Delta F_{\Delta^k} \tag{121}$$

$$\tag{122}$$

where the the partial derivatives $F_{\cdot} \coloneqq \partial_{\cdot} F \equiv \frac{\partial F}{\partial_{\cdot}}$ and $\tau_{\cdot}$ is the step size. We now define the Lagrangian:

$$\begin{aligned}\mathcal{L}(\vartheta) = \int_0^{t_m} \Bigg[& f + \lambda \left(\ddot{\xi}_r + \beta(1+\xi_r^2)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r - \alpha(\dot{\xi}_r + \dot{\xi}_l) \right) \\ & + \eta \left(\ddot{\xi}_l + \beta(1+\xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l - \alpha(\dot{\xi}_r + \dot{\xi}_l) \right) \Bigg] dt \\ & + \mu_l(\xi_l(0) - C_l) + \mu_r(\xi_r(0) - C_r) + \nu_l\dot{\xi}_l(0) + \nu_r\dot{\xi}_r(0) \end{aligned} \tag{123}$$

where λ, η, μ and ν are multipliers. Taking the derivative of the Lagrangian w.r.t. the model parameter α yields:

$$\begin{aligned}\mathcal{L}_\alpha = \int_0^{t_m} \Bigg[& 2\tilde{c}d\mathcal{H}_f^{'}\Big|_{u_0} (\partial_\alpha\xi_l + \partial_\alpha\xi_r) \\ & + \lambda \left(\partial_\alpha\ddot{\xi}_r + 2\beta\dot{\xi}_r\xi_r\partial_\alpha\xi_r + \beta(1+\xi_r^2)\partial_\alpha\dot{\xi}_r + \partial_\alpha\xi_r - \frac{\Delta}{2}\partial_\alpha\xi_r - \alpha(\partial_\alpha\dot{\xi}_r + \partial_\alpha\dot{\xi}_l) - (\dot{\xi}_r + \dot{\xi}_r) \right) \\ & + \eta \left(\partial_\alpha\ddot{\xi}_l + 2\beta\dot{\xi}_l\xi_l\partial_\alpha\xi_l + \beta(1+\xi_l^2)\partial_\alpha\dot{\xi}_l + \partial_\alpha\xi_l + \frac{\Delta}{2}\partial_\alpha\xi_l - \alpha(\partial_\alpha\dot{\xi}_r + \partial_\alpha\dot{\xi}_l) - (\dot{\xi}_r + \dot{\xi}_r) \right) \Bigg] dt \\ & + \mu_l\partial_\alpha\xi_l(0) + \mu_r\partial_\alpha\xi_r(0) + \nu_l\partial_\alpha\dot{\xi}_l(0) + \nu_r\partial_\alpha\dot{\xi}_r(0) \end{aligned} \tag{124}$$

Integrating the term $\lambda\partial_\alpha\ddot{\xi}_r$ by parts twice gives:

$$\int_0^{t_m} \lambda\partial_\alpha\ddot{\xi}_r dt = \int_0^{t_m} \partial_\alpha\xi_r\ddot{\lambda} dt - \partial_\alpha\xi_r\dot{\lambda}\Big|_0^{t_m} + \partial_\alpha\dot{\xi}_r\lambda\Big|_0^{t_m} \tag{125}$$

Defining the estimation residual $R \coloneqq \mathcal{H}_f(u_0) - \frac{A(L)}{\rho c} p_m(t)$, applying the same to $\eta\partial_\alpha\ddot{\xi}_l$, substitution into (124), after simplification yields:

$$\begin{aligned}\mathcal{L}_\alpha = \int_0^{t_m} \Bigg[& \left(\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2} \right) \lambda + 2\tilde{c}dR\mathcal{H}_f^{'}\Big|_{u_0} \right) \partial_\alpha\xi_r \\ & + \left(\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2} \right) \lambda + 2\tilde{c}dR\mathcal{H}_f^{'}\Big|_{u_0} \right) \partial_\alpha\xi_l \\ & + \left(\beta(1+\xi_r^2)\lambda - \alpha(\lambda+\eta) \right) \partial_\alpha\dot{\xi}_r + \left((\beta(1+\xi_l^2)\eta - \alpha(\lambda+\eta) \right) \partial_\alpha\dot{\xi}_l - (\dot{\xi}_r + \dot{\xi}_l)(\lambda+\eta) \Bigg] dt \\ & + (\mu_r + \dot{\lambda})\partial_\alpha\xi_r(0) - \dot{\lambda}\partial_\alpha\xi_r(T) + (\nu_r - \lambda)\partial_\alpha\dot{\xi}_r(0) + \lambda\partial_\alpha\dot{\xi}_r(T) \\ & + (\mu_l + \dot{\eta})\partial_\alpha\xi_l(0) - \dot{\eta}\partial_\alpha\xi_l(T) + (\nu_l - \eta)\partial_\alpha\dot{\xi}_l(0) + \eta\partial_\alpha\dot{\xi}_l(T) \end{aligned} \tag{126}$$

where the term $\mathcal{H}_f^{'}|_{u_0} \approx u_L/u_0$ by linearization. Since the partial derivatives of the displacement ξ w.r.t. the model parameter α are difficult to compute, we cancel out the related terms by setting:

$$\begin{aligned}
&\text{For } 0 < t < t_m : \\
&\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2}\right)\lambda + 2\tilde{c}dR\mathcal{H}_f^{'}\Big|_{u_0} = 0 && (127)\\
&\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2}\right)\eta + 2\tilde{c}dR\mathcal{H}_f^{'}\Big|_{u_0} = 0 && (128)\\
&\beta(1+\xi_r^2)\lambda - \alpha(\lambda+\eta) = 0 && (129)\\
&\beta(1+\xi_l^2)\eta - \alpha(\lambda+\eta) = 0 && (130)\\
& && (131)
\end{aligned}$$

with initial conditions:

$$\begin{aligned}
&\text{At } t = t_m : \\
&\lambda(t_m) = 0 && (132)\\
&\dot{\lambda}(t_m) = 0 && (133)\\
&\eta(t_m) = 0 && (134)\\
&\dot{\eta}(t_m) = 0 && (135)\\
& && (136)
\end{aligned}$$

Consequently, we obtain the derivative of F w.r.t. α:

$$F_\alpha = \int_0^{t_m} -(\dot{\xi}_r + \dot{\xi}_l)(\lambda+\eta)dt \tag{137}$$

Similarly, we obtain the derivatives of F w.r.t. β and Δ

$$F_\beta = \int_0^{t_m} \left((1+\xi_r^2)\dot{\xi}_r\lambda + (1+\xi_l^2)\dot{\xi}_l\eta\right)dt \tag{138}$$

$$F_\Delta = \int_0^{t_m} \frac{1}{2}\left(\xi_l\eta - \xi_r\lambda\right)dt \tag{139}$$

5.4.2. The ADLES-VFT algorithm summarized

The algorithm for solving the parameter estimation problem (70) is outlined below.

1. Integrate (64) and (65) with initial conditions (66), (67), (68) and (69) from 0 to t_m, obtaining ξ_r^k, ξ_l^k, $\dot{\xi}_r^k$ and $\dot{\xi}_l^k$.

2. Solve the forward propagation model (77) for u_L^k, $\mathcal{H}_f^{'}\Big|_{u_0^k}$.

3. Calculate the estimation difference r^k using (85).

4. Solve the backward propagation model (97) for z^k.

5. Update f^k using (116).

6. Integrate (127), (128), (129) and (130) with initial conditions (132), (133), (134) and (135) from t_m to 0, obtaining λ^k, $\dot{\lambda}^k$, η^k and $\dot{\eta}^k$.

7. Update α, β and Δ with (122).

In this solution, we have adopted the simple gradient descent method. However, other gradient-based optimization approaches, such as the conjugate gradient method, can also be used.

5.5. Numerical solution for wave propagation

What remains now is to solve the acoustic wave propagation problems represented by (77) and (97). We derive a finite element solution for these below.

5.5.1. Variational Formulation

First, for the time-dependent system of PDEs, we discretize it along time t with the backward Euler method [45], yielding a sequence of differential equations. We split the time domain T into N uniform length intervals Δt. For time step n, $0 \leq n \leq N-1$, applying the backward Euler method to the left side of (71) gives:

$$\left[D_t D_t^- u\right]^n := D_t D_t^- \left(\frac{\partial^2 u}{\partial t^2}\right) = \frac{u^n - 2u^{n-1} + u^{n-2}}{\Delta t^2} \tag{140}$$

where $D_t D_t^n$ is a finite difference operator w.r.t. time at time step n [45, 46]. Substitution into (71) yields:

$$\left[D_t D_t^- u = c^2 \frac{\partial^2 u}{\partial x^2} + f\right]^n \tag{141}$$

$$\Leftrightarrow \quad u^n = \Delta t^2 c^2 \frac{\partial^2 u^n}{\partial x^2} + \Delta t^2 f^n + 2u^{n-1} - u^{n-2} \tag{142}$$

Next, define the residual at time step n as:

$$R^n = u^n - \Delta t^2 c^2 \frac{\partial^2 u^n}{\partial x^2} + \Delta t^2 f^n + 2u^{n-1} - u^{n-2} \tag{143}$$

Applying Galerkin's method [45, 47] gives:

$$\langle R^n, v \rangle_{W^{k,2}} = 0 \tag{144}$$

where $v \in \mathcal{W}^{k,2}$ ($\mathcal{W}^{k,2}$ is the Sobolev space of functions with bounded L_2 norm and k-th order weak derivatives) is a qualified test function. Galerkin's method orthogonally projects the residual to the function space $\mathcal{W}^{k,2}$. Expanding (144) yields:

$$\int_\Omega u^n v dx - \Delta t^2 c^2 \int_\Omega \frac{\partial^2 u^n}{\partial x^2} v dx = \int_\Omega \left(\Delta t^2 f^n + 2u^{n-1} - u^{n-2}\right) v dx \tag{145}$$

Integration by parts for the second-order term in (145) gives:

$$\int_\Omega \frac{\partial^2 u^n}{\partial x^2} v dx = -\int_\Omega \frac{\partial u^n}{\partial x}\frac{\partial v}{\partial x} dx + \int_\Gamma \frac{\partial u^n}{\partial n_\Gamma} ds \tag{146}$$

where n_Γ is the outward normal unit vector of the boundary Γ, and ds is the 1-form [38] on Γ. For problem (77), applying the boundary condition (74) and substitution (146) back into (145) yield the variational problem:

$$\int_\Omega u^n v dx + \Delta t^2 c^2 \int_\Omega \frac{\partial u^n}{\partial x}\frac{\partial v}{\partial x} dx = \int_\Omega \left(\Delta t^2 f^n + 2u^{n-1} - u^{n-2}\right) v dx \tag{147}$$

For the problem represented by (97), applying the boundary condition (94) and substitution (146) back into (145) yields a similar variational problem:

$$\int_\Omega z^n w dx + \Delta t^2 c^2 \int_\Omega \frac{\partial z^n}{\partial x}\frac{\partial w}{\partial x} dx = \int_\Omega \left(\Delta t^2 f^n + 2z^{n-1} - z^{n-2}\right) w dx + \int_\Gamma r^n w ds \tag{148}$$

We can split the variational problem (147) into two parts:

$$a(u, v) = \int_\Omega uv dx + \Delta t^2 c^2 \int_\Omega \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} dx \tag{149}$$

$$L(v) = \int_\Omega \left(\Delta t^2 f^n + 2u^{n-1} - u^{n-2}\right) v dx \tag{150}$$

where we have interchanged the unknown u^n with u. (149) is the bilinear form, and (150) is the linear form [45].

Our original problem (77) and (97) then reduce to solving:

$$a(u, v) = L(v) \tag{151}$$

for each time step. By the Lax-Milgram Lemma [46], solving (151) is equivalent to solving the functional minimization problem:

$$F(u) = \arg\min_{v \in \mathcal{V}} \frac{1}{2} a(v, v) - L(v) \tag{152}$$

By the calculus of variations, and taking the variation of the functional gives (151), hence the name *variational* form [45, 46].

5.5.2. Finite Element Approximation

For each time step, we solve (151) with finite element method. We discretize the domain Ω with a mesh of uniformly spaced triangular cells. We take the P_2 elements as the basis function space, which contains piece-wise, second-order Lagrange polynomials defined over a cell. Each basis function has a degree-of-freedom (DoF) of 6 over a two-dimensional cell [45, 48]. Each element is associated with a coordinate map that transforms local coordinates to global coordinates and a DoF map that maps local DoF to global DoF [45, 48]. Each cell is essentially a simplex and can be continuously transformed into the physical domain.

Existence of Unique Solution The solution to the variational problem (151) exists and is unique [48].

Approximation Error The Galerkin's method gives the solution u_e with error bounded by $\mathcal{O}(h^3\|D^2u_e\|_{\mathcal{W}^{3,2}})$, where h is the cell size and D is the bounded derivative operator [46, 48].

Assume a solution $u = B + c^j\psi_j$ (using Einstein summation convention) with basis $\psi_j \in P_2$ and coefficients c^j. The function $B(x)$ incorporates the boundary condition and, as an example, can take the form:

$$B(x) = u_g + (u_m - u_g)\,\frac{x^p}{L^p}, \quad p > 0 \tag{153}$$

We also project $B(x)$ over the basis functions P_2 and express it as $B(x) = b^j\psi_j$. As a result, we obtain an unified expression $u = U^j\psi_j$ with U^j incorporating b^j and c^j. Similarly, we have $f^n = F_n^j\psi_j$, $u^{n-1} = U_{n-1}^j\psi_j$, $u^{n-2} = U_{n-2}^j\psi_j$. Set the test function as $v = \hat{\psi}_i$. Substitution into (149) and (150) yields:

$$\begin{aligned} a(u,v) &= \int_\Omega U^j\psi_j\hat{\psi}_i dx + \Delta t^2c^2\int_\Omega U^j\psi_j'\psi_i' dx \\ &= \left(\int_\Omega \hat{\psi}_i\psi_j dx + \Delta t^2c^2\int_\Omega \psi_i'\psi_j' dx\right)U^j \end{aligned} \tag{154}$$

$$\begin{aligned} L(v) &= \int_\Omega \left(\Delta t^2F_n^j\psi_j + 2U_{n-1}^j\psi_j - U_{n-2}^j\psi_j\right)\hat{\psi}_i dx \\ &= \Delta t^2\left(\int_\Omega \hat{\psi}_i\psi_j dx\right)F_n^j + 2\left(\int_\Omega \hat{\psi}_i\psi_j dx\right)U_{n-1}^j - \left(\int_\Omega \hat{\psi}_i\psi_j dx\right)U_{n-2}^j \end{aligned} \tag{155}$$

Setting $M_{i,j} = \int_\Omega \hat{\psi}_i\psi_j dx$, $K_{i,j} = \int_\Omega \psi_i'\psi_j' dx$ and collecting (154) and (155) into matrix-vector form, we obtain:

$$AU = b \tag{156}$$

where $A = M + \Delta t^2c^2K$, and $b = \Delta t^2MF^n + 2MU^{n-1} - MU^{n-2}$. Hence, we have reduced the problem of (151) into solving the linear system (156), with the solution described above. Furthermore, the matrices M (known as the mass matrix) and K (known as the stiffness matrix) can be pre-calculated for efficiency.

In the next section, we demonstrate the usefulness of the ADLES and ADLES-VFT algorithms experimentally.

6. EXPERIMENTAL RESULTS AND INTERPRETATION

Having presented the algorithms for estimating the parameters of the vocal folds model, and the coupled vocal fold-tract models, it not only important to validate them, but also to find ways to interpret the solutions for real-world use. Lacking explicit validation data, our validations comes from the proxy of showing that the solutions obtained are indeed discriminative of fine-level changes in glottal flow dynamics of the phonation process. For this, we first describe some ways to interpret the solutions obtained through these individualized models, extract various feature representations from them, and use them in conjunction with machine learning algorithms for discriminative tasks, such as classification and regression. In this section we present all of these.

6.1. Essential characterizations for analysis of dynamical system models

Having recovered the model parameters by our backward or forward approach, we can solve the models to obtain the time-series corresponding to the oscillations of each vocal fold, as estimated from recorded speech samples.

To interpret these, we can utilize some well-established methods for characterizing dynamical systems, borrowing them from chaos theory and other areas of applied mathematics. We describe some of these below.

The models we have discussed in this paper are essentially dynamical systems represented by coupled nonlinear equations that may not have closed-form solutions, but can be numerically solved.

Definition 6.1 (Dynamical system)**.** A real-time dynamical system is a tuple (T, M, Φ), where T is a monoid (an algebraic construct, such as an open interval in $\mathbb{R}_+$). M is a manifold locally diffeomorphic to a Banach space, usually called the *phase space*. As opposed to the configuration space describing the "position" of a dynamical system, the phase space describes the "states" or "motion" of the dynamical system. It is often defined as the tangent bundle TM or the cotangent bundle T^*M of the underlying manifold. $\Phi : T \times M \supseteq U \to M$, where $\mathrm{proj}_2(U) = M$, is the (continuous) evolution function [49].

A phonation model outputs a phase space trajectory of state variables that describes the movements of the vocal folds. The trajectories tend to fall into orbits with regular or irregular behaviors that explain observed behavior patterns of the vocal folds. The possible types and distributions of these orbits depend on the system parameters.

6.1.1. The evolution of a dynamical system

A dynamical system can be instantiated with ordinary or partial differential equations with initial conditions, and the evolution function Φ is the solution to the ODE or PDE.

Definition 6.2 (Evolution function)**.** Denote the duration of evolution of a dynamical system as $I(x) = \{t \in T \mid (t, x) \in U\}$. The evolution function Φ is a group action of T on M satisfying

1. $\Phi(0, x) = x,$ for all $x \in M$;

2. $\Phi(t_2 + t_1, x) = \Phi(t_2, \Phi(t_1, x)),$ for $t_1, t_2 + t_1 \in I(x), t_2 \in I(\Phi(t_1, x))$.

6.1.2. Trajectory, flow, orbit and invariance

Write $\Phi_x(t) \equiv \Phi^t(x) \equiv \Phi(t, x)$. The map $\Phi^t : M \to M$ is a diffeomorphism (i.e., differentiable, invertible, bijection map between manifolds).

Definition 6.3 (Flow, orbit, invariance)**.** The map $\Phi_x : I(x) \to M$ is the *flow* or *trajectory* through x. The set of all flows $\gamma_x \coloneqq \{\Phi_x \mid t \in I(x)\}$ is the *orbit* through x. Particularly, a subset $S \subseteq M$ is called Φ-*invariant* if $\Phi(t, x) \in S$ for all $x \in S$ and $t \in T$.

6.1.3. Phase space behavior: Attractor

The behaviors of flows can be described by their attractor/attraction sets.

Definition 6.4 (Attractor)**.** An attractor set $A \subseteq M$ in the phase space is a closed subset satisfying for some initial point x, there exists a t_0 such that $\Phi_x(t) \in A$ for any $t > t_0$.

Namely, the orbit γ_x is "trapped" in the interior of A. A dynamical system can have more than one attractor set depending on the choice of initial points (or the choice of parameters, as we will see later). Locally we can talk about a *basin of attraction* $B(A)$, which is a neighborhood of A satisfying for any initial point $x \in B(A)$, and its orbit is eventually trapped in A.

There are different types of attractor sets. Some are shown in Figure 4. The simplest one is a *fixed point* or an *equilibrium point*, to which a trajectory in phase space converges regardless of initial settings of the variables (or their starting point). To study vocal fold behaviors, we are particularly interested in those attractors revealing the periodic motion of the flow in phase space. Such attractors include the *limit cycle* or the *limit torus*, which are isolated periodic or toroidal orbits respectively. Some attractor sets have a fractal structure resultant from a chaotic state of the dynamical system [50, 51], and are called *strange attractors*.

6.1.4. Chaos and exponential divergence of phase space trajectories

Chaos is a characteristic state of a nonlinear dynamical system. There are different definitions of chaos. A simple one is as follows:

Definition 6.5 (Chaos)**.** Equip a distance metric d on the phase space M. Then $C \in M$ is referred to as a chaotic set of Φ if, for any $x, y \in C$, $x \neq y$, we have

$$\lim_{n\to\infty} \inf d(\Phi^n(x), \Phi^n(y)) = 0 \tag{157}$$

$$\lim_{n\to\infty} \sup d(\Phi^n(x), \Phi^n(y)) > 0 \tag{158}$$

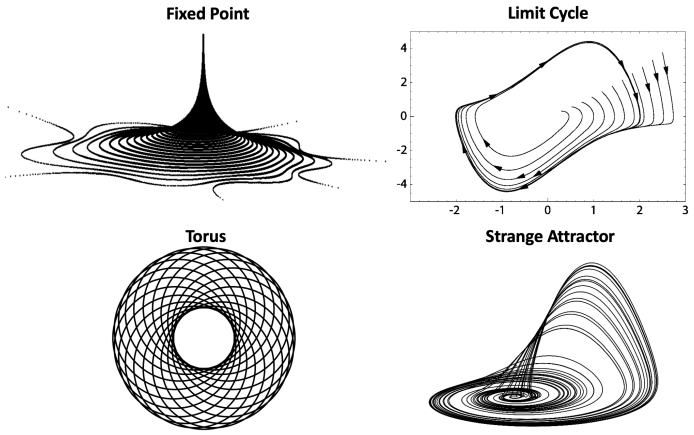


Fig. 4. Illustration of different attractors in a dynamical system.

Thus, by definition, chaos is a state characterized by extreme sensitivity to initial conditions (trajectories starting from any two arbitrarily close initial conditions diverge exponentially). The Lypunov exponent is used to quantify this divergence. It also measures the measures the sensitivity of the evolution of the dynamical system to initial conditions.

6.1.5. Stability

Attractor sets (of all types) also characterize the stability of dynamical systems.

Definition 6.6 (Stability)**.** A compact Φ-invariant subset $A = \Phi(A) \subseteq M$ is called a *Lyapunov stable* attraction set if

1. It has an open basin of attraction $B(A)$;
2. The Lyapunov stability condition is satisfied: every neighborhood U of A contains a smaller neighborhood V such that every iterative forward image $\Phi^n(V)$ is contained in U.

6.1.6. Poincaré map and Poincaré section

To study the orbit structure of dynamical systems, we use the Poincaré map or Poincaré section.

Definition 6.7 (Poincaré map [49])**.** For an n-dimensional phase space with a periodic orbit γ_x, a Poincaré section S is an $(n-1)$-dimensional section (hyper-plane) that is transverse to γ_x. Given an open, connected neighborhood $U \subseteq S$ of x, the Poincaré map on Poincaré section S is a map $P : U \to S$, $x \mapsto \Phi_x(t_s)$ where t_s is the time between the two intersections, satisfying

1. $P(U)$ is a neighborhood of x and $P : U \to P(U)$ is a diffeomorphism;
2. For every point x in U, the positive semi-orbit of x intersects S for the first time at $P(x)$.

6.1.7. Bifurcation

Since the flow of a dynamical system in its phase space is a function of its parameters, the topological structure of the trajectories (including attractor sets) in phase space changes as the parameters change. To see how the topological structure changes with system parameters, we study the bifurcation map of the system.

Definition 6.8 (Bifurcation)**.** A bifurcation occurs when a small smooth change in a system parameter value causes an abrupt change in the topological structure of the trajectory in phase space. A *bifurcation diagram* is a visualization of the system's parameter space showing the number and behavior of attractor sets for each parameter configuration.

At a *bifurcation point*, the system stability may change as the topological structure splits or merges, such as the periodic doubling or halving of a limit cycle.

6.2. Interpreting a system's phase portraits using its bifurcation map

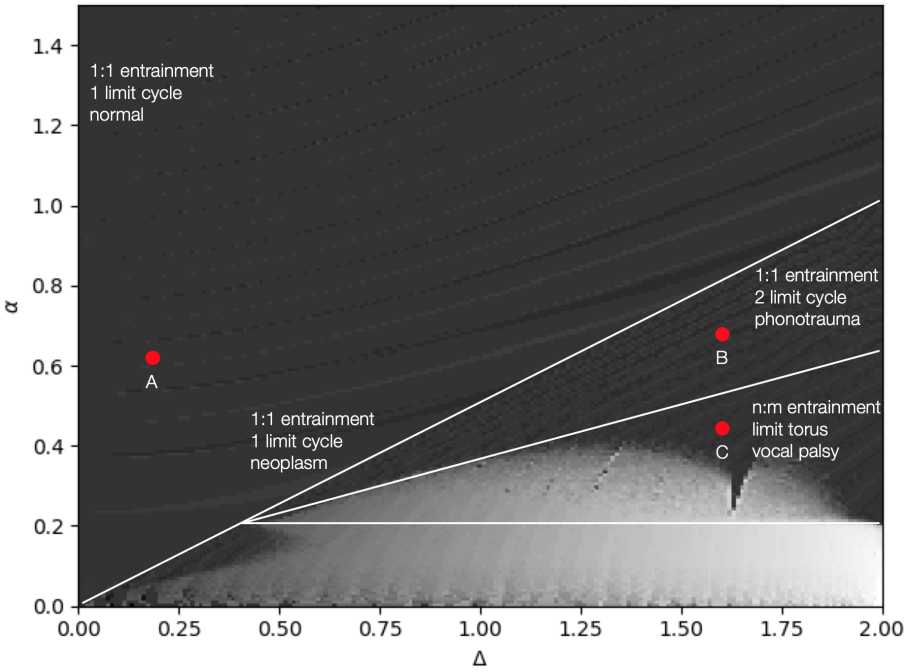
We have introduced the concepts and tools used to study the behaviors (e.g., flow, orbit, attractor, stability, Poincaré map, bifurcation) of nonlinear dynamical systems such as (10) in the previous section. The phase space of the system in (10) (representing vocal fold motion) is four-dimensional and includes states $(\xi_r, \dot{\xi}_r, \xi_l, \dot{\xi}_l)$. For this nonlinear system, it is expected that attractors such as limit cycles or toruses will appear in the phase space. Such phenomena are consequences of specific parameter settings. Specifically, the parameter β determines the periodicity of oscillations; the parameter α and Δ quantify the asymmetry of the displacement of left and right vocal folds and the degree to which one of the vocal folds is out of phase with the other [25, 18]. We can visualize them by plotting the left and right displacements and the phase space portrait.

The coupling of right and left oscillators is described by their *entrainment*; they are in $n : m$ entrainment if their phase θ_r, θ_l satisfy $|n\theta_r - m\theta_l| < C$ where n, m are integers and C is a constant [18]. Such entrainment can be shown by the Poincaré map, where the number of trajectory crossings of the right or left oscillator with the Poincaré section shows the periodicity of its limit cycles. Therefore, their ratio represents the entrainment. We can use the bifurcation diagram to visualize how the entrainment changes with parameters. An example of such a bifurcation diagram is shown in Figure 5 [9, 25]. As we will see later (and as indicated in Figure 5), model parameters can characterize voice pathologies, which will also be visible in phase portrait and bifurcation plots.

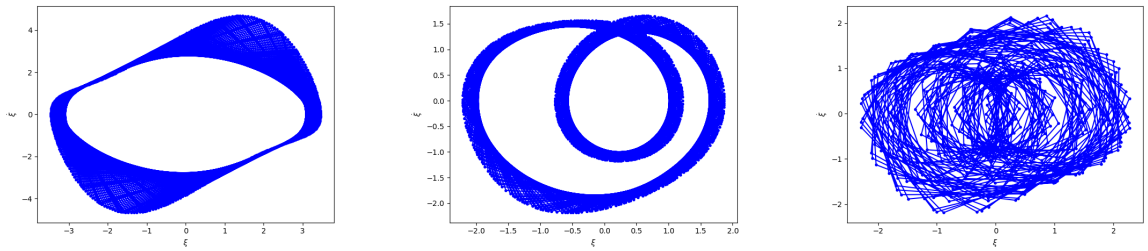
6.3. Experimental results

In the sections above, we have briefly presented some elements of dynamical systems that are important for experimental analysis using the models and solutions proposed in this paper.

We now describe our experimental setup wherein we show the validity of our proposed algorithms ADLES and ADLES-VFT for the analysis of real-world data.



(a) A 3D bifurcation diagram. The third dimension is perpendicular to the parameter plane shown. It shows the entrainment ratio $n : m$ (encoded in different shades of gray) as a function of model parameters α and Δ, where n and m are the number of intersections of the orbits of right and left oscillators across the Poincaré section $\dot{\xi}_{r,l} = 0$ at stable status. This is consistent with the theoretical results in [18].



(b) Phase portraits (phase-space trajectories) for points A (left panel), B (center panel) and C (right panel). The horizontal axis is displacement of a vocal fold, and the vertical axis is its velocity. (b) (c)

Fig. 5. **(a)** Bifurcation diagram of the asymmetric vocal fold model; **(b)** Phase-space trajectories corresponding to the points A, B and C.

6.3.1. Experiment 1: ADLES

We use the ADLES algorithm to estimate the asymmetric model parameters for clinically acquired pathological speech data. The data comprise speech samples collected from subjects suffering from three different vocal pathologies. Our goal is to demonstrate that the individualized phase space trajectories of the asymmetric vocal fold model are discriminative of these disorders.

The data used in our experiments is the FEMH database [52]. It comprises 200 recordings of the sustained vowel `/a:/`. The data were obtained from a voice clinic in a tertiary teaching hospital, and the complete database includes 50 normal voice samples (control set) and 150 samples that represent common voice pathologies. Specifically, the set contains $40/60/50$ samples for glottis neoplasm, phonotrauma (including vocal nodules, polyps, and cysts), and unilateral vocal paralysis, respectively.

Figure 6 shows the glottal flow obtained by inverse filtering and those obtained by the asymmetric model with the parameters estimated by our ADLES method. We observe consistent matches, showing that the ADLES algorithm accurately achieves its objectives in individualizing the asymmetric model to each speaker instance.

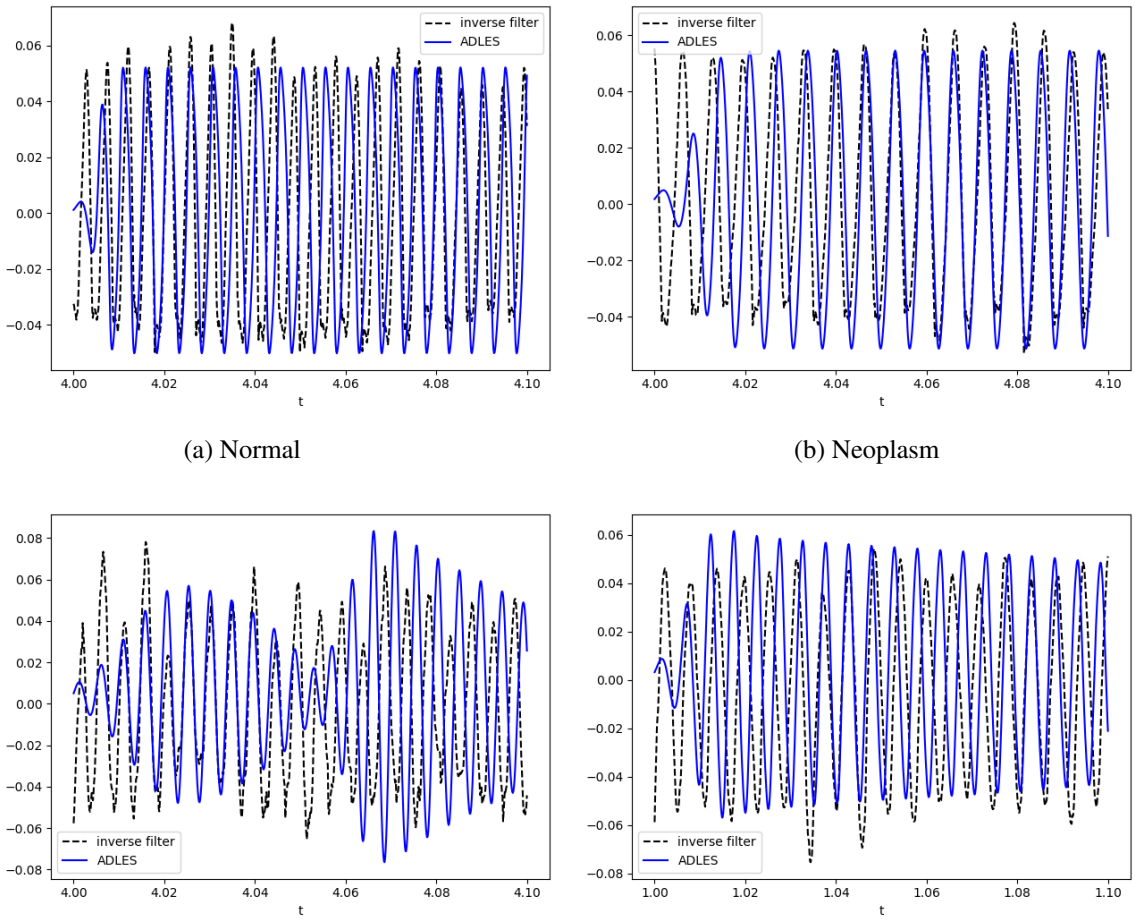
Figure 7 shows some phase portraits of the right and left vocal folds obtained using the ADLES solution. We observe that the attractor behaviors are typical and even visually differentiable for different types of pathologies.

Table 1 shows the results of deducing voice pathologies by simple thresholding of parameter ranges. Specifically, the ranges of model parameters in each row of Table 1 correspond to regions in the bifurcation diagram in Figure 5. Each region has distinctive attractors and phase entrainment, representing distinct vocal fold behaviors and thereby indicating different voice pathologies. By extracting the phase trajectories for the speech signal and, thereby, the underlying system parameters, the ADLES algorithm can place the vocal-fold oscillations in voice production on the bifurcation diagram and thus deduce the pathology.

Δ	α	Phase Space Behavior	Pathology	Accuracy
< 0.5	> 0.25	1 limit cycle, $1:1$ entrain	Normal	0.90
~ 0.6	~ 0.35	1 limit cycle, $1:1$ entrain	Neoplasm	0.82
~ 0.6	~ 0.3	2 limit cycles, $1:1$ entrain	Phonotrauma	0.95
~ 0.85	~ 0.4	toroidal, $n:m$ entrain	Vocal Palsy	0.89

Table 1. Parameters obtained and pathologies identified through ADLES.

Further, we compare the estimation precision of the proposed backward approach and the forward-backward approach. Table 2 shows the mean absolute error (MAE) of calculating glottal flows and parameters for four voice types (normal, neoplasm, phonotrauma, vocal palsy) obtained by backward ADLES (ADLES-B) and forward-backward ADLES, which is the same as ADLES-VFT. The glottal flows obtained by inverse filtering the speech signals are treated as ground truths.



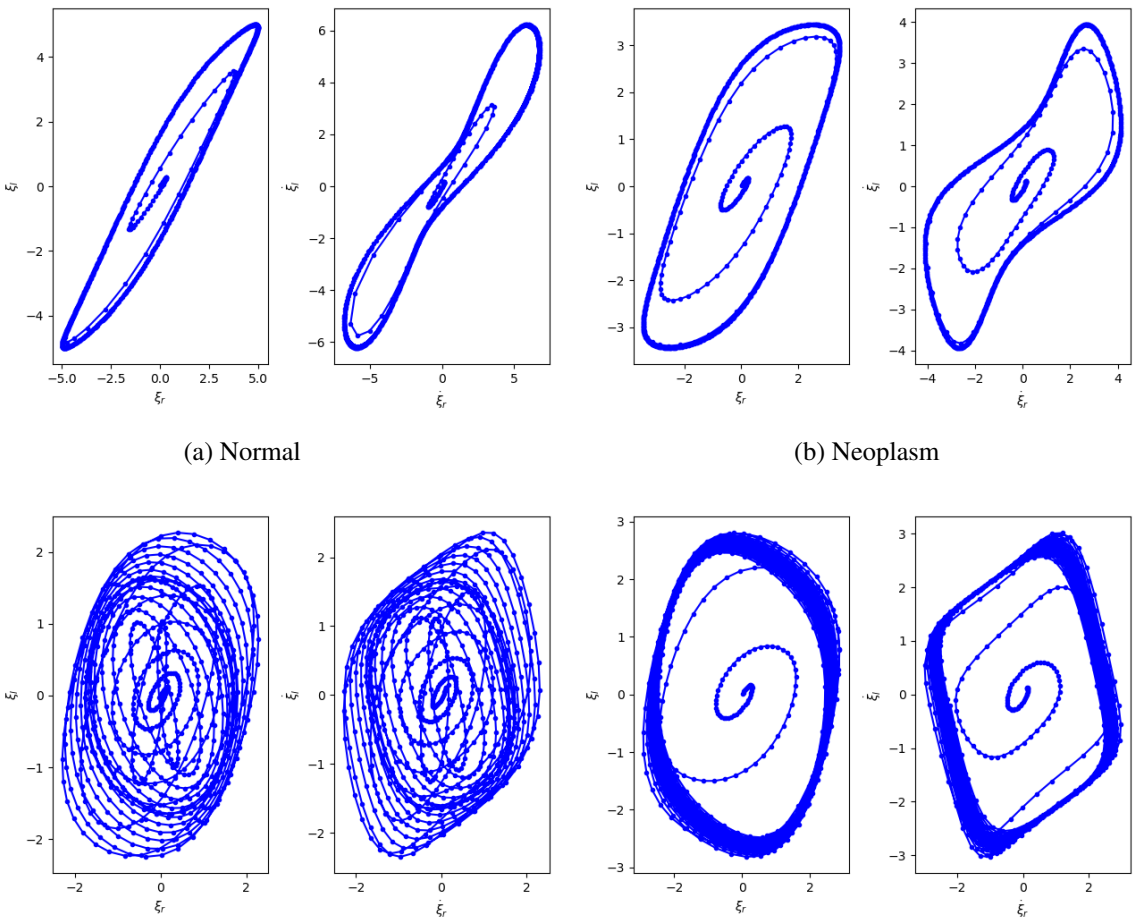
(a) Normal

(b) Neoplasm

(c) Phonotrauma

(d) Vocal palsy

Fig. 6. Glottal flows from inverse filtering and ADLES estimation for **(a)** normal speech (control), **(b)** neoplasm, **(c)** phonotrauma, and **(d)** vocal palsy.



(a) Normal

(b) Neoplasm

(c) Phonotrauma

(d) Vocal palsy

Fig. 7. Phase portraits of left and right oscillators (ADLES-based estimation) for (a) normal speech: 1 limit cycle, (b) neoplasm: 1 limit cycle, (c) phonotrauma: 2 limit cycles, (d) vocal palsy: limit torus. The convergence trajectory is also shown, and the limit cycles can be observed as the emergent geometries in these plots.

Since there is no ground truth for model parameters, we treat the parameters obtained by backward ADLES (ADLES-B) as ground truth. These results suggest that our forward-backward algorithm can effectively recover the vocal tract profile, glottal flow, and model parameters.

	Glottal Flow MAE		Parameter MAE	
	ADLES-B	ADLES-VFT	α	Δ
Normal	0.021	0.022	0.042	0.049
Neoplasm	0.028	0.036	0.055	0.058
Phonotrauma	0.043	0.051	0.083	0.079
Vocal palsy	0.059	0.065	0.102	0.119
All	0.040	0.045	0.074	0.078

Table 2. Estimation error by backward and forward-backward approach.

Further analysis using Lyapunov exponents, Hurst exponents and other primary measurements and secondary characterizations of the derived vocal fold dynamics can be used in other tasks. In the current experiments, the direct use of parameter ranges alone gives high accuracy in the task of classifying the vocal pathologies, and further supplementation is not needed to prove the validity of ADLES and ADLES-VFT in estimating individualized model parameters and consequent phase space behaviors.

6.4. Deriving additional information for voice analysis tasks

A wealth of information can be derived from the solutions of the individualized phonation models to aid machine learning algorithms for specific voice-based detection tasks. For example, features can be derived by performing various measurements on the phase space trajectories represented by the left and right vocal fold oscillations, velocities or accelerations. Such measurements can be performed from statistical, signal processing, information-theoretic, dynamical systems, topological and other perspectives. For example, from an information theoretic perspective, assuming that we find a suitable underlying distribution that fits the vocal fold oscillation trajectories in phase space, we can compute entities such as mutual information of the left and right vocal fold trajectories, conditional entropy of the distribution of the displacement values for the right vocal fold given the displacement values of the left vocal fold and *vice versa*, joint entropy of the distribution of the displacement values for both the vocal folds, rate distortion etc.

In the paragraphs below we mention some easily computable but important features as examples. We do not provide experimental results for these, since it would over-extend this paper. However, we refer to some results later in this section.

Information from a statistical perspective: In addition to standard features such as the amplitudes, range, mean, standard deviation etc. of the displacement data points for the right and left

vocal folds, we can compute many features such as the such as the following:

1. **Pearson correlation coefficient between the right and left vocal folds:** This is given by

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{159}$$

   where r_{xy} is the Pearson correlation coefficient between the left and right vocal fold displacements, x_i and y_i are the left and right vocal fold displacements, respectively, for the ith observation, and $\bar{x}$ and $\bar{y}$ are the sample means of the left and right vocal fold displacements, respectively. The numerator of the formula calculates the covariance between the left and right vocal fold displacements, while the denominator normalizes the covariance by the standard deviations of the left and right vocal fold displacements. The resulting correlation coefficient ranges from -1 to 1, with a value of 1 indicating a perfect positive correlation between the two variables, 0 indicating no correlation, and -1 indicating a perfect negative correlation.

2. **Area of the enclosed region formed by the displacement data points in phase space:** This can be computed using the Shoelace formula, which calculates the area of a polygon by summing the products of the x-coordinates and y-coordinates of adjacent vertices, and then subtracting the products of the x-coordinates and y-coordinates of non-adjacent vertices:

$$A = \frac{1}{2}\left|\sum_{i=1}^{n-1}(x_i y_{i+1} - x_{i+1} y_i) + x_n y_1 - x_1 y_n\right| \tag{160}$$

   where A is the area of the enclosed region, x_i and y_i are the coordinates of the ith data point for the left and right vocal folds, respectively, and n is the total number of data points. The absolute value of the result is taken to ensure that the area is positive, regardless of the order in which the data points are listed.

3. **Slope and intercept of the regression line fitted to the displacement data points of the left and right vocal folds, and error residuals:** These are computed as:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{161}$$

   where b is the slope of the regression line, x_i and y_i are the coordinates of the ith data point for the left and right vocal folds, respectively, and $\bar{x}$ and $\bar{y}$ are the sample means of the left and right vocal fold displacements, respectively. If the slope is positive, it indicates a positive correlation between the two variables, while a negative slope indicates a negative correlation.

   The intercept a of the regression line fitted to the displacement data points is given by $a = \bar{y} - b\bar{x}$. Using this, the error residuals can be calculated as $e_i = y_i - \hat{y}_i$, where e_i is the residual for the ith data point, y_i is the actual value of the right vocal fold displacement for

the ith data point, and $\hat{y}_i$ is the predicted value of the right vocal fold displacement for the ith data point, based on the regression line. The predicted value can be calculated using the equation for the regression line:

$$\hat{y}_i = a + bx_i \tag{162}$$

where a is the intercept of the regression line, b is the slope of the regression line, and x_i is the left vocal fold displacement for the ith data point. These entities are similarly derived for the left vocal fold.

4. **Coefficient of determination (R^2) for the regression line fitted to the displacement data points:** This is given by

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum i = 1^n (y_i - \bar{y})^2}, \tag{163}$$

where R^2 is the coefficient of determination, $\hat{y}_i$ is the predicted value of the right vocal fold displacement for the ith data point, based on the regression line, y_i is the actual value of the right vocal fold displacement for the ith data point, and $\bar{y}$ is the sample mean of the right vocal fold displacements. The numerator of the formula calculates the amount of variation in the right vocal fold displacements that is explained by the regression line, while the denominator calculates the total variation in the right vocal fold displacements. The resulting coefficient of determination ranges from 0 to 1, with higher values indicating a better fit of the regression line to the data. If R^2 is close to 1, it indicates that a large proportion of the variation in the right vocal fold displacements can be explained by the left vocal fold displacements, while if R^2 is close to 0, it indicates that the left vocal fold displacements are not a good predictor of the right vocal fold displacements.

Information from a signal processing perspective: Since the vocal fold oscillations are a time-series, all standard signal-based measurements for time series data can be applied to them. Some fairly obvious features that can be computed without applying data transforms are:

1. **Frequency of the displacement oscillations for the right and left vocal folds, and phase difference between them:** $f_r = \frac{1}{T_r}$ and $f_l = \frac{1}{T_l}$, where f_r and f_l are the frequencies of the displacement oscillations, and T_r and T_l are the periods of the oscillations for the right and left vocal folds, respectively. The phase difference between the right and left vocal folds $\Delta\phi$ is given by:

$$\Delta\phi = \frac{2\pi\Delta t}{T}, \tag{164}$$

where Δt is the time difference between the peak displacements of the left and right vocal folds, and T is the period of the oscillations. The phase difference represents the amount by which the displacement of the right vocal fold lags behind the displacement of the left vocal

fold, and is typically measured in radians. A positive phase difference indicates that the right vocal fold displacement lags behind the left vocal fold displacement, while a negative phase difference indicates that the right vocal fold displacement leads the left vocal fold displacement.

2. **Amplitude ratio of the right to left vocal folds:** We can compute the amplitude ratio, A_r, of the displacement data points of the right vocal fold to the displacement data points of the left vocal fold as follows:

$$A_r = \frac{A_{r,max}}{A_{l,max}}, \tag{165}$$

where $A_{r,max}$ and $A_{l,max}$ are the maximum displacements of the right vocal and left vocal folds respectively. This is 1 when right and left vocal fold displacements are equal in magnitude.

Information from a topological perspective: Topological measurements provide insights into the geometric and structural properties of the vocal fold displacement data. Some common ones are listed below.

1. **Fractal dimension:** This is a measure of the complexity or self-similarity of the data, given by the scaling exponent of the number of points or simplices needed to cover the data as a function of the size or resolution.

   We can compute the fractal dimension of the displacement data points of the left and right vocal folds using the box-counting method as follows:

$$D_{left} = \lim_{\epsilon \to 0} \frac{\log(N_\epsilon^l)}{\log(\frac{1}{\epsilon})} \tag{166}$$

$$D_{right} = \lim_{\epsilon \to 0} \frac{\log(N_\epsilon^r)}{\log(\frac{1}{\epsilon})} \tag{167}$$

   where D_{left} (or D_{right}) is the fractal dimension of the left (or right) vocal fold displacement data, N_ϵ^l (or N_ϵ^r) is the number of ϵ-sized boxes needed to cover the left (or right) vocal fold displacement data, and ϵ is the size of the boxes.

   Note: This implicit assumption here is that the displacement data points of the vocal folds can be treated as fractal objects and that their fractal dimension can be computed using the box-counting method. In actual implementations, an appropriate box size must be chosen.

2. **Homology:** This is a measure of the topological structure of the data, which characterizes the number and type of holes and voids in the data. This can be computed using algebraic topology methods such as persistent homology, which captures the topological features that persist over a range of threshold values.

To compute the homology of the displacement data points of the left and right vocal folds, and of a graph of the displacement data points of the left vocal fold versus the displacement data points of the right vocal fold, we first need to define the simplicial complex associated with the data points.

Let $L = (x_1, y_1), \ldots, (x_n, y_n)$ be the set of n data points for the left vocal fold displacement, and $R = (z_1, w_1), \ldots, (z_m, w_m)$ be the set of m data points for the right vocal fold displacement. We can construct a simplicial complex K_L associated with L as follows:

For each data point $(x_i, y_i) \in L$, we include a 0-simplex $[i]$ in K_L. For each pair of data points (x_i, y_i) and (x_j, y_j) in L such that $|x_i - x_j| + |y_i - y_j| \leq \epsilon$, where ϵ is a small positive number, we include a 1-simplex $[i, j]$ in K_L. For each triple of data points $(x_i, y_i), (x_j, y_j), (x_k, y_k)$ in L such that $|x_i - x_j| + |y_i - y_j| \leq \epsilon$, $|x_j - x_k| + |y_j - y_k| \leq \epsilon$, and $|x_i - x_k| + |y_i - y_k| \leq \epsilon$, we include a 2-simplex $[i, j, k]$ in K_L.

Similarly, we can construct a simplicial complex K_R associated with R.

To compute the homology of the displacement data points of the left and right vocal folds, we can compute the homology groups $H_i(K_L)$ and $H_i(K_R)$, respectively, using any standard method, such as the Smith normal form algorithm.

To compute the homology of the graph of the displacement data points of the left vocal fold versus the displacement data points of the right vocal fold, we can construct a new simplicial complex K as follows:

For each pair of data points $(x_i, y_i) \in L$ and $(z_j, w_j) \in R$ such that $|x_i - z_j| + |y_i - w_j| \leq \epsilon$, we include a 1-simplex $[i, j]$ in K.

We can then compute the homology groups $H_i(K)$ using any standard method.

3. **Betti numbers:** These are a set of integer-valued topological invariants that count the number of k-dimensional holes or loops in the data, where k ranges from 0 (connected components) to 2 (voids or cavities).

   The Betti numbers of the displacement data points of the left and right vocal folds can be computed using the persistent homology approach. Let X be a finite set of points in $\mathbb{R}^n$ representing the displacement data points. The k-th Betti number, denoted as β_k, is defined as the rank of the k-th homology group of a simplicial complex constructed from X.

   The construction of the simplicial complex can be done using the Vietoris-Rips complex, which forms a complex by connecting any two points in X if their distance is less than or equal to a given parameter ϵ. The k-th homology group is obtained by taking the quotient of the k-th cycle group and the k-th boundary group, where a k-cycle is a collection of k-dimensional simplices that form a closed loop, and a $(k+1)$-boundary is the set of $(k+1)$-dimensional simplices that form the boundary of the k-cycle.

   For the graph of the displacement data points of the left vocal fold versus the displacement data points of the right vocal fold, the Betti numbers can be computed by constructing a

simplicial complex on the product space of the left and right vocal fold point sets. Let X_L and X_R be the left and right vocal fold point sets, respectively. The simplicial complex is constructed by connecting any two points (x_L, x_R) and (y_L, y_R) if the distance between (x_L, x_R) and (y_L, y_R) is less than or equal to ϵ. The Betti numbers are then computed as above, by taking the homology groups of the resulting simplicial complex.

4. **Euler characteristic:** a scalar topological invariant that captures the overall shape and connectivity of the data, given by the alternating sum of the number of vertices, edges, and faces (or higher-dimensional simplices) in the data. This can be computed using methods such as discrete Morse theory or Euler calculus.

   The Euler characteristic χ of a graph of the displacement data points of the left vocal fold versus the displacement data points of the right vocal fold can be computed using the formula:

   $$\chi = V - E + F$$

   where V is the number of vertices, E is the number of edges, and F is the number of faces in the graph.

   In this case, each point in the left vocal fold point set is paired with a corresponding point in the right vocal fold point set to form a vertex in the graph. Hence, the number of vertices V is equal to the cardinality of the left vocal fold point set, which we can denote as $|X_L|$.

   Each pair of points (x_L, x_R) and (y_L, y_R) in the left and right vocal fold point sets that are connected by an edge in the graph satisfies the condition that the distance between them is less than or equal to ϵ. Hence, an edge in the graph corresponds to a pair of points in the left and right vocal fold point sets that are within a distance of ϵ of each other. The number of edges E is equal to the number of such pairs of points, which we can denote as $|X_L \times X_R \cap B_\epsilon|$ where B_ϵ is the ϵ-ball in $\mathbb{R}^n$.

   Finally, the faces in the graph correspond to cycles of edges that form closed loops. The number of faces F can be computed using the formula:

   $F = E - V + C$

   where C is the number of connected components in the graph. In this case, since the graph is undirected, C is equal to the number of connected components in the graph when considered as an undirected graph.

   Once the values of V, E, F, and C have been computed, the Euler characteristic χ can be computed using the formula above.

Information from a dynamical systems perspective can give insights about the underlying mechanisms and principles that govern the vocal fold dynamics. Examples of features in this

category are recurrence analysis features, Lyapunov exponents, Hurst exponents etc. These are mentioned in earlier sections of this paper.

Some of the features mentioned above have been used in real-world applications and proven to be effective. For example, in [53], the authors hypothesize that since COVID-19 impairs the respiratory system, effects on the phonation process could be expected, and signatures of COVID-19 could manifest in the vibration patterns of the vocal folds. In this paper, features have been derived from a signal processing perspective.

This study used the ADLES method to estimate the asymmetric vocal folds model parameters. It further used the parameters and estimation residuals as features to other binary classifiers such as logistic regression, support vector machine, decision tree, and random forest, achieving around 0.8 ROC-AUC (area under the ROC curve) in discriminating positive COVID-19 cases from negative instances, on clinically collected and curated data. The data used contained recordings of extended vowel sounds from affected speakers and control subjects. The authors also discovered that COVID-19 positive individuals display different phase space behaviors from negative individuals: the phase space trajectories for negative individuals were found to be more regular and symmetric across the two vocal folds, while the trajectories for positive patients were more chaotic, implying a lack of synchronization and a higher degree of asymmetry in the vibrations of the left and right vocal folds.

In a companion study, the authors in [54] used the ADLES-estimated glottal flows as features to CNN-based two-step attention neural networks. The neural model detects differences in the estimated and actual glottal flows and predicts two classes corresponding to COVID-19 positive and negative cases. This achieved 0.9 ROC-AUC (normalized) on clinically collected vowel sounds. Yet another study used higher order statistics derived from parameters, and Lyapunov and Hurst exponents derived from the phase space trajectories of the individualized asymmetric models, to detect Amyotrophic Lateral Sclerosis (ALS) from voice with high accuracy (normalized ROC-AUC of 0.82 to 0.99)) [55].

7. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we have presented a dynamical system perspective for physical process modeling and phase space characterization of phonation, and proposed a framework wherein these can be derived for individual speakers from recorded speech samples. The oscillatory dynamics of vocal folds provide a tool to analyze different phonation phenomena in many real-world task settings. We have proposed a backward approach for modeling vocal fold dynamics, and an efficient algorithm (the ADLES algorithm) to solve the inverse problem of estimating model parameters from speech observations. Further, we have integrated the vocal tract and vocal folds models, and have presented a forward-backward paradigm (the ADLES-VFT algorithm) for effectively solving the inverse problem for the coupled vocal fold-tract model.

We have shown that the parameters estimated by these algorithms allow the models to closely emulate the vocal fold motion of individual speakers. Features and statistics derived from the model dynamics are (at least) discriminative enough for use in regular machine-learning based

classification algorithms to accurately identify various voice pathologies from recorded speech samples. In future, these approaches are expected to be helpful in deducing many other underlying influences on the speaker's vocal production mechanism.

Finally, extensions of these approaches can use other physical models of voice production, and other physical processes including phonation. The phase space characterization presented in this paper is based on phase space trajectories (a topological perspective). Another direction of suggested research is characterizing the phase space from algebraic perspectives. We can recast the study of the topological structures of the phase space to the study of its algebraic constructs, such as homotopy groups and homology/cohomology groups, which are easier to classify. For example, algebraic invariants can characterize the homeomorphisms between phase spaces (e.g., evolution maps, poincaré maps) and reveal large-scale structures and global properties (e.g., existence and structure of orbits). We can also explore and build upon the deep connection between dynamical systems and deep neural models. We can study deep learning approaches for solving and analyzing dynamical systems, and explore the integration of dynamical systems with deep neural models to analyze and interpret the behaviors of the vocal folds.

8. REFERENCES

[1] L. Cveticanin, "Review on mathematical and mechanical models of the vocal cord," *Journal of Applied Mathematics*, 2012.

[2] I. R. Titze, "The physics of small-amplitude oscillation of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1536–1552, 1988.

[3] R. Singh, *Profiling humans from their voice*. Springer, 2019.

[4] J. Flanagan and L. Landgraf, "Self-oscillating source for vocal-tract synthesizers," *IEEE Transactions on Audio and Electroacoustics*, vol. 16, no. 1, pp. 57–64, 1968.

[5] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell system technical journal*, vol. 51, no. 6, pp. 1233–1268, 1972.

[6] Z. Zhang, J. Neubauer, and D. A. Berry, "The influence of subglottal acoustics on laboratory models of phonation," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, pp. 1558–1569, 2006.

[7] W. Zhao, C. Zhang, S. H. Frankel, and L. Mongeau, "Computational aeroacoustics of phonation, part i: Computational methods and sound generation mechanisms," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2134–2146, 2002.

[8] C. Zhang, W. Zhao, S. H. Frankel, and L. Mongeau, "Computational aeroacoustics of phonation, part ii: Effects of flow parameters and ventricular folds," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2147–2154, 2002.

[9] J. C. Lucero, "Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region," *The Journal of the Acoustical Society of America*, vol. 94, no. 6, pp. 3104–3111, 1993.

[10] J. C. Lucero and J. Schoentgen, "Modeling vocal fold asymmetries with coupled van der pol oscillators," in *Proceedings of Meetings on Acoustics ICA2013*, vol. 19, no. 1. ASA, 2013, p. 060165.

[11] F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," *The Journal of the Acoustical Society of America*, vol. 108, no. 6, pp. 3003–3012, 2000.

[12] A. Yang, M. Stingl, D. A. Berry, J. Lohscheller, D. Voigt, U. Eysholdt, and M. Döllinger, "Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model," *The Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. 948–964, 2011.

[13] B. A. Pickup and S. L. Thomson, "Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models," *Journal of biomechanics*, vol. 42, no. 14, pp. 2219–2225, 2009.

[14] J. J. Jiang, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," *Journal of the Acoustical Society of America*, vol. 10, no. 4, pp. 2120–2128, 2001.

[15] I. R. Titze, "Nonlinear source–filter coupling in phonation: Theory," *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. 1902–1915, 2008.

[16] B. H. Story and I. R. Titze, "Voice simulation with a body-cover model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1249–1260, 1995.

[17] R. W. Chan and I. R. Titze, "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2351–2362, 2006.

[18] J. C. Lucero, J. Schoentgen, J. Haas, P. Luizard, and X. Pelorson, "Self-entrainment of the right and left vocal fold oscillators," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 2036–2046, 2015.

[19] S. Maeda, "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech production and speech modelling*. Springer, 1990, pp. 131–149.

[20] P. Birkholz and B. J. Kröger, "Simulation of vocal tract growth for articulatory speech synthesis," in *Proceedings of the 16th international congress of phonetic sciences*, 2007, pp. 377–380.

[21] J. Dang and K. Honda, "Construction and control of a physiological articulatory model," *The Journal of the Acoustical Society of America*, vol. 115, no. 2, pp. 853–870, 2004.

[22] M. R. Portnoff, "A quasi-one-dimensional digital simulation for the time-varying vocal tract." Ph.D. dissertation, Massachusetts Institute of Technology, 1973.

[23] D. R. Allen and W. J. Strong, "A model for the synthesis of natural sounding vowels," *The Journal of the Acoustical Society of America*, vol. 78, no. 1, pp. 58–69, 1985.

[24] K. Motoki, X. Pelorson, P. Badin, and H. Matsuzaki, "Computation of 3-d vocal tract acoustics based on mode-matching technique," in *Sixth International Conference on Spoken Language Processing*, 2000.

[25] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *The Journal of the Acoustical Society of America*, vol. 97, no. 3, pp. 1874–1884, 1995.

[26] W. Zhao and R. Singh, "Speech-based parameter estimation of an asymmetric vocal fold oscillation model and its application in discriminating vocal fold pathologies," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 7344–7348.

[27] P. Mergell, H. Herzel, and I. R. Titze, "Irregular vocal-fold vibration—high-speed observation and modeling," *The Journal of the Acoustical Society of America*, vol. 108, no. 6, pp. 2996–3002, 2000.

[28] C. Tao, Y. Zhang, D. G. Hottinger, and J. J. Jiang, "Asymmetric airflow and vibration induced by the coanda effect in a symmetric model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 2270–2278, 2007.

[29] V. Isakov, *Inverse problems for partial differential equations*. Springer, 2006, vol. 127.

[30] C. Tao, Y. Zhang, G. Du, and J. J. Jiang, "Estimating model parameters by chaos synchronization," *Physical Review E*, vol. 69, no. 3, p. 036204, 2004.

[31] Y. Zhang, C. Tao, and J. J. Jiang, "Parameter estimation of an asymmetric vocal-fold system from glottal area time series using chaos synchronization," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 16, no. 2, p. 023118, 2006.

[32] S. J. Rupitsch, J. Ilg, A. Sutor, R. Lerch, and M. Döllinger, "Simulation based estimation of dynamic mechanical properties for viscoelastic materials used for vocal fold models," *Journal of Sound and Vibration*, vol. 330, no. 18-19, pp. 4447–4459, 2011.

[33] C. Tao, Y. Zhang, and J. J. Jiang, "Extracting physiologically relevant parameters of vocal folds from high-speed video image series," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 794–801, 2007.

[34] P. Birkholz, D. Jackèl, and B. J. Kroger, "Construction and control of a three-dimensional vocal tract model," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 1. IEEE, 2006, pp. I–I.

[35] J. Mullen, D. M. Howard, and D. T. Murphy, "Real-time dynamic articulations in the 2-d waveguide mesh vocal tract model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 577–585, 2007.

[36] B. D. Erath and M. W. Plesniak, "An investigation of jet trajectory in flow through scaled vocal fold models with asymmetric glottal passages," *Experiments in fluids*, vol. 41, no. 5, pp. 735–748, 2006.

[37] P. Alku, "Glottal inverse filtering analysis of human voice production—a review of estimation and parameterization methods of the glottal excitation and their applications," *Sadhana*, vol. 36, no. 5, pp. 623–650, 2011.

[38] M. P. Do Carmo and J. Flaherty Francis, *Riemannian geometry*. Springer, 1992, vol. 6.

[39] P. M. Morse and K. U. Ingard, *Theoretical acoustics*. Princeton university press, 1986.

[40] I. R. Titze and D. W. Martin, "Principles of voice production," 1998.

[41] L. V. Kantorovich and G. P. Akilov, *Functional analysis*. Elsevier, 2016.

[42] K. Zhu, *Operator theory in function spaces*. American Mathematical Soc., 2007, no. 138.

[43] M. B. Giles and E. Süli, "Adjoint methods for pdes: a posteriori error analysis and postprocessing by duality," *Acta numerica*, vol. 11, pp. 145–236, 2002.

[44] C. Dong and Y. Jin, "Mimo nonlinear ultrasonic tomography by propagation and backpropagation method," *IEEE transactions on image processing*, vol. 22, no. 3, pp. 1056–1069, 2012.

[45] H. P. Langtangen and K.-A. Mardal, *Introduction to numerical methods for variational problems*. Springer Nature, 2019, vol. 21.

[46] W. F. Ames, *Numerical methods for partial differential equations*. Academic press, 2014.

[47] V. Thomée, *Galerkin finite element methods for parabolic problems*. Springer, 1984, vol. 1054.

[48] M. G. Larson and F. Bengzon, "The finite element method: theory, implementation, and practice," *Texts in Computational Science and Engineering*, vol. 10, pp. 23–44, 2010.

[49] G. D. Birkhoff, *Dynamical systems*. American Mathematical Soc., 1927, vol. 9.

[50] J. J. Jiang, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," *The Journal of the Acoustical Society of America*, vol. 110, no. 4, pp. 2120–2128, 2001.

[51] J. J. Jiang and Y. Zhang, "Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2127–2133, 2002.

[52] C. Bhat and S. K. Kopparapu, "Femh voice data challenge: Voice disorder detection and classification using acoustic descriptors," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 5233–5237.

[53] M. Al Ismail, S. Deshmukh, and R. Singh, "Detection of covid-19 through the analysis of vocal fold oscillations," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1035–1039.

[54] S. Deshmukh, M. Al Ismail, and R. Singh, "Interpreting glottal flow dynamics for detecting covid-19 from voice," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1055–1059.

[55] J. Zhang, "Vocal fold dynamics for automatic detection of amyotrophic lateral sclerosis from voice," Master's thesis, Computational Biology Department, Carnegie Mellon University, USA, 5 2022, undergraduate thesis.