Article

Deriving Vocal Fold Oscillation Information from Recorded Voice Signals Using Models of Phonation

Wayne Zhao [1] and Rita Singh [2,*]

1   Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA; wenbozhao@cmu.edu
2   School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
*   Correspondence: rsingh@cs.cmu.edu

Abstract: During phonation, the vocal folds exhibit a self-sustained oscillatory motion, which is influenced by the physical properties of the speaker's vocal folds and driven by the balance of bio-mechanical and aerodynamic forces across the glottis. Subtle changes in the speaker's physical state can affect voice production and alter these oscillatory patterns. Measuring these can be valuable in developing computational tools that analyze voice to infer the speaker's state. Traditionally, vocal fold oscillations (VFOs) are measured directly using physical devices in clinical settings. In this paper, we propose a novel analysis-by-synthesis approach that allows us to infer the VFOs directly from recorded speech signals on an individualized, speaker-by-speaker basis. The approach, called the ADLES-VFT algorithm, is proposed in the context of a joint model that combines a phonation model (with a glottal flow waveform as the output) and a vocal tract acoustic wave propagation model such that the output of the joint model is an estimated waveform. The ADLES-VFT algorithm is a forward-backward algorithm which minimizes the error between the recorded waveform and the output of this joint model to estimate its parameters. Once estimated, these parameter values are used in conjunction with a phonation model to obtain its solutions. Since the parameters correlate with the physical properties of the vocal folds of the speaker, model solutions obtained using them represent the individualized VFOs for each speaker. The approach is flexible and can be applied to various phonation models. In addition to presenting the methodology, we show how the VFOs can be quantified from a dynamical systems perspective for classification purposes. Mathematical derivations are provided in an appendix for better readability.

Keywords: vocal fold oscillation; phonation models; dynamical systems; parameter estimation; 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 vibration [1,2]. During this process, depending on the physical state of the vocal folds, their eigenmodes of vibration synchronize, or strive to do so. This self-sustained motion of the vocal folds during phonation is highly sensitive to perturbations caused by many possible influencing factors, which may affect the speaker during speech production. In recent years, there has been a surge of interest in building voice-based diagnostic aids—computational models based on artificial intelligence and machine learning that can infer the speaker's state (and thereby the factors that are affecting the speaker) from voice. Such applications can benefit greatly from being able to deduce the fine-level nuances in the motion of the vocal folds, and being able to measure the response to various perturbing factors to infer their nature. However, doing this using traditional methods on an individual basis for each speaker is very difficult. Traditional methods to observe and record vocal fold oscillations

(VFOs) are based on actual physical measurements taken using various instruments in clinical settings, e.g., [3–5].

The primary focus of this paper is to derive the VFOs for a speaker directly from recorded voice signals, alleviating the need for taking physical measurements. The solution we propose is an analysis-by-synthesis approach, based on physical models of phonation. We propose a methodology to deduce the parameters of a chosen phonation model from a speaker's voice recording, which can then be substituted into a VFO model to obtain speaker-specific solutions, which represent the vocal fold oscillations of the speaker.

In the paragraphs below, we first review some relevant facts about the process of phonation. In the sections that follow, we present our proposed methodology in two stages: in the first, we show how we can infer the parameters of a physics-based model of phonation from measurements of glottal excitation. Since the parameters of such models represent the physical properties of the speaker's vocal folds, using them to obtain solutions to the models gives us estimate of the speaker's VFOs. We subsequently show how to extend the model to include the physics of mucosal wave propagation in the vocal tract, and propose a forward-backward algorithm to estimate the parameters of the joint model. These can then be used in the corresponding phonation model to obtain its solutions.

Later on, we also explain how the solutions of these models (which are the deduced VFOs) can be characterized from a dynamical systems perspective to derive discriminative information in the form of features that are useful for classification tasks.

In this context, it is important to note at the outset that the models we use to demonstrate our methodology are simple and well-established models in the literature. It is *not* the goal of this paper to propose new models of phonation, but rather to propose a novel methodology to derive their parameters from speech signals, so that they can be solved to yield VFOs on a speaker-by-speaker basis. The main contribution of this paper lies in the derivation of the parameters of these models, and their individualized VFO solutions. The viability of these solutions is demonstrated experimentally using classification experiments.

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 phonation 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 physical phenomena such as the Bernoulli effect and the Coandă effect [6–8]. Figure 1 illustrates the interaction between these effects that is largely thought to drive the oscillations of the vocal folds.

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) [9].

The exact physics of the airflow through the glottis during phonation is well studied, e.g., [2,10–14]. The current understanding from these, from the airflow perspective, is that the glottis 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 build-up 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.
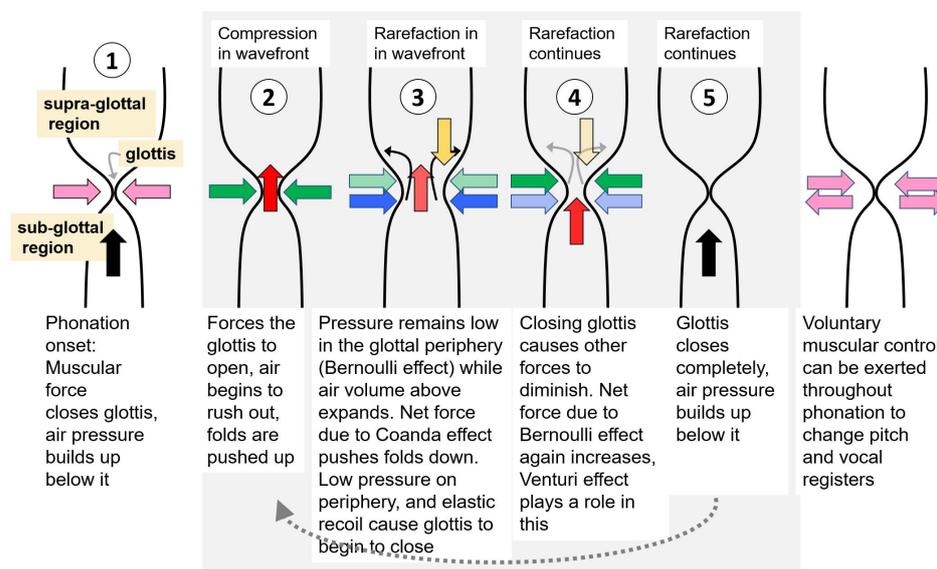

Figure 1. Schematic of the balance of some key forces through one cycle of the self-sustained vibrations of the vocal folds. Sequential "snapshots" of the cycle are numbered 1–5. Arrows depict the direction and type of forces in the glottal area. The color codes for the arrows depict net forces due to 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 weaker forces. Figure from [9] with permission.

1.2. General Approaches to Phonation Modeling

Physical models of phonation, e.g., [9,11,15–20], 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 speechsounds, 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 [21]. Each sub-process model has associated parameters that determine the model output, given an input.

Following the above division of the process, we identify the two following model types, each modeling one of the sub-processes: (i) *Vocal fold oscillation (VFO) models* (also called *vocal folds models*, or *oscillation models*), and (ii) *vocal tract (VT) models*.

The VFO 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., [2,16,22–24], two-mass models, e.g., [11,15], multi-mass models [18], and finite element models, e.g., [17].

Each of these has proven to be useful in different contexts. The **VT** 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, e.g., [25], geometric models, e.g., [26] and biomechanical models, e.g., [27].

In addition, different models are also applied to describe the aero-acoustic interaction of the glottal airflow and the vocal tract. Some examples of these are reflection-type line analog models and transmission line circuit analog models, e.g., [28], hybrid time-frequency domain models, e.g., [29] and finite-element models, e.g., [30].

2. The Problem of Parameter Estimation

Each of the models mentioned in Section 1.2 includes a set of parameters that determine its *state*, and *output*, given an input. For instance, given the parameters for a VFO model, the glottal flow waveform can be determined; given the glottal flow waveform as an input, and the parameters for a VFO model, the acoustic signal can be determined.

The problem we tackle in this paper is the *inverse* problem: given a model output, we must estimate the model parameters from it.

This can be of great practical use. For example, with speaker-specific parameter setting, the output of these models can be used as a proxy for the actual vocal fold motion of the corresponding speaker. To obtain the parameters of such models, the traditional solution has been to take actual physical measurements of the vocal fold oscillations, or of the glottal flow using techniques such as high-speed videostroboscopy, as mentioned in Section 1. This is not always feasible.

On the other hand, the inverse problem of parameter estimation of phonation models is quite difficult to solve through purely computational means. For example, in order to estimate the parameters of a vocal tract (VT) 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 becomes eventually intractable. Some 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.

This paper proposes a methodology for solving the inverse problem of phonation models through purely computational means. As mentioned earlier, the methodology follows an analysis-by-synthesis approach. We explain this by first reviewing our previously proposed Adjoint Least Squares Estimation (ADLES) algorithm [31] that estimates the parameters of a VFO model by minimizing the error between a reference glottal flow waveform and the signal generated by the physical VFO model. We then describe our proposed ADLES-VFT algorithm to estimate the parameters of a *joint* VFO and VT model (also called a body-cover model). Instead of comparing the model-generated excitation at the glottis to a reference glottal flow, the flow is propagated through the vocal tract to generate a signal at the lips, which is compared to a recorded voice signal which is used as a reference. The algorithm proposed iteratively re-estimates the model parameters by minimizing the error between the reference voice sample and this generated signal.

Once estimated, these parameters are used with the VFO model to generate the speaker's VFOs.

3. Vocal Folds, Vocal Tract and Joint Models

In this section we describe the VFO and VT models that we use as examples in this paper. We also explain the formulation of the joint model that we ultimately use to solve the inverse problem of parameter estimation.

3.1. The VFO Model

A schematic illustration of a general mass-spring oscillator model for the vocal folds is shown in Figure 2. This is used to model the phonation process as described below.
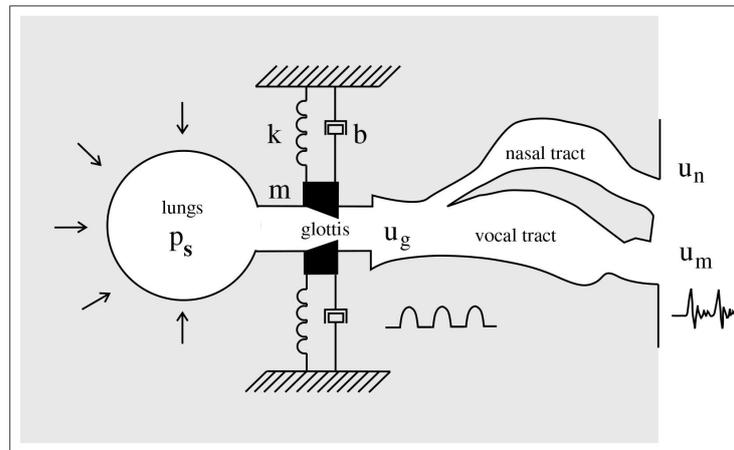
Figure 2. Approximating the vocal folds with mass-spring oscillators in the phonation process. Airflow from the lungs, driven by the subglottal pressure P_s, passes through the glottis, and vocal folds are set into a state of self-sustained vibration, producing the glottal flow u_g which is a quasi-periodic 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.

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].

Other models, namely two-mass, multi-mass and finite element models can also be used as the basis for the VFO model, and are described briefly in the Appendix A for reference.

For our paper, we adopt the version of the one-mass model of the vocal folds proposed in [24], illustrated in Figure 3. This is an asymmetric body-cover model which models the left and right vocal folds individually as one-mass components of a coupled dynamical system. It incorporates an asymmetry parameter, which can emulate the asymmetry in the vibratory motions of left and right vocal folds, and hence is also ideally suited to modeling pathological or atypical phonation [32].

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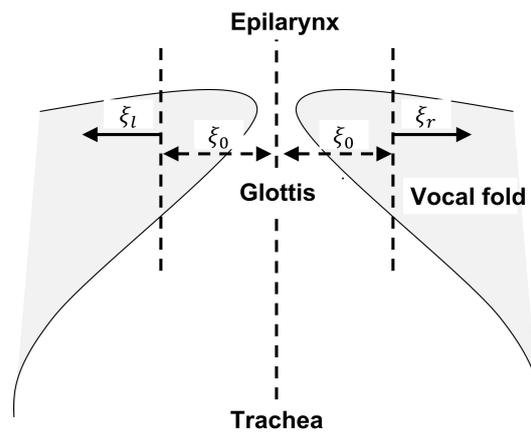(e)   The small-amplitude body-cover assumption is applicable (see definition below).

Figure 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.

Assumption 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.

In the one-mass asymmetric model proposed in [24], with reference to Figure 3, the center-line of the glottis is denoted 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:

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

$$\ddot{\xi}_l + \beta(1 + \xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{3}$$

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 $\Delta = 0$.

3.2. The VT Model

The literature describes a number of different approaches to modeling the vocal-tract, including bio-mechanical models, statistical models, and geometric models. For reference, they are described briefly in the Appendix B.

In our work we use an acoustic wave propagation model described by PDEs for the vocal tract. The vocal tract itself is represented as tube of length L, beginning at the glottis and ending at the lips. Representing the distance along the central axis of the vocal tract as x, $x \in (0, L)$ (where $x = 0$ at the glottis and $x = L$ at the lips) and the time-varying volume velocity of air at any position x along the vocal tract as $u(x, t)$, it can be shown that the PDE that describes $u(x, t)$ is given by

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} + f(x, t) \tag{4}$$

where $f(x, t)$ represents the *vocal tract profile*, which models the characteristics of the vocal tract, including the effect of the nonuniform yielding wall on the acoustic flow dynamics, the effect of vocal tract coupling, lip radiation, etc., and must also be estimated by our algorithm. The derivation of Equation (4) is given in Appendix B. Note that if the vocal tract is assumed to be a rigid tube $f(x, t) = 0$ and Equation (4) reduces to the well-known Webster-Horn equation [33]. In deriving Equation (4) we have assumed a static vocal tract, i.e., that the cross-sectional area $A(x)$ of the vocal tract at any position x is constant. This assumption is valid during phonation, in particular during the steady state of sustained phonation; however our solution can also be extended to consider time-varying vocal tracts $A(x, t)$, although we have not done so in this paper.

The oscillation of the vocal folds results in the movement of air with a time-varying volume velocity $u_0(t) = u(0, t)$ at the glottis. The vocal tract modulates this to result in the volume velocity $u_L(t) = u(L, t)$ at the lips and the corresponding pressure wave $p_L(t)$, where [34]

$$u_L(t) = \frac{A(L)}{\rho c} p_L(t) \tag{5}$$

where $A(L)$ is the opening area at the lip, c is the speed of sound, and ρ is the ambient air density.

4. Estimation of Model Parameters: Solving the Inverse Problem

Our objective is to derive vocal fold oscillations during phonation directly from the speech signal. In order to do so, we will utilize the VFO and VT models.

The *VFO* model represents the actual dynamics of the vocal folds. Given the model parameters, which are α, β and Δ for the coupled one-mass model of Equation (3), it can be used to compute the vocal fold oscillations and the volume velocity of air at the glottis. The *VT* model represents the dynamics of the vocal tract. Given an excitation (at the glottis), it can be used to compute the pressure wave at the lips, which manifests as sound.

Ours is the *inverse* problem: given the the pressure wave at the output of the vocal tract (i.e., the recorded speech signal) we must estimate the VFO parameters that could generate it. This is an *analysis-by-synthesis* problem: in order to analyze the given voice signal, we must identify the model parameters that synthesize the closest (in a metric sense) facsimile to it.

We present the solution in a two-step manner. In the first, given the actual excitation of the vocal tract to produce the voice signal, the parameters of the VFO model are estimated to minimize the error between its output and the reference excitation signal. We refer to this as the *backward* approach (and the corresponding estimation as the "backward" problem), since the reference excitation signal itself must first be derived by passing the voice signal through an inverse model of the vocal tract, i.e., "backward" through the vocal tract. We have previously described our solution to the backward problem in [31], and restate it in Section 4.1 for completeness.

The second, more complete solution considers the joint model, i.e., both the motions of the vocal folds and the propagation of the resulting air flow (the excitation) through the vocal tract. The model parameters are estimated by comparing the signal produced at the lips by the joint model to the recorded voice signal. We refer to this as the "forward-backward" approach since this requires forward propagating the output of the VFO through the VT model, prior to applying the backward approach. The solution to this problem is the primary contribution of this paper.

The two approaches are illustrated in Figure 4.

Once estimated, the VFO model parameters can be used to compute the oscillations of the vocal folds and their phase-space trajectories.
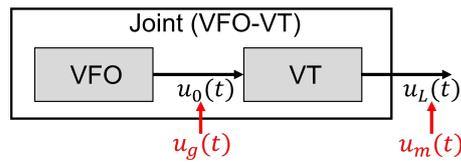
Figure 4. The VFO model models the generation of glottal signals by the movements of the vocal folds. The VT model models the transformation of the glottal signal generated by the vocal folds to the final voice signal. The joint VFO-VT model combines the two, using the output of the VFO model as the input to the VT model. ADLES compares the glottal signal $u_0(t)$ generated by the VFO model to a reference glottal signal $u_g(t)$ to estimate VFO parameters. ADLES-VFT compares the output of the joint model, $u_L(t)$, to a reference signal $u_m(t)$ obtained from an actual voice recording, to estimate both VFO and VT parameters. The output of the VFO model is the desired vocal-fold oscillation.

4.1. Estimating VFO Parameters from the Excitation Signal: The Backward Approach (ADLES)

As the first step, we describe the backward problem: how to derive the VFO model parameters that best explain the glottal excitation for a given phonated signal. We use the approach proposed in [31]—we estimate the VFO model parameters to minimize the error between the volume velocity of air predicted by the model and a reference signal representing the actual glottal excitation to the vocal tract. If the VFO model were to be considered in isolation, this reference could be obtained through actual physical measurements, e.g., through photography [5], physical or numerical simulations [7,17], or by inverse filtering the speech signal using a technique such as [35] (the approach used in [31]).

For the purpose of our discussion in this section, however, we do not specify where this reference excitation is obtained from, since the estimation of VFO model parameters from a given glottal excitation is only a waypoint towards estimation from the joint model that includes both the VFO and VT components. As we will see in Section 4.2, this does not in fact require explicit knowledge of the reference signal at the glottis.

Let $u_g(t)$ be the reference signal representing the *true* air-volume velocity at the glottis that excites the vocal tract. The volume velocity of air $u_0(t)$ (we remind the reader that $u_0(t) = u(0, t)$) at the glottis can also be computed from the vocal fold opening at the glottis as

$$u_0(t) = \tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t)) \tag{6}$$

where ξ_0 is the half glottal width at rest, $2\xi_0 + \xi_l(t) + \xi_r(t)$ is the complete glottal opening, d is the length of the vocal fold, and $\tilde{c}$ is the air particle velocity at the midpoint of the vocal folds.

We assume that the movement of the vocal folds follows the VFO model of Section 3.1. Correspondingly, $\xi_l(t)$ and $\xi_r(t)$ must obey Equation (3), subject to boundary conditions. The model parameters α, β and Δ can hence be computed to minimize the difference between the air volume velocity $u_0(t)$ predicted by the model and the reference $u_0^g(t)$.

We define the instantaneous *residual R* as the error between $u_0(t)$ and $u_g(t)$:

$$\begin{aligned} R(t) &= u_0(t) - u_g(t) \\ &= \tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t)) - u_g(t) \end{aligned} \tag{7}$$

The overall ℓ_2 error between $u_0(t)$ and $u_g(t)$ is given by the integral

$$F(\xi_l, \xi_r; \vartheta) = \int_0^T R^2(t)dt \tag{8}$$

where $\vartheta = [\alpha, \beta, \Delta]$ represents the parameters of the VFO model, and T represents the complete length of the reference signal.

The actual estimation can now be stated as

$$\alpha^*, \beta^*, \Delta^* = \arg\min_{\alpha,\beta,\Delta} F(\xi_l, \xi_r \vartheta) \tag{9}$$

subject to

$$\ddot{\xi}_l + \beta(1 + \xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{10}$$

$$\ddot{\xi}_r + \beta(1 + \xi_r^2)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{11}$$

$$\xi_r(0) = C_r \tag{12}$$
$$\xi_l(0) = C_l \tag{13}$$
$$\dot{\xi}_r(0) = 0 \tag{14}$$
$$\dot{\xi}_l(0) = 0 \tag{15}$$

where C_r and C_l are constants representing the quiescent positions of the vocal folds, and the folds are assumed to be at rest prior to the onset of phonation. For the computation we set ξ_0 to a typical value of 0.1 cm. The length of the vocal folds d may be set to 17.5 mm (which is within the range of normal lengths for both male and female subjects), and the air particle velocity $\tilde{c}$ to 5000 cm/s [2].

Note that given α, β and Δ the differential equations of model Equations (11)–(15) (the constraints) can be solved by any ODE solver to obtain $\xi_l(t)$ and $\xi_r(t)$. So, in principle, we could solve the constrained optimization problem of Equation (9)–(15) by a grid search over (α, β, Δ) to identify the specific values that minimize the squared error of Equation (8). This would, however, be highly inefficient.

Instead we propose the ADLES ("ADjoint LEast Squares") algorithm, which restates the constraints (11)–(15) as Lagrangians on the objective, and derives a gradient descent solution. The detailed derivation of ADLES is given in Appendix C. We summarize the key steps below.

Incorporating constraints (11)–(15) into the objective, we define the Lagrangian:

$$
\begin{aligned}
\mathcal{L}(\vartheta) = \\
\int_0^T \Bigg[R^2 + \lambda\left(\ddot{\xi}_r + \beta\left(1 + \xi_r^2\right)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r - \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right)\right) \\
+ \eta\left(\ddot{\xi}_l + \beta\left(1 + \xi_l^2\right)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l - \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right)\right) \Bigg] dt \\
+ \mu_l(\xi_l(0) - C_l) + \mu_r(\xi_r(0) - C_r) + \nu_l\dot{\xi}_l(0) + \nu_r\dot{\xi}_r(0)
\end{aligned}
\tag{16}
$$

where λ, η, μ_l, μ_r, ν_l and ν_r are Lagrangian multipliers. Note that λ and η are also functions of time (we have not explicitly shown the "(t)" above for brevity of notation).

We obtain the Lagrangian parameters λ and η as the solution to the following equations: For $0 < t < T$:

$$\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2}\right)\lambda + 2\tilde{c}dR = 0 \tag{17}$$

$$\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2}\right)\eta + 2\tilde{c}dR = 0 \tag{18}$$

$$\beta\left(1 + \xi_r^2\right)\lambda - \alpha(\lambda + \eta) = 0 \tag{19}$$

$$\beta\left(1 + \xi_l^2\right)\eta - \alpha(\lambda + \eta) = 0 \tag{20}$$

with initial conditions at $t = T$:

$$\lambda(T) = 0 \tag{21}$$

$$\dot{\lambda}(T) = 0 \tag{22}$$

$$\eta(T) = 0 \tag{23}$$

$$\dot{\eta}(T) = 0 \tag{24}$$

Note that given α, β, Δ, ξ_r and ξ_l, Equations (17)–(24) represent a differential-algebraic system of equations and can be solved by any DAE solver to obtain λ and η.

Given ξ_r, ξ_l, λ and η, the derivatives of $F(\xi_l, \xi_r, \vartheta)$ w.r.t. α, β and Δ can now be obtained as:

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

$$F_\beta = \int_0^T \left(\left(1 + \xi_r^2\right)\dot{\xi}_r \lambda + \left(1 + \xi_l^2\right)\dot{\xi}_l \eta \right)dt \tag{25}$$

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

The derivatives from Equation (25) are plugged into a gradient descent update rule for the model parameters:

$$\alpha \longleftarrow \alpha - \tau^\alpha F_\alpha$$

$$\beta \longleftarrow \beta - \tau^\beta F_\beta \tag{26}$$

$$\Delta \longleftarrow \Delta - \tau^\Delta F_\Delta$$

where $\tau^{\cdot}$ is the step-size.

The overall ADLES algorithm is summarized in Algorithm 1:

Algorithm 1 ADLES algorithm

1: Initialize α, β and Δ
2: **while** F not converged **do**   ▷ *Iterate until the error converges*
3:      Solve (11)–(10) with initial conditions (12)–(15), using the current estimates of α, β and Δ, obtaining ξ_r, ξ_l, $\dot{\xi}_r$ and $\dot{\xi}_l$.
4:      Solve (17)–(20) with the initial conditions (21)–(24), using the current values of ξ_l, ξ_r, α, β and Δ, obtaining λ, $\dot{\lambda}$, η and $\dot{\eta}$.
5:      Compute F_α, F_β and F_Δ from Equation (25).
6:      Update α, β and Δ with (26).
7: **end while**

4.2. Estimating VFO Parameters from the Speech Signal: The Forward-Backward Approach (ADLES-VFT)

The backward approach, solved by the ADLES algorithm in Section 4.1, derives the VFO parameters by minimizing the error between the output of the VFO model $u_0(t)$ and the glottal excitation $u_g(t)$. However, in general, $u_g(t)$ is not available, and this error cannot actually be computed.

Instead, in the forward-backward approach, we further *propagate* the generated excitation $u_0(t)$ through the vocal tract, represented by the VT model of Equation (4) to obtain a signal $u_L(t) = u(L, t)$ at the lips. This is the output of the joint VFO and VT models. We compute the error between the generated signal $u_L(t)$ and the air velocity measurement derived from the recorded voice signal, which *is* available, and propagate this error backward through the vocal tract, to obtain the error at the glottis. The VT and VFO model parameters are estimated to minimize this error. Thus, the algorithm itself proceeds through iterations of two steps: a forward step in which the VFO-generated excitation is propagated through the VT model to generate an output signal, and a backward step in which the error between

the generated signal and the recorded speech is propagated backward through the VT to adjust the model parameters. We explain the entire procedure below.

The recorded voice signal is, in fact, a pressure wave and records the pressure wave emitted at the lips. Let $p_m(t)$ be the measured acoustic pressure at the lip. The corresponding volume velocity is given by [34]

$$u_m(t) = \frac{A(L)}{\rho c} p_m(t) \tag{27}$$

$u_m(t)$ is now our reference signal at the lips to which $u_L(t)$ must be matched, in order to estimate model parameters.

The propagation of $u_0(t) = u(0, t)$ through the vocal tract is assumed to follow the dynamics of Equation (4). Let $\mathcal{H}_f$ be the nonlinear operator representing acoustic wave propagation through the vocal tract from the glottis to the lip. The subscript f in $\mathcal{H}_f$ represents the vocal-tract profile $f(x, t)$ in Equation (4) and indicates the dependence of $\mathcal{H}_f$ on $f(x, t)$. Thus the vocal-tract output $u_L(t)$ is given by

$$u_L(t) = \mathcal{H}_f(u_0(t)) = \mathcal{H}_f(\tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t))) \tag{28}$$

Our objective is to minimize the difference between the measured volume velocity $u_m(t)$ and the predicted volume velocity $u_L(t)$ near the lip subject to constraint that the dynamics of the vocal folds must follow the VFO model of Equation (3). Note that the parameters of the joint model include the VFO model parameters α, β and Δ, and the vocal tract profile $f(x, t)$ required by the VT model. Although we only require the VFO model parameters to determine vocal fold oscillation, the minimization must be performed against all of these. Thus the estimation problem becomes

$$\min \int_0^T (u_L(t) - u_m(t))^2 dt$$
$$\Leftrightarrow \quad \min \int_0^T \left(u_L(t) - \frac{A(L)}{\rho c} p_m(t) \right)^2 dt \tag{29}$$

subject to

$$\ddot{\xi}_r + \beta \left(1 + \xi_r^2 \right) \dot{\xi}_r + \xi_r - \frac{\Delta}{2} \xi_r = \alpha \left(\dot{\xi}_r + \dot{\xi}_l \right) \tag{30}$$

$$\ddot{\xi}_l + \beta \left(1 + \xi_l^2 \right) \dot{\xi}_l + \xi_l + \frac{\Delta}{2} \xi_l = \alpha \left(\dot{\xi}_r + \dot{\xi}_l \right) \tag{31}$$

$$\text{(I.C.1)} \ \xi_r(0) = C_r \tag{32}$$
$$\text{(I.C.2)} \ \xi_l(0) = C_l \tag{33}$$
$$\text{(I.C.3)} \ \dot{\xi}_r(0) = 0 \tag{34}$$
$$\text{(I.C.4)} \ \dot{\xi}_l(0) = 0 \tag{35}$$

where, as before, (30) and (31) represent the asymmetric vocal folds displacement model (3), I.C. stands for initial condition, and Cs are constants. The minimization is performed against the complete set of parameters of the joint VFO-VT model, i.e., α, β, Δ and $f(x, t)$.

Unlike in Equations (11)–(15), this cannot be solved, even in principle, by simply scanning for the optimal α, β and Δ, since $\mathcal{H}_f$ is characterized by $f(x, t)$ which is also unknown and must be determined.

To solve the optimization problem of (29)–(35), we derive an efficient gradient-descent solution which we term the ADLES-VFT algorithm. The essential steps of the solution are given below. The details of the derivation are in Appendix D.

4.2.1. Forward Pass

First, note that, as before, the constraint Equations (30)–(35) are ordinary differential equations with initial conditions that, given α, β and Δ, can be solved by any ODE solver. The solution will give us the VFO model generated glottal excitation $u_0(t)$.

Next, we propagate the generated excitation $u_0(t)$ through the VT model. For this, we must solve

$$\frac{\partial^2 u(x,t)}{\partial t^2} = c^2 \frac{\partial^2 u(x,t)}{\partial x^2} + f(x,t) \tag{36}$$

subject to

$$\text{(B.C.1)} \qquad u(0,t) = u_0(t) \tag{37}$$

$$\text{(B.C.2)} \qquad \frac{\partial u}{\partial n_\Gamma} = 0 \tag{38}$$

$$\text{(I.C.1)} \qquad u(x,0) = 0 \tag{39}$$

$$\text{(I.C.2)} \qquad \frac{\partial u(x,0)}{\partial t} = 0 \tag{40}$$

where B.C. stands for boundary condition, and I.C. stands for initial condition. The vocal tract is assumed to be circular at the glottis and the lips. Here, n_Γ is the outward unit normal to the vocal tract boundary Γ, at the glottis.

Equations (36)–(40) represent a set of partial differential equations. The boundary conditions relate to the air volume velocity at the glottal end of the vocal tract. The initial conditions relate to air volume velocity at the initial time, $t = 0$, when the generated glottal signal $u_0(t)$ enters the vocal tract. Given $u_0(t)$ and $f(x,t)$ (36)–(40) can be solved using a PDE solver. In our work we use the finite-element method described in Appendix E.

Solving (36)–(40) gives us $u(x,t)$ at all positions $x \in (0,L)$ and time $t \in (0,T)$. In the process, it also gives us $\mathcal{H}_f(u_0(t)) = u_L(t) = u(L,t)$.

4.2.2. Backward Pass

The backward pass updates all model parameters including the VT term $f(x,t)$, and VFO parameters based on the error at the output.

Updating f:

We denote the estimation residual as:

$$r(t) = u_m(t) - \mathcal{H}_f(u_0(t)) \tag{41}$$

We must propagate this residual backward through the VT model. To do so, we use a time reversal technique [36] and backpropagate the difference (41) into the vocal tract, which gives:

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2} + f(x,t) \tag{42}$$

subject to

$$\text{(B.C.1)} \qquad \frac{\partial z}{\partial n_\Gamma} = r \tag{43}$$

$$\text{(I.C.1)} \qquad z(x,T) = 0 \tag{44}$$

$$\text{(I.C.2)} \qquad \frac{\partial z(x,T)}{\partial t} = 0 \tag{45}$$

where z is the time reversal of u. Note that the boundary conditions and initial conditions in (42)–(45) are now defined at the lip, and the equation itself propagates backward through the vocal tract.

As before, Equations (42)–(45) can be solved by the finite-element method of Appendix E to give us $z(x, t)$. The gradient update rule for $f(x, t)$ is then obtained as

$$f \longleftarrow f + \iota\left(\frac{z}{c^2} + u\right).\tag{46}$$

where ι is a learning rate parameter (see Appendix D).

Updating α, β, Δ:

As in the case of the backward approach of Section 4.1 we define a residual

$$R(t) = u_L(t) - u_m(t) = \mathcal{H}_f(\tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t))) - \frac{A(L)}{\rho c}p_m(t)\tag{47}$$

Note that unlike in Section 4.1 the residual in Equation (47) is defined at the lips, rather than at the glottis. As before, we can define the total squared residual error as

$$F(\xi_l, \xi_r; \vartheta) = \int_0^T R^2(t)dt\tag{48}$$

where $\vartheta = [\alpha, \beta, \Delta]$ are the parameters of the vocal folds model (3). F must be minimized with respect to ϑ, subject to the constraints imposed by the VFO model.

Once again, as in Section 4.1 we fold in the constraints into the objective through Lagrangian multipliers as

$$
\begin{aligned}
\mathcal{L}(\vartheta) = \int_0^T &\left[R^2 + \lambda\left(\ddot{\xi}_r + \beta\left(1 + \xi_r^2\right)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r - \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right)\right)\right.\\
&\left.+\eta\left(\ddot{\xi}_l + \beta\left(1 + \xi_l^2\right)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l - \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right)\right)\right]dt\\
&+ \mu_l(\xi_l(0) - C_l) + \mu_r(\xi_r(0) - C_r) + \nu_l\dot{\xi}_l(0) + \nu_r\dot{\xi}_r(0)
\end{aligned}\tag{49}
$$

where λ, η, μ and ν are multipliers, and λ and η are themselves functions of time. Optimization requires minimization of $\mathcal{L}(\vartheta)$ with respect to α, β and Δ.

This leads us (Appendix D) to the following set of equations for λ and η:

$$\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2}\right)\lambda + 2\tilde{c}dR\frac{u_L}{u_0} = 0\tag{50}$$

$$\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2}\right)\eta + 2\tilde{c}dR\frac{u_L}{u_0} = 0\tag{51}$$

$$\beta\left(1 + \xi_r^2\right)\lambda - \alpha(\lambda + \eta) = 0\tag{52}$$

$$\beta\left(1 + \xi_l^2\right)\eta - \alpha(\lambda + \eta) = 0\tag{53}$$

with initial conditions (at $t = T$, i.e., at the lips):

$$\lambda(T) = 0\tag{54}$$

$$\dot{\lambda}(T) = 0\tag{55}$$

$$\eta(T) = 0\tag{56}$$

$$\dot{\eta}(T) = 0\tag{57}$$

Given $R(t)$, $u_0(t)$ and $u_L(t)$ Equations (50)–(57) represent a set of differential-algebraic equations and can be solved with a DAE solver.

We finally obtain the derivative of F w.r.t. α, β and Δ (represented below as F_α, F_β and F_Δ, respectively) as

$$F_\alpha = \int_0^T -(\dot{\xi}_r + \dot{\xi}_l)(\lambda + \eta)dt \tag{58}$$

$$F_\beta = \int_0^T \left(\left(1 + \xi_r^2\right)\dot{\xi}_r\lambda + \left(1 + \xi_l^2\right)\dot{\xi}_l\eta \right)dt \tag{59}$$

$$F_\Delta = \int_0^T \frac{1}{2}(\xi_l\eta - \xi_r\lambda)dt \tag{60}$$

The gradient descent update rules for the VFO model parameters are finally obtained as

$$\alpha \longleftarrow \alpha - \tau^\alpha F_\alpha \tag{61}$$

$$\beta \longleftarrow \beta - \tau^\beta F_\beta \tag{62}$$

$$\Delta \longleftarrow \Delta - \tau^\Delta F_\Delta \tag{63}$$

4.2.3. The ADLES-VFT Algorithm Summarized

The overall ADLES-VFT algorithm for solving the parameter estimation problem (29)–(35) is summarized in Algorithm 2.

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.

We note here that Algorithm 2 requires several terms to be specified. In our implementation, the quiescent positions of the vocal folds, C_r and C_l were set to 0. We initialize $[\alpha, \beta, \Delta] = [0.8, 0.32, 1.0]$—these values were empirically found to work best. $f(x, t)$ is initialized to 0. This effectively initializes Equation (4) to the Webster Horn equation. The step sizes τ^α, τ^β and τ^Δ are all adaptively set to $0.01/max(F_\alpha, F_\beta, F_\Delta)$, and ι is set to 1. The actual objective minimize, Equation (29), requires scaling $p_m(t)$ by $A(L)/\rho c$ prior to comparison to $u_L(t)$. In practice, since $p_m(t)$ is derived from a voice signal recorded at a distance from the lips, the unknown transmission loss between the lips and the microphone must also be considered. To deal with this, we simply normalize both sequences to 0 mean and unit variance, and do not apply any additional scaling.

Algorithm 2 ADLES-VFT algorithm

1: Initialize α, β, Δ and $f(x, t)$.
2: **while** F not converged **do**   ▷ *Iterate until the error converges*
3:     Solve (30)–(31) with initial conditions (32)–(35) with an ODE solver, using the current estimates of α, β and Δ, obtaining ξ_r, ξ_l, $\dot{\xi}_r$ and $\dot{\xi}_l$ and $u_0(t)$.
4:     Using current estimate of $f(x, t)$ and $u_0(t)$, solve the forward propagation model (36)–(40) for $u_L(t)$ with a PDE solver, e.g., the finite-element method of Appendix E.
5:     Calculate the estimation difference $r(t)$ using (41).
6:     Using the current estimate of $f(x, t)$ and $r(t)$, solve the backward propagation model (42)–(45) for $z(x, t)$ with a PDE solver (Appendix E).
7:     Update $f(x, t)$ using (46).
8:     Solve (50)–(53) with initial conditions (54)–(57) using a DAE solver to obtain λ, $\dot{\lambda}$, η and $\dot{\eta}$.
9:     Compute (58)–(60) through numerical integration to obtain derivatives F_α, F_β and F_Δ from
10:     Update α, β and Δ with (63).
11: **end while**

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

5. Experimental Results and Interpretation

Unfortunately, until the time of writing this paper, we could not obtain actual electroglottographic measurements or video data of vocal fold motion to compare our derived VFOs to. However, from a computational perspective the algorithms proposed can still be validated in different ways. We explain these below.

5.1. Validation 1

Our first validation approach is to use the proxy of showing that the solutions obtained are indeed discriminative of fine-level changes in glottal flow dynamics of the phonation process.

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. We note that 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.

To interpret these in discriminative settings, we can utilize some well-established methods for characterizing dynamical systems, borrowing them from chaos theory and other areas of applied mathematics (e.g., flow, orbit, attractor, stability, Poincaré map, bifurcation, Lyapunov exponents, etc.). These are described in Appendix F.

Interpreting a System's Phase Portraits Using Its Bifurcation Map

Appendix F describes the concepts and tools used to study the behaviors (e.g., flow, orbit, attractor, stability, Poincaré map, bifurcation) of nonlinear dynamical systems such as Equation (3). The phase space of the system in Equation (3) (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 [24,37]. 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 phases θ_r, θ_l satisfy $|n\theta_r - m\theta_l| < C$ where n, m are integers, and C is a constant [24]. Such entrainment can be revealed by the Poincaré map, where the number of trajectory crossings of the right or left oscillators within the Poincaré section indicates 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 [15,37].

As we will see later (and as indicated in Figure 5), model parameters can characterize voice pathologies, and these can be visually evident in in phase portraits and bifurcation plots.

We use the backward 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 [38]. It comprises 200 recordings of the sustained vowel /aː/. 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.
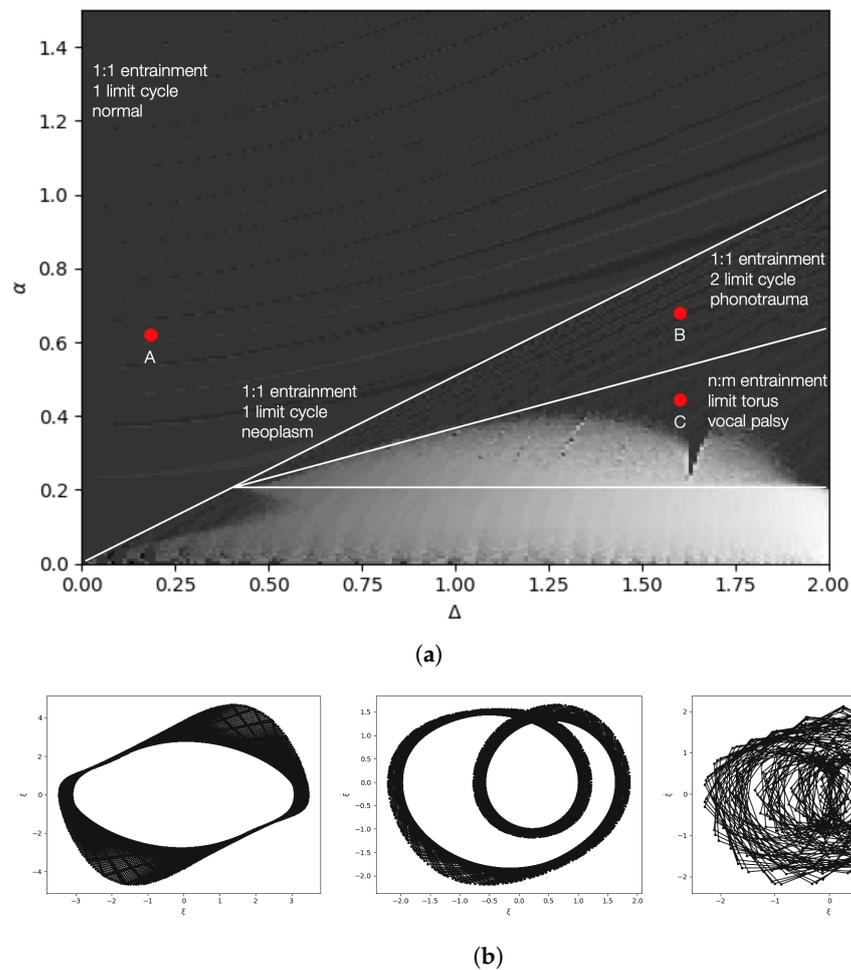
(**a**)

(**b**)

Figure 5. (**a**) A 3D Bifurcation diagram of the asymmetric vocal fold model. The third dimension is perpendicular to the parameter plane shown, and depicts 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 $\dot{\zeta}_{r,l} = 0$ at stable status. This is consistent with the theoretical results in [24]); (**b**) Phase-space trajectories (or phase portraits) corresponding to the 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.

Figure 6 shows some phase portraits showing the coupling 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.
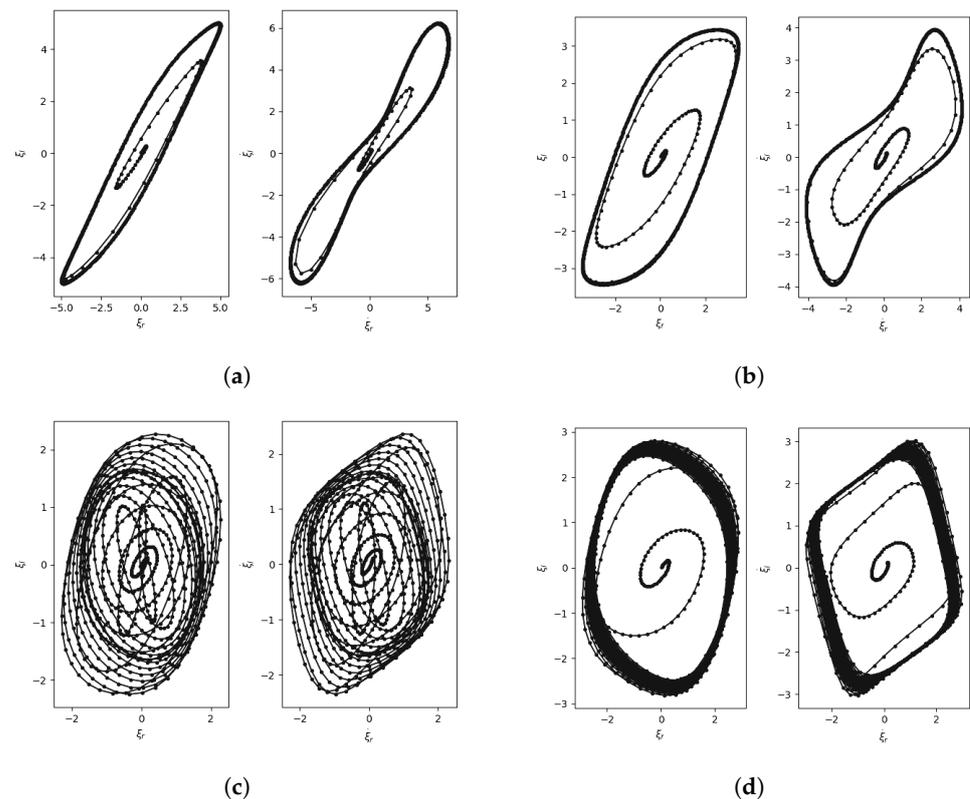
(a)

(b)

(c)

(d)

Figure 6. Phase portraits showing the coupling of the 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.

Table 1. Parameters obtained and pathologies identified through ADLES.

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

5.2. Validation 2

Our second validation approach is to compare the excitation signal obtained through inverse filtering with the glottal flow signal (VFO) obtained through the backward or forward-backward algorithm. The rationale behind this is that within reasonable bounds of error, the glottal flow signal obtained through our model is expected to conform to the oscillation patterns seen in the excitation signal for each speaker.

Figure 7 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 does achieve its objectives in individualizing the asymmetric model to each speaker instance.
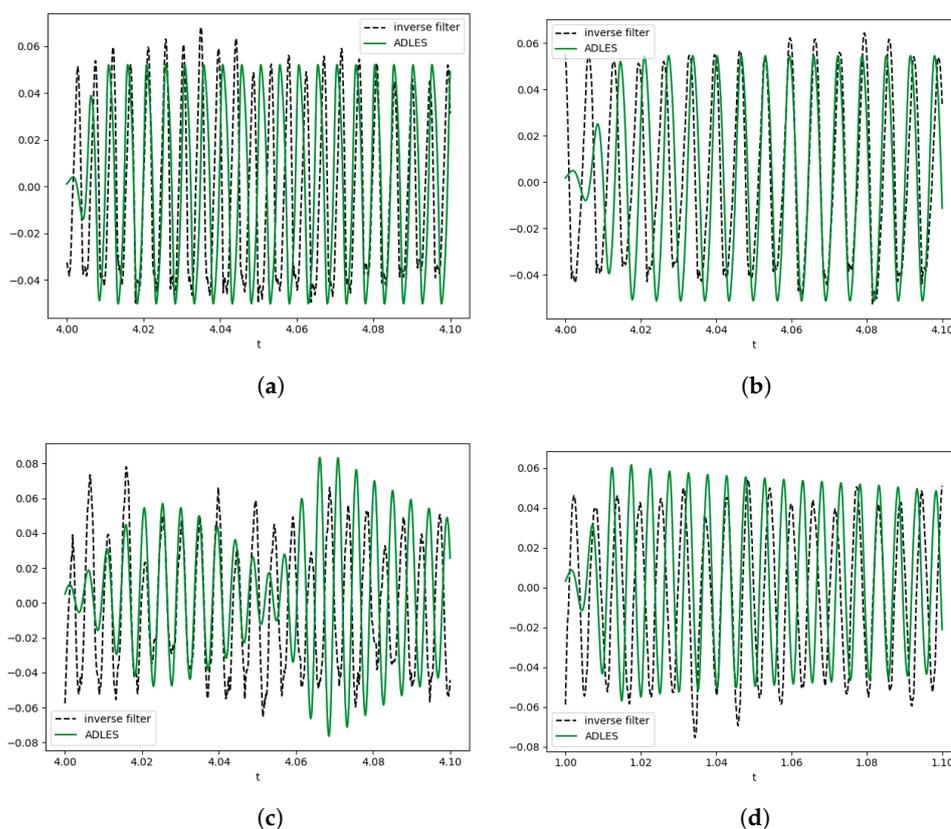
Figure 7. Glottal flows from inverse filtering and ADLES estimation for (**a**) normal speech (control), (**b**) neoplasm, (**c**) phonotrauma, and (**d**) vocal palsy.

5.3. Validation 3

Our third validation approach is to compare the estimation precision of the 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) and forward-backward (ADLES-VFT) algorithms. The glottal flows obtained by inverse filtering the speech signals are treated as ground truths. Since there is no ground truth for model parameters, we treat the parameters obtained by ADLES as ground truth. These results suggest that our forward-backward algorithm can effectively recover the vocal tract profile, glottal flow, and model parameters.

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

	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

5.4. Validation 4

Our fourth validation comes indirectly from prior studies. 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 Appendix F.

Some of these features have been used in real-world applications and proven to be effective. For example, in [39], 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 [40] 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) [41].

6. 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. Extensions of these approaches can use other physical models of voice production, and other physical processes including phonation.

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.

The phase space characterization presented in this paper is based on phase space trajectories (a topological perspective)—the left and right vocal fold oscillations, velocities or accelerations. Measurements can also be performed from statistical, signal processing, information-theoretic and other perspectives. 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 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. We delegate these explorations to future work.

Appendix A

The following is a brief description of various VFO models:

- **Two-mass models**:
  Two-mass models describe vocal fold motion as two coupled mass-damper-spring oscillators:

$$M_1 \ddot{x}_1 + B_1 \dot{x}_1 + K(x_1 - x_2) + R_1 = F_1 \tag{A1}$$

$$M_2 \ddot{x}_2 + B_2 \dot{x}_2 + K(x_2 - x_1) + R_2 = F_2 \tag{A2}$$

  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 R_i is the elastic restoring force [15]. 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.

- **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 on mass-spring-damper motion dynamics which are widely used in multiple problem settings (e.g., [42]). For the i-th mass component, the equation of motion is:

$$M_i \ddot{x}_i = F_i^A + F_i^V + F_i^L + F_i^C + F_i^D \tag{A3}$$

  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 [18]. In [18], 50 masses are used.

- **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 [17].

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

$$\sigma = S\epsilon \tag{A4}$$

where σ is the stress tensor, ϵ is the strain tensor, and S is the stiffness matrix consisting of Young's modulus, shear modulus, and Poisson's ratio [17]. The relation between stress and displacement is governed by:

$$\sigma_x = C_1 \mu \frac{\partial u}{\partial x} + C_2 \mu \frac{\partial w}{\partial z} \tag{A5}$$

$$\sigma_z = C_2 \mu \frac{\partial u}{\partial x} + C_1 \mu \frac{\partial w}{\partial z} \tag{A6}$$

$$\tau_{xy} = \mu' \frac{\partial u}{\partial y} \tag{A7}$$

$$\tau_{yz} = \mu' \frac{\partial w}{\partial y} \tag{A8}$$

$$\tau_{zx} = \mu \left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) \tag{A9}$$

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 [17]. This system of partial differential equations can be efficiently solved by finite element methods.

The following is a brief description of various VT models:

- **Bio-mechanical 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, bio-mechanical, and visco-elastic properties of muscles [27]. They are more scalable and accurate, and allow for the modeling of more fine-grained control over muscular forces, articulator positions, and movements.
- **Statistical models**:
  These 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 a combination of, components and vocal tract parameters [25].
- **Geometric models**:
  These 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. [26]. 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.

Appendix B

Appendix B.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 [43]. A vector field is a smooth section on TM, $\Gamma(TM) \ni X : M \to TM$. It

associates each point $p \in M$ with a tangent vector $\bar{v}(p) := X|_p : C^\infty(M) \overset{\sim}{\to} \mathbb{R}$ [43]. Let $\gamma(t) : \mathbb{R} \supseteq I \to M$ be a maximal integral curve [43] through p at t_0, which is a solution to:

$$\begin{aligned} \gamma'(t) &= X(\gamma(t)) \\ \gamma(t_0) &= p \end{aligned} \tag{A10}$$

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

The planar motion of the pressure wave in the vocal tract is governed by the equations [36]:

$$\frac{1}{\rho c^2} \frac{\partial \hat{p}}{\partial t} + \text{div } v = 0 \tag{A11}$$

$$\rho \frac{\partial v}{\partial t} + \text{grad } \hat{p} = 0 \tag{A12}$$

where $\hat{p}(p,t)$ 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 (A11) describes the conservation of mass, and (A12) describes the conservation of momentum [36].

For notational convenience, we use cylindrical coordinates $p = (r, \theta, x)$, where the 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:

$$A(x) = \int_0^{2\pi} d\theta \int_0^{R(\theta,x)} r dr \tag{A13}$$

The average acoustic pressure is:

$$p(x,t) = \frac{1}{A(x)} \int_0^{2\pi} d\theta \int_0^{R(\theta,x)} \hat{p} r dr \tag{A14}$$

and the volume velocity is:

$$u(x,t) = \int_0^{2\pi} d\theta \int_0^{R(\theta,x)} v_x r dr \tag{A15}$$

where v_x is the x component of v. Integrating (A11) over the volume of vocal tract bounded by cross sections at x_0 and x gives:

$$0 = \int_M \frac{1}{\rho c^2} \frac{\partial \hat{p}}{\partial t} + \text{div } v \tag{A16}$$

$$= \int_{x_0}^x \left[\int_0^{2\pi} d\theta \int_0^R \frac{1}{\rho c^2} \frac{\partial \hat{p}}{\partial t} r dr \right] dx' + \int_M \text{div } v \tag{A17}$$

$$= \frac{1}{\rho c^2} \int_{x_0}^x A(x') \frac{\partial p(x',t)}{\partial t} dx' + \iint_\Sigma n_v d\sigma + u(x,t) - u(x_0,t) \tag{A18}$$

where we substitute Equations (A17) and (A18) into (A14) and (A15), and apply Stokes' theorem [28,36]; n_v is the component of v normal and outward to the inner surface Σ.

The element of area $d\sigma$ is given by [28,36]:

$$d\sigma = S(\theta, x)d\theta dx \tag{A19}$$

where $S d\theta dx$ is a top 2-form on Σ [43]. Substituting (A19) into (A18) and differentiating w.r.t. x yields:

$$\frac{A(x)}{\rho c^2} \frac{\partial p}{\partial t} + \frac{\partial u}{\partial x} + \int_0^{2\pi} n_v(\theta, x, t) S(\theta, x) d\theta = 0 \tag{A20}$$

Following similar steps, integrating the x component of (A12) over the cross section at x yields:

$$\rho\frac{\partial u}{\partial t} + A(x)\frac{\partial p}{\partial x} + \int_0^{2\pi}(p(x,t) - p_w(\theta, x, t))\frac{\partial}{\partial x}\left(\frac{1}{2}R^2\right)d\theta = 0 \tag{A21}$$

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

Appendix B.2. The Integrated Vocal Tract Model

To simplify our problem, we combine the wave Equations (A20) and (A21) into a single vocal tract model. Differentiating (A20) w.r.t x and (A21) w.r.t. t, and canceling out the pressure term gives:

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

$$+ c^2\partial_x\int_0^{2\pi}n_v(\theta, x, t)S(\theta, x)d\theta \tag{A22}$$

$$= c^2\frac{\partial^2 u}{\partial x^2} + f(x,t) \tag{A23}$$

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.

Appendix C

In this appendix, we derive the ADLES algorithm to estimate VFO model parameters from a given glottal excitation signal.

Let $u_g(t)$ be the actual air volume velocity at the glottis. We can can also derive $u_0(t)$ from the displacement of the vocal folds as

$$u_0(t) = \tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t)) \tag{A24}$$

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

We assume the movements of the vocal folds to follow the dynamics specified by Equation (3) in Section 3.1. Our objective is then to estimate the parameters of the model to minimize the difference:

$$\min\int_0^T\left(u_0(t) - u_g(t)\right)^2 dt \Leftrightarrow \tag{A25}$$

$$\min\int_0^T\left(\tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t)) - u_g(t)\right)^2 dt \tag{A26}$$

such that the movements of the vocal folds follow the VFO model:

$$\ddot{\xi}_r + \beta(1 + \xi_r^2)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{A27}$$

$$\ddot{\xi}_l + \beta(1 + \xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l = \alpha(\dot{\xi}_r + \dot{\xi}_l) \tag{A28}$$

$$\xi_r(0) = C_r \tag{A29}$$

$$\xi_l(0) = C_l \tag{A30}$$

$$\dot{\xi}_r(0) = 0 \tag{A31}$$

$$\dot{\xi}_l(0) = 0 \tag{A32}$$

where C_r and C_l are constants representing the quiescent positions of the vocal folds.

The Adjoint Least Squares (ADLES) Solution

To solve the functional least squares in (A26), we require the gradients of (A26) 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

$$R = \tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t)) - u_g(t) \tag{A33}$$

We have

$$f(\xi_l, \xi_r; \vartheta) = R^2 \tag{A34}$$

and

$$F(\xi_l, \xi_r; \vartheta) = \int_0^T f(\xi_l, \xi_r; \vartheta)dt \tag{A35}$$

where $\vartheta = [\alpha, \beta, \Delta]$.

Incorporating the constraints into the objective using Lagrangian multipliers, we define the Lagrangian:

$$
\begin{aligned}
\mathcal{L}(\vartheta) = & \\
\int_0^T & \left[f + \lambda \left(\ddot{\xi}_r + \beta\left(1 + \xi_r^2\right)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r - \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right) \right) \right. \\
& \left. + \eta \left(\ddot{\xi}_l + \beta\left(1 + \xi_l^2\right)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l - \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right) \right) \right] dt \\
& + \mu_l(\xi_l(0) - C_l) + \mu_r(\xi_r(0) - C_r) + \nu_l\dot{\xi}_l(0) + \nu_r\dot{\xi}_r(0)
\end{aligned}
\tag{A36}
$$

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

$$
\begin{aligned}
\mathcal{L}_\alpha = \int_0^T & \left[2\tilde{c}dR(\partial_\alpha\xi_l + \partial_\alpha\xi_r) \right. \\
& + \lambda \left(\partial_\alpha\ddot{\xi}_r + 2\beta\dot{\xi}_r\xi_r\partial_\alpha\xi_r + \beta\left(1 + \xi_r^2\right)\partial_\alpha\dot{\xi}_r \right. \\
& \left. + \partial_\alpha\xi_r - \frac{\Delta}{2}\partial_\alpha\xi_r - \alpha\left(\partial_\alpha\dot{\xi}_r + \partial_\alpha\dot{\xi}_l\right) - \left(\dot{\xi}_r + \dot{\xi}_l\right) \right) \\
& + \eta \left(\partial_\alpha\ddot{\xi}_l + 2\beta\dot{\xi}_l\xi_l\partial_\alpha\xi_l + \beta\left(1 + \xi_l^2\right)\partial_\alpha\dot{\xi}_l \right. \\
& \left. \left. + \partial_\alpha\xi_l + \frac{\Delta}{2}\partial_\alpha\xi_l - \alpha\left(\partial_\alpha\dot{\xi}_r + \partial_\alpha\dot{\xi}_l\right) - \left(\dot{\xi}_r + \dot{\xi}_l\right) \right) \right] dt \\
& + \mu_l\partial_\alpha\xi_l(0) + \mu_r\partial_\alpha\xi_r(0) + \nu_l\partial_\alpha\dot{\xi}_l(0) + \nu_r\partial_\alpha\dot{\xi}_r(0)
\end{aligned}
\tag{A37}
$$

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

$$\int_0^T \lambda\partial_\alpha\ddot{\xi}_r dt = \int_0^T \partial_\alpha\xi_r\ddot{\lambda}dt - \partial_\alpha\xi_r\dot{\lambda}\Big|_0^T + \partial_\alpha\dot{\xi}_r\lambda\Big|_0^T \tag{A38}$$

Applying the same to $\eta\partial_\alpha\ddot{\xi}_l$, substituting into (A37) and simplifying the final expression we obtain:

$$
\begin{aligned}
\mathcal{L}_\alpha = \int_0^T &\left[\left(\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2} \right)\lambda + 2\tilde{c}dR \right)\partial_\alpha\xi_r \right.\\
&+ \left(\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2} \right)\lambda + 2\tilde{c}dR \right)\partial_\alpha\xi_l \\
&+ \left(\beta\left(1+\xi_r^2\right)\lambda - \alpha(\lambda+\eta) \right)\partial_\alpha\dot{\xi}_r \\
&+ \left(\left(\beta\left(1+\xi_l^2\right)\eta - \alpha(\lambda+\eta) \right)\partial_\alpha\dot{\xi}_l \right.\\
&\left. - (\dot{\xi}_r + \dot{\xi}_l)(\lambda+\eta) \right]dt \\
&+ (\mu_r+\dot{\lambda})\partial_\alpha\xi_r(0) - \dot{\lambda}\partial_\alpha\xi_r(T) \\
&+ (\nu_r-\lambda)\partial_\alpha\dot{\xi}_r(0) + \lambda\partial_\alpha\dot{\xi}_r(T) \\
&+ (\mu_l+\dot{\eta})\partial_\alpha\xi_l(0) - \dot{\eta}\partial_\alpha\xi_l(T) \\
&+ (\nu_l-\eta)\partial_\alpha\dot{\xi}_l(0) + \eta\partial_\alpha\dot{\xi}_l(T)
\end{aligned}
\tag{A39}
$$

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

For $0 < t < T$:

$$
\ddot{\lambda} + \left(2\beta\xi_r\dot{\xi}_r + 1 - \frac{\Delta}{2} \right)\lambda + 2\tilde{c}dR = 0 \tag{A40}
$$

$$
\ddot{\eta} + \left(2\beta\xi_l\dot{\xi}_l + 1 + \frac{\Delta}{2} \right)\eta + 2\tilde{c}dR = 0 \tag{A41}
$$

$$
\beta\left(1+\xi_r^2\right)\lambda - \alpha(\lambda+\eta) = 0 \tag{A42}
$$

$$
\beta\left(1+\xi_l^2\right)\eta - \alpha(\lambda+\eta) = 0 \tag{A43}
$$

with initial conditions at $t = T$:

$$
\lambda(T) = 0 \tag{A44}
$$

$$
\dot{\lambda}(T) = 0 \tag{A45}
$$

$$
\eta(T) = 0 \tag{A46}
$$

$$
\dot{\eta}(T) = 0 \tag{A47}
$$

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

$$
F_\alpha = \int_0^T -(\dot{\xi}_r + \dot{\xi}_l)(\lambda+\eta)dt \tag{A48}
$$

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

$$
F_\beta = \int_0^T \left(\left(1+\xi_r^2\right)\dot{\xi}_r\lambda + \left(1+\xi_l^2\right)\dot{\xi}_l\eta \right)dt \tag{A49}
$$

$$
F_\Delta = \int_0^T \frac{1}{2}(\xi_l\eta - \xi_r\lambda)dt \tag{A50}
$$

Having calculated the gradients of F w.r.t. the model parameters, we can now apply the gradient-descent rule to optimize our objective (A26):

$$
\begin{aligned}
\alpha^{k+1} &= \alpha^k - \tau^\alpha F_\alpha \\
\beta^{k+1} &= \beta^k - \tau^\beta F_\beta \\
\Delta^{k+1} &= \Delta^k - \tau^\Delta F_\Delta
\end{aligned}
\tag{A51}
$$

where $\tau^\cdot$ is the step-size.

The overall algorithm is summarized as follows:

1. Integrate (A27) and (A28) with initial conditions (A29)–(A32) from 0 to T, obtaining ξ_r, ξ_l, $\dot{\xi}_r$ and $\dot{\xi}_l$.
2. Integrate (A40)–(A43) with the initial conditions (A44)–(A47) from T to 0, obtaining λ, $\dot{\lambda}$, η and $\dot{\eta}$.
3. Update α, β and Δ with (A51).

Appendix D

Appendix D.1. The Combined Vocal Folds-Tract Model: Formulation of the Inverse Problem

In this appendix we formulate the problem of estimating the parameters of the combined vocal fold-tract model from speech measurements. Let $\Omega \times \mathcal{T}$ be the domain of volume velocity u, where Ω is the spatial domain, and $\mathcal{T}$ is the time domain. In the one-dimensional case, $\Omega = [0, L]$ where L is the length of vocal tract, and $\mathcal{T} = [0, T]$. Given measured acoustic pressure $p_m(t)$ at the lip, the corresponding volume velocity is given by [34]:

$$
u_m(t) = \frac{A(L)}{\rho c} p_m(t)
\tag{A52}
$$

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 $u_0(t) := u(0, t)$, $u_L(t) := u(L, t)$. The glottal flow $u_0(t)$ can be derived from the vocal folds displacement model (3) by:

$$
u_0(t) = \tilde{c} d (2\xi_0 + \xi_l(t) + \xi_r(t))
\tag{A53}
$$

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:

$$
\begin{aligned}
\mathcal{H} : \mathcal{L}^2(\Omega \times \mathcal{T}) \times \mathcal{L}^2(\Gamma \times \mathcal{T}) &\to \mathcal{L}^2(\Gamma \times \mathcal{T}) \\
(f, u_0) &\mapsto u_L
\end{aligned}
\tag{A54}
$$

where f is the vocal tract profile in (A23), and $\Gamma = \partial \Omega$ is the boundary.

We can split Γ into two parts $\Gamma = \Gamma_0 \bigcup \Gamma_1$, $\Gamma_0 \bigcap \Gamma_L = \varnothing$ corresponding to $x = 0$ and $x = L$. However, we neglect the difference to ease our derivation. Note that in the 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:

$$
\begin{aligned}
\mathcal{H}_f : \mathcal{L}^2(\Gamma \times \mathcal{T}) &\to \mathcal{L}^2(\Gamma \times \mathcal{T}) \\
u_0 &\mapsto u_L
\end{aligned}
\tag{A55}
$$

and

$$
\begin{aligned}
\mathcal{F} := \mathcal{H}_{u_0} : \mathcal{L}^2(\Omega \times \mathcal{T}) &\to \mathcal{L}^2(\Gamma \times \mathcal{T}) \\
f &\mapsto u_L
\end{aligned}
\tag{A56}
$$

Note that both $\mathcal{H}_f$ and $\mathcal{F}$ are bounded. Our objective is to (estimate model parameters α, β and Δ such that they) minimize the difference between the measured volume velocity u_m and the predicted volume velocity u_L near the lip subject to constraints:

$$\min \int_0^T \left(\mathcal{H}_f(u_0(t)) - u_m(t) \right)^2 dt \tag{A57}$$

$$\Leftrightarrow \quad \min \int_0^T \left(\mathcal{H}_f(\tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t))) - \frac{A(L)}{\rho c} p_m(t) \right)^2 dt \tag{A58}$$

$$\text{subject to} \quad \ddot{\xi}_r + \beta\left(1 + \xi_r^2\right)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r = \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right) \tag{A59}$$

$$\ddot{\xi}_l + \beta\left(1 + \xi_l^2\right)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l = \alpha\left(\dot{\xi}_r + \dot{\xi}_l\right) \tag{A60}$$

$$\text{(I.C.1)} \; \xi_r(0) = C_r \tag{A61}$$

$$\text{(I.C.2)} \; \xi_l(0) = C_l \tag{A62}$$

$$\text{(I.C.3)} \; \dot{\xi}_r(0) = 0 \tag{A63}$$

$$\text{(I.C.4)} \; \dot{\xi}_l(0) = 0 \tag{A64}$$

where (30) and (A60) represent the asymmetric vocal folds displacement model (3), I.C. stands for initial condition, and Cs are constants.

Appendix D.2. Solving the Inverse Problem for the Vocal Folds-Tract Model: The ADLES-VFT Algorithm

We solve the inverse problem for the combined vocal folds-tract model using the 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 (A60)–(A64), first we need to estimate the vocal tract profile f in $\mathcal{H}$ and (A23). Specifically, we need to solve:

$$\frac{\partial^2 u(x,t)}{\partial t^2} = c^2 \frac{\partial^2 u(x,t)}{\partial x^2} + f(x,t) \tag{A65}$$

subject to

$$\text{(B.C.1)} \quad u(0,t) = u_g(t) \tag{A66}$$

$$\text{(B.C.2)} \quad \frac{\partial u}{\partial n_\Gamma} = 0 \tag{A67}$$

$$\text{(I.C.1)} \quad u(x,0) = 0 \tag{A68}$$

$$\text{(I.C.2)} \quad \frac{\partial u(x,0)}{\partial t} = 0 \tag{A69}$$

where B.C. stands for boundary condition, u_g is the volume velocity at the glottis and n_Γ is the outward unit normal to the boundary Γ. We now derive the solution to (A65)–(A69).

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

$$f^{k+1} = f^k + \tau \delta f^k \tag{A70}$$

where $\delta f^k \in \mathcal{L}^2(\Omega \times \mathcal{T})$ is a small variation, and τ is a step size. A Taylor expansion of $\mathcal{F}$ (A56) at f^k gives:

$$\mathcal{F}\left(f^k + \delta f^k\right) = \mathcal{F}\left(f^k\right) + \mathcal{F}'\left(f^k\right)\delta f^k + O\left(\left(\delta f^k\right)^2\right) \tag{A71}$$

where $\mathcal{F}'$ is the Fréchet derivative [44]. Omitting higher order terms, we obtain:

$$\mathcal{F}'\left(f^k\right)\delta f^k = \mathcal{F}\left(f^k + \delta f^k\right) - \mathcal{F}\left(f^k\right) \tag{A72}$$

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

$$\begin{aligned} \mathcal{F}'(f) :&\mathcal{L}^2(\Omega \times \mathcal{T}) \quad \to \mathcal{L}^2(\Gamma \times \mathcal{T}) \\ &\delta f \mapsto \delta u_L \end{aligned} \tag{A73}$$

Correspondingly, the adjoint operator [44–46] is:

$$\begin{aligned} \mathcal{F}'(f)^* :&\mathcal{L}^2(\Gamma \times \mathcal{T}) \quad \to \mathcal{L}^2(\Omega \times \mathcal{T}) \\ &\delta u_L \mapsto \delta f \end{aligned} \tag{A74}$$

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:

$$\min \left\| \delta f^k \right\|_2^2$$

$$\text{subject to} \quad \mathcal{F}'\left(f^k\right)\delta f^k = u_m - \mathcal{F}\left(f^k\right)$$

It is simple to obtain the solution to (A75):

$$\delta f^k = -\mathcal{F}'\left(f^k\right)^* \left[\mathcal{F}'\left(f^k\right)\mathcal{F}'\left(f^k\right)^*\right]^{-1}\left(\mathcal{F}\left(f^k\right) - u_m\right) \tag{A75}$$

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 γI where I is the identity matrix.

We denote the estimation residual as:

$$r^k := u_m - \mathcal{F}\left(f^k\right) \tag{A76}$$

We have

$$\delta f^k = \frac{1}{\gamma}\mathcal{F}'\left(f^k\right)^* r^k \tag{A77}$$

Now consider the wave Equation (A65). Let $u + \delta u$ be a solution with variation $f + \delta f$. Substitution into (A65) yields:

$$\frac{\partial^2 (u + \delta u)}{\partial t^2} = c^2 \frac{\partial^2 (u + \delta u)}{\partial x^2} + f + \delta f \tag{A78}$$

Subtracting (A65) yields:

$$\frac{\partial^2 \delta u}{\partial t^2} = c^2 \frac{\partial^2 \delta u}{\partial x^2} + \delta f \tag{A79}$$

subject to

$$(\text{B.C.1}) \, \frac{\partial \delta u}{\partial n_\Gamma} = 0 \tag{A80}$$

$$(\text{I.C.1}) \, \delta u(x,0) = 0 \tag{A81}$$

$$(\text{I.C.2}) \, \frac{\partial \delta u(x,0)}{\partial t} \tag{A82}$$

Next, we use a time reversal technique [36] and backpropagate the difference (A76) into the vocal tract, which gives:

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2} + f(x,t) \tag{A83}$$

subject to

$$\text{(B.C.1)} \qquad \frac{\partial z}{\partial n_\Gamma} = r \tag{A84}$$

$$\text{(I.C.1)} \qquad z(x,T) = 0 \tag{A85}$$

$$\text{(I.C.2)} \qquad \frac{\partial z(x,T)}{\partial t} = 0 \tag{A86}$$

where z is the time reversal of u. It follows [47] that:

$$\langle \delta f, z \rangle_{\Omega \times T} = \int_0^T \int_\Omega \delta f z \, dx \, dt \tag{A87}$$

$$= \int_0^T \int_\Omega \left(\frac{\partial^2 \delta u}{\partial t^2} - c^2 \frac{\partial^2 \delta u}{\partial x^2} \right) z \, dx \, dt \tag{A88}$$

$$= \int_0^T \int_\Omega \left(\frac{\partial^2 \delta u}{\partial t^2} - c^2 \frac{\partial^2 \delta u}{\partial x^2} \right) z \, dx \, dt - \int_0^T \int_\Omega \left(\frac{\partial^2 z}{\partial t^2} - c^2 \frac{\partial^2 z}{\partial x^2} - f \right) \delta u \, dx \, dt \tag{A89}$$

$$= \int_0^T \int_\Omega \left(\frac{\partial^2 \delta u}{\partial t^2} z - \frac{\partial^2 z}{\partial t^2} \delta u \right) dx \, dt - c^2 \int_0^T \int_\Omega \left(\frac{\partial^2 \delta u}{\partial x^2} z - \frac{\partial^2 z}{\partial x^2} \delta u \right) dx \, dt$$
$$+ \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A90}$$

$$= \int_\Omega \left(\frac{\partial \delta u}{\partial t} z - \frac{\partial z}{\partial t} \delta u \right) \Big|_0^T dx \, dt - c^2 \int_0^T \int_\Omega \left(\frac{\partial^2 \delta u}{\partial x^2} z - \frac{\partial^2 z}{\partial x^2} \delta u \right) dx \, dt$$
$$+ \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A91}$$

$$= -c^2 \int_0^T \int_\Omega \left(\frac{\partial^2 \delta u}{\partial x^2} z - \frac{\partial^2 z}{\partial x^2} \delta u \right) dx \, dt + \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A92}$$

$$= -c^2 \int_0^T \int_\Omega \left(z d \frac{\partial \delta u}{\partial x} - \delta u d \frac{\partial z}{\partial x} \right) dt + \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A93}$$

$$= -c^2 \int_0^T \left(\int_\Gamma z \frac{\partial \delta u}{\partial n_\Gamma} ds - \int_\Omega \frac{\partial \delta u}{\partial x} \frac{\partial z}{\partial x} dx - \int_\Gamma \delta u \frac{\partial z}{\partial n_\Gamma} ds + \int_\Omega \frac{\partial \delta u}{\partial x} \frac{\partial z}{\partial x} dx \right) dt$$
$$+ \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A94}$$

$$= c^2 \int_0^T \int_\Gamma \delta u \frac{\partial z}{\partial n_\Gamma} ds \, dt + \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A95}$$

$$= c^2 \int_0^T \int_\Gamma \delta u r \, ds \, dt + \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A96}$$

$$= c^2 \int_0^T \int_\Gamma \mathcal{F}'(f) \delta f r \, ds \, dt + \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A97}$$

$$= c^2 \int_0^T \int_\Omega \delta f \mathcal{F}'(f)^* r \, dx \, dt + \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A98}$$

$$= c^2 \int_0^T \int_\Omega \delta f \mathcal{F}'(f)^* r \, dx \, dt - \int_0^T \int_\Omega f \delta u \, dx \, dt \tag{A99}$$

$$= c^2 \int_0^T \int_\Omega \delta f \left(\mathcal{F}'(f)^* r - u \right) dx \, dt \tag{A100}$$

where we substitute from (A87)–(A89) into (A79) and (A83); from (A89)–(A92) we apply initial conditions (A81), (A82), (A85) and (A86); from (A92)–(A94) we integrate by parts;

from (A94)–(A95) we apply boundary condition (A80); from (A95)–(A96) we use boundary condition (A84); from (A96)–(A97) we use definition (A73); from (A97)–(A98) we use definition (A74) and the duality property

$$\langle \mathcal{F}'(f)\delta f, r \rangle_{\Gamma \times \mathcal{T}} = \langle \delta f, \mathcal{F}'(f)^* r \rangle_{\Omega \times \mathcal{T}} \tag{A101}$$

from (A98)–(A99), we assume that the second-order variation is small, i.e.,

$$\langle f + \delta f, u + \delta u \rangle = \langle f, u \rangle + \langle f, \delta u \rangle + \langle \delta f, u \rangle + \langle \delta f, \delta u \rangle \approx \langle f, u \rangle \tag{A102}$$

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

$$z = c^2 \left(\mathcal{F}'(f)^* r - u \right) \tag{A103}$$

and hence

$$\mathcal{F}'(f)^* r = \frac{z}{c^2} + u \tag{A104}$$

Substitution into (A77) and (A70) yields:

$$f^{k+1} = f^k + \iota \left(\frac{z^k}{c^2} + u^k \right) \tag{A105}$$

where $\iota = \tau/\gamma$.

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

Appendix D.2.1. Estimating model parameters via the adjoint least squares method

Now, we derive solution to the parameter estimation problem (A60)–(A64) using the adjoint least squares method of Appendix C as follows.

Denote the estimation error as:

$$f(\xi_l, \xi_r; \vartheta) = \left(\mathcal{H}_f(\tilde{c}d(2\xi_0 + \xi_l(t) + \xi_r(t))) - \frac{A(L)}{\rho c} p_m(t) \right)^2 \tag{A106}$$

and

$$F(\xi_l, \xi_r; \vartheta) = \int_0^T f(\xi_l, \xi_r; \vartheta) dt \tag{A107}$$

where $\vartheta = [\alpha, \beta, \Delta]$ are the parameters of the vocal folds model (3). We would like to obtain update rules for the model parameters α, β, and Δ, i.e.,

$$\alpha^{k+1} = \alpha^k - \tau^\alpha F_{\alpha^k} \tag{A108}$$
$$\beta^{k+1} = \beta^k - \tau^\beta F_{\beta^k} \tag{A109}$$
$$\Delta^{k+1} = \Delta^k - \tau^\Delta F_{\Delta^k} \tag{A110}$$

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

$$\begin{aligned} \mathcal{L}(\vartheta) = \int_0^T & \left[f + \lambda \left(\ddot{\xi}_r + \beta \left(1 + \xi_r^2 \right) \dot{\xi}_r + \xi_r - \frac{\Delta}{2} \xi_r - \alpha \left(\dot{\xi}_r + \dot{\xi}_l \right) \right) \right. \\ & \left. + \eta \left(\ddot{\xi}_l + \beta \left(1 + \xi_l^2 \right) \dot{\xi}_l + \xi_l + \frac{\Delta}{2} \xi_l - \alpha \left(\dot{\xi}_r + \dot{\xi}_l \right) \right) \right] dt \\ & + \mu_l (\xi_l(0) - C_l) + \mu_r (\xi_r(0) - C_r) + \nu_l \dot{\xi}_l(0) + \nu_r \dot{\xi}_r(0) \end{aligned} \tag{A111}$$

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

$$
\begin{aligned}
\mathcal{L}_\alpha = \int_0^T \Bigg[& 2\tilde{c}d\mathcal{H}'_f\Big|_{u_0} (\partial_\alpha \xi_l + \partial_\alpha \xi_r) \\
& + \lambda \Big(\partial_\alpha \ddot{\xi}_r + 2\beta \dot{\xi}_r \xi_r \partial_\alpha \xi_r + \beta(1 + \xi_r^2)\partial_\alpha \dot{\xi}_r + \partial_\alpha \xi_r \\
& - \frac{\Delta}{2} \partial_\alpha \xi_r - \alpha(\partial_\alpha \dot{\xi}_r + \partial_\alpha \dot{\xi}_l) - (\dot{\xi}_r + \dot{\xi}_l) \Big) \\
& + \eta \Big(\partial_\alpha \ddot{\xi}_l + 2\beta \dot{\xi}_l \xi_l \partial_\alpha \xi_l + \beta(1 + \xi_l^2)\partial_\alpha \dot{\xi}_l + \partial_\alpha \xi_l \\
& + \frac{\Delta}{2} \partial_\alpha \xi_l - \alpha(\partial_\alpha \dot{\xi}_r + \partial_\alpha \dot{\xi}_l) - (\dot{\xi}_r + \dot{\xi}_l) \Big) \Bigg] dt \\
& + \mu_l \partial_\alpha \xi_l(0) + \mu_r \partial_\alpha \xi_r(0) + \nu_l \partial_\alpha \dot{\xi}_l(0) + \nu_r \partial_\alpha \dot{\xi}_r(0)
\end{aligned}
\tag{A112}
$$

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

$$
\int_0^T \lambda \partial_\alpha \ddot{\xi}_r dt = \int_0^T \partial_\alpha \xi_r \ddot{\lambda} dt - \partial_\alpha \xi_r \dot{\lambda}\Big|_0^T + \partial_\alpha \dot{\xi}_r \lambda\Big|_0^T
\tag{A113}
$$

Defining the estimation residual $R := \mathcal{H}_f(u_0) - \frac{A(L)}{\rho c}p_m(t)$, applying this to $\eta \partial_\alpha \ddot{\xi}_l$, and substituting into (A112), after simplification yields:

$$
\begin{aligned}
\mathcal{L}_\alpha = \int_0^T \Bigg[& \Big(\ddot{\lambda} + \Big(2\beta \xi_r \dot{\xi}_r + 1 - \frac{\Delta}{2}\Big)\lambda + 2\tilde{c}dR\mathcal{H}'_f\Big|_{u_0} \Big)\partial_\alpha \xi_r \\
& + \Big(\ddot{\eta} + \Big(2\beta \xi_l \dot{\xi}_l + 1 + \frac{\Delta}{2}\Big)\lambda + 2\tilde{c}dR\mathcal{H}'_f\Big|_{u_0} \Big)\partial_\alpha \xi_l \\
& + \Big(\beta(1 + \xi_r^2)\lambda - \alpha(\lambda + \eta) \Big)\partial_\alpha \dot{\xi}_r + \Big(\beta(1 + \xi_l^2)\eta - \alpha(\lambda + \eta) \Big)\partial_\alpha \dot{\xi}_l - \\
& (\dot{\xi}_r + \dot{\xi}_l)(\lambda + \eta) \Bigg] dt \\
& + (\mu_r + \dot{\lambda})\partial_\alpha \xi_r(0) - \dot{\lambda}\partial_\alpha \xi_r(T) + (\nu_r - \lambda)\partial_\alpha \dot{\xi}_r(0) + \lambda\partial_\alpha \dot{\xi}_r(T) \\
& + (\mu_l + \dot{\eta})\partial_\alpha \xi_l(0) - \dot{\eta}\partial_\alpha \xi_l(T) + (\nu_l - \eta)\partial_\alpha \dot{\xi}_l(0) + \eta\partial_\alpha \dot{\xi}_l(T)
\end{aligned}
\tag{A114}
$$

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:

For $0 < t < T$:

$$
\ddot{\lambda} + \Big(2\beta \xi_r \dot{\xi}_r + 1 - \frac{\Delta}{2}\Big)\lambda + 2\tilde{c}dR\mathcal{H}'_f\Big|_{u_0} = 0
\tag{A115}
$$

$$
\ddot{\eta} + \Big(2\beta \xi_l \dot{\xi}_l + 1 + \frac{\Delta}{2}\Big)\eta + 2\tilde{c}dR\mathcal{H}'_f\Big|_{u_0} = 0
\tag{A116}
$$

$$
\beta\Big(1 + \xi_r^2\Big)\lambda - \alpha(\lambda + \eta) = 0
\tag{A117}
$$

$$
\beta\Big(1 + \xi_l^2\Big)\eta - \alpha(\lambda + \eta) = 0
\tag{A118}
$$

with initial conditions:

At $t = T$:

$$\lambda(T) = 0 \tag{A119}$$
$$\dot{\lambda}(T) = 0 \tag{A120}$$
$$\eta(T) = 0 \tag{A121}$$
$$\dot{\eta}(T) = 0 \tag{A122}$$

Assuming the boundary conditions on ξ_l and ξ_r to hold—an assumption we empirically find to be valid for $C_r = C_l = 0$, we obtain the derivative of F w.r.t. α:

$$F_\alpha = \int_0^T -\left(\dot{\xi}_r + \dot{\xi}_l\right)(\lambda + \eta)dt \tag{A123}$$

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

$$F_\beta = \int_0^T \left(\left(1 + \xi_r^2\right)\dot{\xi}_r \lambda + \left(1 + \xi_l^2\right)\dot{\xi}_l \eta\right)dt \tag{A124}$$

$$F_\Delta = \int_0^T \frac{1}{2}(\xi_l \eta - \xi_r \lambda)dt \tag{A125}$$

Appendix D.2.2. The ADLES-VFT Algorithm Summarized

The algorithm for solving the parameter estimation problem (A60)–(A64) is outlined below.

1. Integrate (A59) and (A60) with initial conditions (A61)–(A64) from 0 to T, obtaining $\xi_r^k, \xi_l^k, \dot{\xi}_r^k$ and $\dot{\xi}_l^k$.
2. Solve the forward propagation model (A65)–(A69) for $u_L^k, \mathcal{H}'_f\big|_{u_0^k}$.
3. Calculate the estimation difference r^k using (A76).
4. Solve the backward propagation model (A83)–(A86) for z^k.
5. Update f^k using (A105).
6. Integrate (A115)–(A118) with initial conditions (A119)–(A122) from T to 0, obtaining $\lambda^k, \dot{\lambda}^k, \eta^k$ and $\dot{\eta}^k$.
7. Update α, β and Δ with (A110).

Appendix E

Appendix E.1. Numerical Solution for Wave Propagation

In this appendix we describe a finite-element solution for time-dependent systems of PDEs examplified by the wave propagation equation:

$$\frac{\partial^2 u(x,t)}{\partial t^2} = c^2 \frac{\partial^2 u(x,t)}{\partial x^2} + f(x,t) \tag{A126}$$

subject to

(B.C.1) $u(0,t) = u_g(t)$ \hfill (A127)

(B.C.2) $\dfrac{\partial u}{\partial n_\Gamma} = 0$ \hfill (A128)

(I.C.1) $u(x,0) = 0$ \hfill (A129)

(I.C.2) $\dfrac{\partial u(x,0)}{\partial t} = 0$ \hfill (A130)

and the reverse propagation equation:

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2} + f(x, t) \tag{A131}$$

subject to

$$(\text{B.C.1}) \qquad \frac{\partial z}{\partial n_\Gamma} = r \tag{A132}$$

$$(\text{I.C.1}) \qquad z(x, T) = 0 \tag{A133}$$

$$(\text{I.C.2}) \qquad \frac{\partial z(x, T)}{\partial t} = 0 \tag{A134}$$

where the solution must be defined for $x \in (0, L)$ and $t \in (0, T)$.

Appendix E.1.1. Variational Formulation

First, for the time-dependent system of PDEs, we discretize it along time t with the backward Euler method [48], yielding a sequence of differential equations. We split the time domain T into N uniform length intervals Δt. For time step n, $0 \le n \le N - 1$, applying the backward Euler method to the left side of (A65) gives:

$$\left[D_t D_t^- u\right]^n := D_t D_t^- \left(\frac{\partial^2 u}{\partial t^2}\right)\bigg|_{t=n\Delta t} = \frac{u^n - 2u^{n-1} + u^{n-2}}{\Delta t^2} \tag{A135}$$

where $D_t D_t^-$ is a finite difference operator w.r.t. time, and the superscript n indicates that the differencing is performed at time step n [48,49]. Substitution into (36) yields:

$$\left[D_t D_t^- u = c^2 \frac{\partial^2 u}{\partial x^2} + f\right]^n \tag{A136}$$

$$\Leftrightarrow \quad u^n = \Delta t^2 c^2 \frac{\partial^2 u^n}{\partial x^2} + \Delta t^2 f^n + 2u^{n-1} - u^{n-2} \tag{A137}$$

Next, we define the residual at time step n as:

$$R^n = u^n - \Delta t^2 c^2 \frac{\partial^2 u^n}{\partial x^2} + \Delta t^2 f^n + 2u^{n-1} - u^{n-2} \tag{A138}$$

Applying Galerkin's method [48,50] gives:

$$\langle R^n, v \rangle_{W^{k,2}} = 0 \tag{A139}$$

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 (A139) yields:

$$\int_\Omega u^n v dx - \Delta t^2 c^2 \int_\Omega \frac{\partial^2 u^n}{\partial x^2} v dx = \int_\Omega \left(\Delta t^2 f^n + 2u^{n-1} - u^{n-2}\right) v dx \tag{A140}$$

Integration by parts for the second-order term in (A140) gives:

$$\int_\Omega \frac{\partial^2 u^n}{\partial x^2} v dx = -\int_\Omega \frac{\partial u^n}{\partial x} \frac{\partial v}{\partial x} dx + \int_\Gamma \frac{\partial u^n}{\partial n_\Gamma} ds \tag{A141}$$

where n_Γ is the outward normal unit vector of the boundary Γ, and ds is the 1-form [43] on Γ. For problem (A126)–(A130), applying the boundary condition (A128) and substituting (A141) back into (A140) yields the variational problem:

$$\int_\Omega u^n v dx + \Delta t^2 c^2 \int_\Omega \frac{\partial u^n}{\partial x} \frac{\partial v}{\partial x} dx = \int_\Omega \left(\Delta t^2 f^n + 2u^{n-1} - u^{n-2}\right) v dx \tag{A142}$$

For the reverse problem represented by (A131)–(A134), applying the boundary condition (A132) and substituting (A141) back into (A140) yields a similar variational problem:

$$\int_\Omega z^n w dx + \Delta t^2 c^2 \int_\Omega \frac{\partial z^n}{\partial x}\frac{\partial w}{\partial x} dx = \int_\Omega \left(\Delta t^2 f^n + 2z^{n-1} - z^{n-2}\right) w dx + \int_\Gamma r^n w ds \quad \text{(A143)}$$

We can split the variational problem (A142) into two parts:

$$a(u,v) = \int_\Omega uv dx + \Delta t^2 c^2 \int_\Omega \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} dx \quad \text{(A144)}$$

$$L(v) = \int_\Omega \left(\Delta t^2 f^n + 2u^{n-1} - u^{n-2}\right) v dx \quad \text{(A145)}$$

where we have interchanged the unknown u^n with u. (A144) is the bilinear form, and (A145) is the linear form [48].

Our original problem (A126)–(A130) and (A131)–(A134) then reduce to solving:

$$a(u,v) = L(v) \quad \text{(A146)}$$

for each time step. By the Lax-Milgram Lemma [49], solving (A146) is equivalent to solving the functional minimization problem:

$$F(u) = \arg\min_{v \in \mathcal{V}} \frac{1}{2} a(v,v) - L(v) \quad \text{(A147)}$$

By the calculus of variations, and taking the variation of the functional gives (A146), hence the name *variational* form [48,49].

Appendix E.1.2. Finite Element Approximation

For each time step, we solve (A146) 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 [48,51].

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 [48,51]. Each cell is essentially a simplex and can be continuously transformed into the physical domain. We use the following two previously established points in this context, and proceed as explained thereafter.

1. **Existence of Unique Solution**: The solution to the variational problem (A146) exists and is unique [51].
2. **Approximation Error**: The Galerkin's method gives the solution u_e with error bounded by $\mathcal{O}(h^3 \|D^2 u_e\|_{\mathcal{W}^{3,2}})$, where h is the cell size and D is the bounded derivative operator [49,51].

Assume a solution $u = B + c^j \psi_j$ (using Einstein summation convention) with basis $\psi_j \in P_2$ and coefficients c^j. Here the ψ_j are piece-wise second-order Lagrange polynomials [48]. Also, the mesh of cells, P_2 elements of basis functions, coordinate map, DoF map, and the assembly of the linear systems, are implemented and solved with the FEniCS software [52].

The function $B(x)$ incorporates the boundary condition and, as an example, can take the form:

$$B(x) = u_g + (u_m - u_g)\frac{x^p}{L^p}, \quad p > 0 \quad \text{(A148)}$$

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 $v = \hat{\psi}_i$ (where $\hat{\psi}_i$ is the ith Lagrange polynomial basis in a test cell). Substitution into (A144) and (A145) yields:

$$
\begin{aligned}
a(u,v) &= \int_\Omega U^j \psi_j \hat{\psi}_i dx + \Delta t^2 c^2 \int_\Omega U^j \psi_j' \psi_i' dx \\
&= \left(\int_\Omega \hat{\psi}_i \psi_j dx + \Delta t^2 c^2 \int_\Omega \psi_i' \psi_j' dx \right) U^j
\end{aligned}
\tag{A149}
$$

$$
\begin{aligned}
L(v) &= \int_\Omega \left(\Delta t^2 F_n^j \psi_j + 2 U_{n-1}^j \psi_j - U_{n-2}^j \psi_j \right) \hat{\psi}_i dx \\
&= \Delta t^2 \left(\int_\Omega \hat{\psi}_i \psi_j dx \right) F_n^j + 2 \left(\int_\Omega \hat{\psi}_i \psi_j dx \right) U_{n-1}^j - \left(\int_\Omega \hat{\psi}_i \psi_j dx \right) U_{n-2}^j
\end{aligned}
\tag{A150}
$$

where ψ_i' is the derivative of ψ_i w.r.t. x.

Setting $M_{i,j} = \int_\Omega \hat{\psi}_i \psi_j dx$, $K_{i,j} = \int_\Omega \psi_i' \psi_j' dx$ and collecting (A149) and (A150) into matrix-vector form, we obtain:

$$
AU = b
\tag{A151}
$$

where $A = M + \Delta t^2 c^2 K$, and $b = \Delta t^2 M F^n + 2 M U^{n-1} - M U^{n-2}$. Hence, we have reduced the problem of (A146) into solving the linear system (A151), 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.

Appendix F

This appendix provides some essential definitions used for the characterization of dynamical system models for analysis.

Definition A1 (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 [53].*

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.

Appendix F.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 A2 (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 $t_1, t_2 + t_1 \in I(x), t_2 \in I(\Phi(t_1, x))$.*

Appendix F.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 A3 (Flow, orbit, invariance). *The map $\Phi_x : I(x) \to M$ is the* flow *or* trajectory *through x. The set of all flows $\gamma_x := \{\Phi_x \mid t \in I(x)\}$ 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 $t \in T$.*

Appendix F.3. Phase Space Behavior: Attractor

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

Definition A4 (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 A1. 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 [20,54], and are called *strange attractors*.
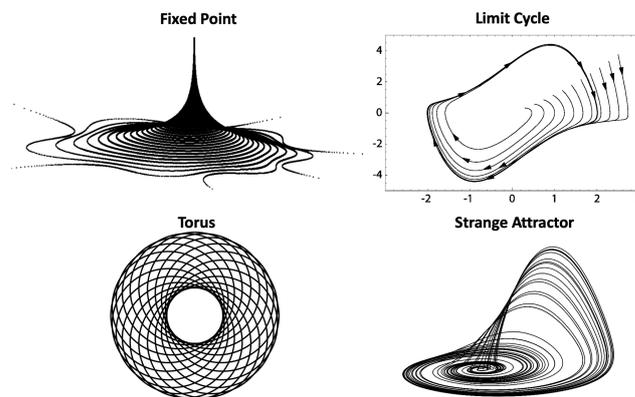

Figure A1. Illustration of different attractors in a dynamical system.

Appendix F.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 A5 (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*

$$\lim_{n \to \infty} \inf d(\Phi^n(x), \Phi^n(y)) = 0 \tag{A152}$$

$$\lim_{n \to \infty} \sup d(\Phi^n(x), \Phi^n(y)) > 0 \tag{A153}$$

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.

Appendix F.5. Stability

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

Definition A6 (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.*

Appendix F.6. Poincaré Map and Poincaré Section

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

Definition A7 (Poincaré map [53]). *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 $P(x)$.*

Appendix F.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 A8 (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.

References

1. Cveticanin, L. Review on Mathematical and Mechanical Models of the Vocal Cord. *J. Appl. Math.* **2012**, *2012*, 928591. [CrossRef]
2. Titze, I.R. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* **1988**, *83*, 1536–1552. [CrossRef] [PubMed]
3. Döllinger, M.; Gómez, P.; Patel, R.R.; Alexiou, C.; Bohr, C.; Schützenberger, A. Biomechanical simulation of vocal fold dynamics in adults based on laryngeal high-speed videoendoscopy. *PLoS ONE* **2017**, *12*, e0187486. [CrossRef] [PubMed]
4. Herbst, C.T.; Fitch, W.; Švec, J.G. Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively. *J. Acoust. Soc. Am.* **2010**, *128*, 3070–3078. [CrossRef] [PubMed]
5. Mergell, P.; Herzel, H.; Titze, I.R. Irregular vocal-fold vibration—High-speed observation and modeling. *J. Acoust. Soc. Am.* **2000**, *108*, 2996–3002. [CrossRef]
6. Zhang, Z. Mechanics of human voice production and control. *J. Acoust. Soc. Am.* **2016**, *140*, 2614–2635. [CrossRef]
7. Tao, C.; Zhang, Y.; Hottinger, D.G.; Jiang, J.J. Asymmetric airflow and vibration induced by the Coanda effect in a symmetric model of the vocal folds. *J. Acoust. Soc. Am.* **2007**, *122*, 2270–2278. [CrossRef]
8. Erath, B.D.; Plesniak, M.W. The occurrence of the Coanda effect in pulsatile flow through static models of the human vocal folds. *J. Acoust. Soc. Am.* **2006**, *120*, 1000–1011. [CrossRef]
9. Singh, R. *Profiling Humans from Their Voice*; Springer-Nature: Singapore, 2019.
10. Flanagan, J.; Landgraf, L. Self-oscillating source for vocal-tract synthesizers. *IEEE Trans. Audio Electroacoust.* **1968**, *16*, 57–64. [CrossRef]
11. Ishizaka, K.; Flanagan, J.L. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst. Tech. J.* **1972**, *51*, 1233–1268. [CrossRef]

12. Zhang, Z.; Neubauer, J.; Berry, D.A. The influence of subglottal acoustics on laboratory models of phonation. *J. Acoust. Soc. Am.* **2006**, *120*, 1558–1569. [CrossRef] [PubMed]
13. Zhao, W.; Zhang, C.; Frankel, S.H.; Mongeau, L. Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms. *J. Acoust. Soc. Am.* **2002**, *112*, 2134–2146. [CrossRef] [PubMed]
14. Zhang, C.; Zhao, W.; Frankel, S.H.; Mongeau, L. Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds. *J. Acoust. Soc. Am.* **2002**, *112*, 2147–2154. [CrossRef] [PubMed]
15. Lucero, J.C. Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region. *J. Acoust. Soc. Am.* **1993**, *94*, 3104–3111. [CrossRef]
16. Lucero, J.C.; Schoentgen, J. Modeling vocal fold asymmetries with coupled van der Pol oscillators. *Proc. Mtgs. Acoust* **2013**, *19*, 060165.
17. Alipour, F.; Berry, D.A.; Titze, I.R. A finite-element model of vocal-fold vibration. *J. Acoust. Soc. Am.* **2000**, *108*, 3003–3012. [CrossRef]
18. Yang, A.; Stingl, M.; Berry, D.A.; Lohscheller, J.; Voigt, D.; Eysholdt, U.; Döllinger, M. Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model. *J. Acoust. Soc. Am.* **2011**, *130*, 948–964. [CrossRef]
19. Pickup, B.A.; Thomson, S.L. Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models. *J. Biomech.* **2009**, *42*, 2219–2225. [CrossRef]
20. Jiang, J.J.; Zhang, Y.; Stern, J. Modeling of chaotic vibrations in symmetric vocal folds. *J. Acoust. Soc. Am.* **2001**, *110*, 2120–2128. [CrossRef]
21. Titze, I.R. Nonlinear source—Filter coupling in phonation: Theory. *J. Acoust. Soc. Am.* **2008**, *123*, 1902–1915. [CrossRef]
22. Story, B.H.; Titze, I.R. Voice simulation with a body-cover model of the vocal folds. *J. Acoust. Soc. Am.* **1995**, *97*, 1249–1260. [CrossRef] [PubMed]
23. Chan, R.W.; Titze, I.R. Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics. *J. Acoust. Soc. Am.* **2006**, *119*, 2351–2362. [CrossRef]
24. Lucero, J.C.; Schoentgen, J.; Haas, J.; Luizard, P.; Pelorson, X. Self-entrainment of the right and left vocal fold oscillators. *J. Acoust. Soc. Am.* **2015**, *137*, 2036–2046. [CrossRef] [PubMed]
25. Maeda, S. 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: Berlin/Heidelberg, Germany, 1990; pp. 131–149.
26. Birkholz, P.; Kröger, B.J. Simulation of vocal tract growth for articulatory speech synthesis. In Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6–10 August 2007; pp. 377–380.
27. Dang, J.; Honda, K. Construction and control of a physiological articulatory model. *J. Acoust. Soc. Am.* **2004**, *115*, 853–870. [CrossRef]
28. Portnoff, M.R. A Quasi-One-Dimensional Digital Simulation for the Time-Varying Vocal Tract. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1973.
29. Allen, D.R.; Strong, W.J. A model for the synthesis of natural sounding vowels. *J. Acoust. Soc. Am.* **1985**, *78*, 58–69. [CrossRef]
30. Motoki, K.; Pelorson, X.; Badin, P.; Matsuzaki, H. Computation of 3-D vocal tract acoustics based on mode-matching technique. In Proceedings of the Sixth International Conference on Spoken Language Processing, Beijing, China, 16–20 October 2000.
31. Zhao, W.; Singh, R. Speech-based parameter estimation of an asymmetric vocal fold oscillation model and its application in discriminating vocal fold pathologies. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 7344–7348.
32. Erath, B.D.; Plesniak, M.W. An investigation of jet trajectory in flow through scaled vocal fold models with asymmetric glottal passages. *Exp. Fluids* **2006**, *41*, 735–748. [CrossRef]
33. Eisner, E. Complete solutions of the "Webster" horn equation. *J. Acoust. Soc. Am.* **1967**, *41*, 1126–1146. [CrossRef]
34. Titze, I.R.; Martin, D.W. Principles of voice production. *Acoust. Soc. Am. J.* **1998**, *104*, 1148. [CrossRef]
35. Alku, P. Glottal inverse filtering analysis of human voice production—A review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana* **2011**, *36*, 623–650. [CrossRef]
36. Morse, P.M.; Ingard, K.U. *Theoretical Acoustics*; Princeton University Press: Princeton, NJ, USA, 1986.
37. Steinecke, I.; Herzel, H. Bifurcations in an asymmetric vocal-fold model. *J. Acoust. Soc. Am.* **1995**, *97*, 1874–1884. [CrossRef]
38. Bhat, C.; Kopparapu, S.K. FEMH Voice Data Challenge: Voice disorder Detection and Classification using Acoustic Descriptors. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5233–5237.
39. Al Ismail, M.; Deshmukh, S.; Singh, R. Detection of COVID-19 through the analysis of vocal fold oscillations. In Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1035–1039.
40. Deshmukh, S.; Al Ismail, M.; Singh, R. Interpreting glottal flow dynamics for detecting COVID-19 from voice. In Proceedings of the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1055–1059.
41. Zhang, J. Vocal Fold Dynamics for Automatic Detection of Amyotrophic Lateral Sclerosis from Voice. Master's Thesis, Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA, 2022.

42. Lee, K.B.; Kim, J.H. Mass-spring-damper motion dynamics-based particle swarm optimization. In Proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–6 June 2008; pp. 2348–2353.
43. Do Carmo, M.P.; Flaherty Francis, J. *Riemannian Geometry*; Springer: Berlin/Heidelberg, Germany, 1992; Volume 6.
44. Kantorovich, L.V.; Akilov, G.P. *Functional Analysis*; Elsevier: Amsterdam, The Netherlands, 2016.
45. Zhu, K. *Operator Theory in Function Spaces*; American Mathematical Soc.: Providence, RI, USA, 2007; No. 138.
46. Giles, M.B.; Süli, E. Adjoint methods for PDEs: A posteriori error analysis and postprocessing by duality. *Acta Numer.* **2002**, *11*, 145–236. [CrossRef]
47. Dong, C.; Jin, Y. MIMO nonlinear ultrasonic tomography by propagation and backpropagation method. *IEEE Trans. Image Process.* **2012**, *22*, 1056–1069. [CrossRef] [PubMed]
48. Langtangen, H.P.; Mardal, K.A. *Introduction to Numerical Methods for Variational Problems*; Springer Nature: Berlin/Heidelberg, Germany, 2019; Volume 21.
49. Ames, W.F. *Numerical Methods for Partial Differential Equations*; Academic Press: Cambridge, MA, USA, 2014.
50. Thomée, V. *Galerkin Finite Element Methods for Parabolic Problems*; Springer: Berlin/Heidelberg, Germany, 1984; Volume 1054.
51. Larson, M.G.; Bengzon, F. The finite element method: Theory, implementation, and practice. *Texts Comput. Sci. Eng.* **2010**, *10*, 23–44.
52. Alnæs, M.; Blechta, J.; Hake, J.; Johansson, A.; Kehlet, B.; Logg, A.; Richardson, C.; Ring, J.; Rognes, M.E.; Wells, G.N. The FEniCS project version 1.5. *Arch. Numer. Softw.* **2015**, *3*.
53. Birkhoff, G.D. *Dynamical Systems*; American Mathematical Soc.: Providence, RI, USA, 1927; Volume 9.
54. Jiang, J.J.; Zhang, Y. Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds. *J. Acoust. Soc. Am.* **2002**, *112*, 2127–2133. [CrossRef] [PubMed]