


Article

Voice Simulation: The Next Generation

Ingo R. Titze ^{1,*}  and Jorge C. Lucero ² ¹ Utah Center for Vocology, University of Utah, Salt Lake City, UT 84112, USA² Department of Computer Science, University of Brasília, Brasília 70910-900, Brazil

* Correspondence: ingo.titze@utah.edu; Tel.: +1-801-646-6520

Abstract: Simulation of the acoustics and biomechanics of sound production in humans and animals began half a century ago. The three major components are the mechanics of tissue under self-sustained oscillation, the transport of air from the lungs to the lips, and the propagation of sound in the airways. Both low-dimensional and high-dimensional computer models have successfully predicted control of pitch, loudness, spectral content, vowel production, and many other features of speaking and singing. However, the problems of computational efficiency, validity, and accuracy have not been adequately addressed. Low-dimensional models are often more revealing of nonlinear phenomena in coupled oscillators, but the simplifying assumptions are not always validated. High-dimensional models can provide more accuracy, but interpretations of results are sometimes clouded by computational redundancy and uncertainty of parameters. The next generation will likely combine pre-calculations and machine learning with abbreviated critical calculations.

Keywords: computer simulation; voice; speech; vocalization; phonation



Citation: Titze, I.R.; Lucero, J.C. Voice Simulation: The Next Generation. *Appl. Sci.* **2022**, *12*, 11720. <https://doi.org/10.3390/app122211720>

Academic Editor: Sten Ternström

Received: 25 October 2022

Accepted: 16 November 2022

Published: 18 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For studies of vocalization, choices for media of experimentation are human subjects, animal models, physical models, and computer simulation. Among these media, computer simulation is rapidly becoming the first choice of experimentation, especially if the physical plant is involved. It is expected that simulation will continue to play a major role in scientific validation of voice therapies, surgery, and voice training. Computer-assisted surgery began several decades ago and continues to be refined with individual-specific simulations. Better understanding of the biology of vocal fatigue, injury, and recovery will lead to damage risk criteria that can be quantified computationally. On the general health improvement side, benefits derived from non-abusive vocalization can also be quantified in terms of healing effects of self-induced tissue vibration. In the vocal performing arts, life span changes, voice quality distinctions, enhancement of the sound source by the vocal tract, and amplified vs. unamplified singing can be addressed computationally.

The strongest rationale for simulation is that nothing can physically be broken or injured by experimentation, even if natural limits are exceeded. Breakdown of the system is manifested simply by an error message on an unreasonable number, requiring nothing more than a new input variable, perhaps some reprogramming, and a re-start. Furthermore, repeatability is so exact that ordinary statistics are mostly superfluous. Perhaps the greatest advantage that simulation has over human and animal experimentation is the ability to change one variable at a time. It helps to discriminate what we see and hear in a voice based on individual components that often co-vary. Simulation also allows extreme or asymptotic conditions to be tested. Even though these may never occur in nature, they are incredibly important for conceptualization. For example, source-filter interaction can be eliminated, a condition that is impossible to achieve with human subjects. Lastly, simulation is not burdened by the problem of “distortion by measurement.” Multiple instruments attached to a human or an animal, especially invasive instruments, can distort natural behavior. In

simulation, multiple outputs or heavy monitoring of the simulation may slow down the process, but they do not distort the outcome.

The strongest rationale against simulation is that the result is never more accurate than the weakest physical law or weakest parameter entered into the code. Assumptions are made by the programmers, those who decide on the appropriate cause-effect relations in the code. Additionally, the code is often not fully debugged before it gets to the user. This is not necessarily an indictment of the programmer's ability, but simply a reality that not all conditions or applications can be imagined *a priori*. Simulation is an inside-out approach, in contrast to measurement, which is an outside-in approach. Given that biology generally offers multiple solutions to produce the same output, i.e., biology works on a many-to-one transformation of variables, the combination of simulation and measurement is critical. Repeated iteration between measurement and simulation can provide much confidence in the outcome. Finally, simulation requires long-term commitment by one or more investigators. Too often, programmers and engineers offer significant, if not brilliant, improvements in simulation, but they move on to other areas of science or engineering. The authors of this article have given a lifetime commitment to this work and will continue to do so. They understand their code and have programmed most of it themselves.

2. Materials and Methods—Criteria for Selecting Simulation Models

It is probably wise to distinguish between a *simulation model* and a *computational prototype*. A model usually simplifies the real world. It does not need to be a virtual representation that can replace the original system in all its dimensionality and all its complexity. It is often a caricature of a phenomenon or feature of interest, specific to a question. Finally, and perhaps most importantly, a model has the potential of being predictive of what cannot be observed or measured. In that sense, a model is highly educational. A computational prototype can have all these features, but it attempts to include most of the structural detail with equal weighting so that a physical implementation is possible. It does not reduce complexity non-uniformly to feature a specific phenomenon.

Some preliminary considerations are appropriate for selecting a simulation model. How accurate do the results need to be to answer the question? Sometimes ratios of variables are good enough, such as ratios of impedances, ratios of physical dimensions, or ratios of energies. Much of current simulation is based on imaging, from which ratios of physical dimensions are much easier to obtain than absolute dimensions. Further, if the simulation contains a parameter with questionable accuracy, a sensitivity analysis should be done on this parameter to avoid contamination of the results. For intelligent interpretation of simulation results, a simplified analytical analysis can be done to determine how variables relate to each based on dimensional analysis. For this reason, low-dimensional models are insightful for guiding high-dimensional computational prototypes.

All digital models require decisions about temporal and spatial resolution. The Courant Criteria for time and space discretization are useful when wave propagation in elastic media (air or tissue) is involved. In bounded media, modes of vibration are defined by standing waves, which in turn are described by wavelengths. The Courant Criteria are:

$$\Delta x < \lambda/20 \quad \text{for spatial accuracy} \quad (1)$$

$$\Delta t < \Delta x/2c \quad \text{for stability} \quad (2)$$

The $\lambda/20$ spatial discretization is considered an intermediate mesh for quantifying a full period of a sinusoid, while $\lambda/10$ is considered a course mesh and $\lambda/30$ a fine mesh, e.g., [1]. The $\lambda/20$ rule provides on the order of 1% accuracy in sinusoidal wave displacement at all points. For example, if only one half-wavelength is of interest in vocal fold vibration in both the anterior-posterior and inferior-superior directions [known as the (1,1) normal mode on the surface], then 10 sections, or 10 finite elements, are needed for sufficient accuracy in the anterior-posterior direction. If the elastic wave propagates at 100 cm/s in the vertical direction and the vocal fold thickness is 0.5 cm, the sampling interval Δt for stability

is 0.00025 s. Mesh and time-sampling independence, often mandated in computational prototypes, can sometimes be relaxed if approximate solutions are sufficient. This is the case with all low-dimensional vocal fold models.

Another example is spatial discretization in vocal tract wave propagation modeling. The number of sections needed depends on the number of modes (formants) of interest. Figure 1 shows that 15 sections are sufficient for the first formant, while at least 40–50 sections are needed for an equivalent accuracy for the fourth formant.

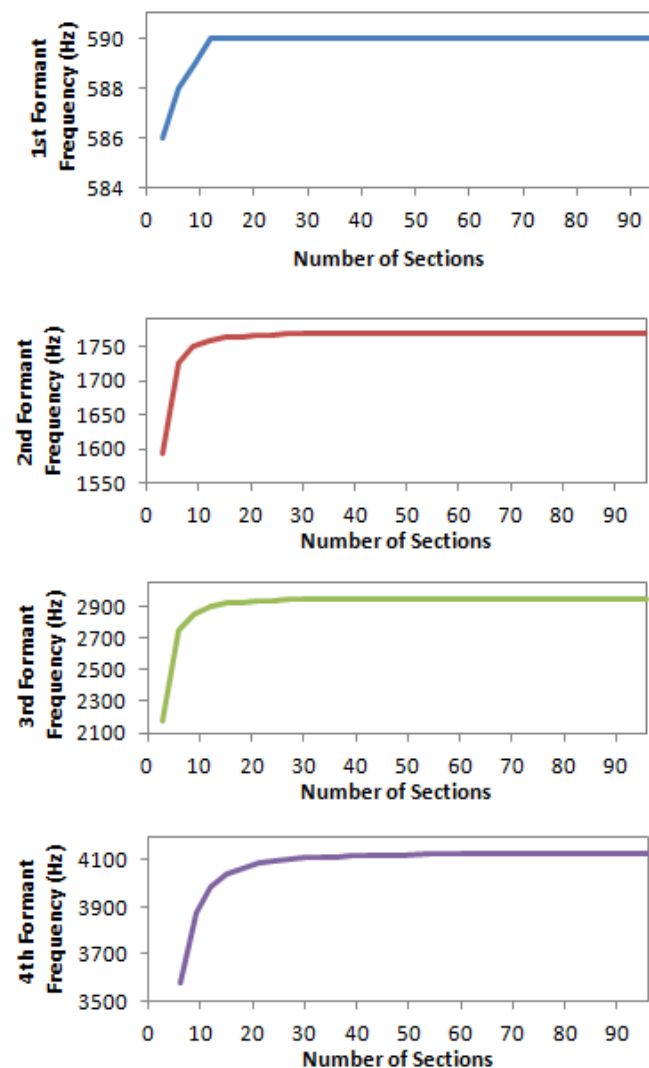


Figure 1. Accuracy of formant frequency calculation with number of sections (spatial discretization) along the vocal tract [2].

Airflow through the glottis is characterized by some fundamental dimensionless numbers [3,4]. The *Strouhal number* (St) pertains to oscillatory flows. St is the ratio of the time Δt for a flow particle to travel along the channel length to the period T of the oscillation, and is a measure of the unsteadiness of the flow in that channel:

$$St = \frac{\Delta t}{T} = \frac{f_o H}{v}, \quad (3)$$

where f_o is the fundamental frequency of the oscillation, H is the length of the glottal channel and v is the flow speed. Typical values for the vocal folds are in the range 0.01 – $0.02 \ll 1$, which means that the oscillation is slow, compared to the flow velocity. During the time a particle travels along the glottal channel, the channel dimensions have barely changed;

thus, the time-varying flow may often be analyzed as a succession of steady-states. The *Reynolds number* (Re) is the ratio of inertial to viscous forces in the fluid, and is given by

$$\text{Re} = \frac{\rho v h}{\mu}, \quad (4)$$

where ρ is the air density, v is the air particle velocity, h is the separation between the vocal folds, and μ is the air viscosity. A typical value for the vocal folds is $\text{Re} \approx 1800 \gg 1$, which implies that the bulk of the glottal airflow is determined by inertial forces (including turbulence), and any viscous drag is limited to a thin layer in the vicinity of the tissue walls. Finally, the *Mach number* (Ma) is the ratio of inertial to elastic forces in the fluid, given by

$$\text{Ma} = \frac{v}{c}, \quad (5)$$

where c is the speed of sound. Typical values for the vocal folds are $\text{Ma} < 0.1$; values of $\text{Ma} < 0.3$ imply that any compressibility effects on the flow in the glottis may be disregarded.

Fundamental frequency of oscillation is either explicit or implicit in all the tissue and air dynamic characteristics listed above in Equations (1)–(5). When fundamental frequency exceeds about 500 Hz, as in singing, calling, infant cry, or many animal vocalizations, some of the approximations do not hold. Nevertheless, low-dimensional modeling is still useful to capture many salient features.

3. Results

3.1. Low-Dimensional Airflow and Intraglottal Pressure Modeling

With relatively low Strouhal, Reynolds, and Mach numbers, the glottal airflow may be considered quasi-steady, inviscid, and incompressible. The exception is in the brief periods near glottal closure, in which both the air particle velocity v and the vocal fold separation h decrease, becoming zero at full glottal closure. Then, the Reynolds number decreases and the fluid becomes viscous [5]. Further, the Strouhal number increases and the assumption of a quasi-steady flow is not valid. Nevertheless, those periods are brief compared to the full glottal cycle and an empirical correction may be possible. Another relevant parameter is the ratio of vocal fold separation to vocal fold length, $h/L \ll 1$, which implies that the flow within the glottal channel may be considered two-dimensional. However, this reduced dimensionality applies only to a simple mode of vocal fold vibration that offers a single flow channel. When islands of partial vocal fold contact occur, the airflow becomes highly complex and multi-directional, as illustrated later.

Under the above assumptions and simplifications, single-channel glottal airflow is predictable by a modified Bernoulli's equation that includes flow separation from the glottal walls and vortex formation for a divergent glottis. An expression for the mean glottal pressure p_g can then be written in the form

$$p_g = p_e + \frac{p_s - p_e}{k_t} \left(1 - \frac{a_2}{a_1} \right), \quad (6)$$

where p_e is the supraglottal pressure (at the epiglottis), p_s is the subglottal pressure (at the upper end of the trachea), k_t is a transglottal pressure coefficient that includes the effects of flow separation and vorticity. The area a_1 is the glottal entry area and a_2 is the flow separation area for a divergent glottis. Generally, a_2 is assumed to be the exit area for a convergent glottis [6]. The supraglottal pressure p_e can further be approximated as

$$p_e = I dU/dt, \quad (7)$$

where I is the vocal tract acoustic inertance and dU/dt is the derivative of the glottal exit volume flow. The above analytic equations contain two independent mechanisms of driving pressure on the vocal folds for self-sustained vocal fold oscillation, one which

is independent of the airway (second term in Equation (6)), and one which is critically dependent on the airway (first term in Equations (6) and (7)). The relative contributions of these two mechanisms, labeled the mucosal wave (MW) mechanism and the supraglottal inertance (SI) mechanism, have recently been reported [7].

3.2. Low-Dimensional Tissue Modeling

Motion of the vocal folds may be considered as a composition of two principal motions: an oscillation in the lateral direction (perpendicular to the airflow), which opens and closes the glottis, plus a surface wave that propagates in the direction of the airflow [6,7]. The first component may be described by a second order lumped representation of a mechanical oscillator:

$$m\ddot{x} + h(x)\dot{x} + kx = p_g(x, \dot{x}) \tag{8}$$

where m is the vocal fold mass per unit area of the medial surface, k is its stiffness, x is the tissue displacement from its rest position, $h(x)$ is a positive function that accounts for all energy dissipation effects due to tissue deformation, collision between the opposite vocal folds and other factors, and $p_g(x, \dot{x})$ is the glottal pressure. The surface-wave component may be added through a simple kinematic description in d’Alembert’s form

$$\zeta(y, t) = x\left(t - \frac{y}{c}\right) \tag{9}$$

where ζ is the lateral displacement of the tissues at any position y along the glottal channel, measured from its midpoint, and c is the surface wave speed [7,8]. The Van der Pol model is obtained by combining the above equations, assuming that the approximation $x(t - y/c) \approx x(t) - (y/c)\dot{x}(t)$ holds, and that the supraglottal pressure p_e is negligible. Then, we obtain

$$m\ddot{x} + h(x)\dot{x} + kx = \frac{p_s}{v_{\text{char}}}\dot{x} \tag{10}$$

where v_{char} is a characteristic glottal coefficient [9]. Defining further a quadratic energy dissipation function $h(x) = r + sx^2$, and using a suitable transformation of variables, a van der Pol’s type equation is obtained [10]:

$$\ddot{x} - \beta(\lambda - x^2)\dot{x} + x = 0 \tag{11}$$

While this equation does not include source-filter interaction, it allows for the application of an extensive theory already available on van der Pol oscillators, e.g., [11] that partially applies to phonation. In fact, this well-known equation has been applied to characterize a number of systems in case of human physiology, from the heartbeat [12] to Parkinsonian tremor [13], and also to the vocalization of songbirds, e.g., [14]. For instance, the theory tells us that a Hopf bifurcation occurs at $\lambda = 0$, which implies $p_s = rv_{\text{char}}$. This value of p_s is known as the phonation threshold pressure and it represents a measure of the effort required to produce voice [15]. Depending on the shape of the dissipation function $h(x)$, the bifurcation may assume the subcritical form and cause an oscillation hysteresis phenomenon [16] with different threshold values for oscillation onset vs. offset. However, the assumption that the supraglottal pressure is negligible does not hold if the larynx canal (the epilarynx tube) is narrow and the mucosal wave velocity is large [7]. Then, self-sustained vocal fold oscillation is also dependent on the airway configuration, also known as source-filter interaction.

Figure 2 shows a bifurcation diagram for voice onset and voice offset based on van der Pol’s equation. The predicted hysteresis is an important finding in vocal fold mechanics. It guides further exploration with higher dimensional models that include the fluid transport and wave propagation in the entire airway.

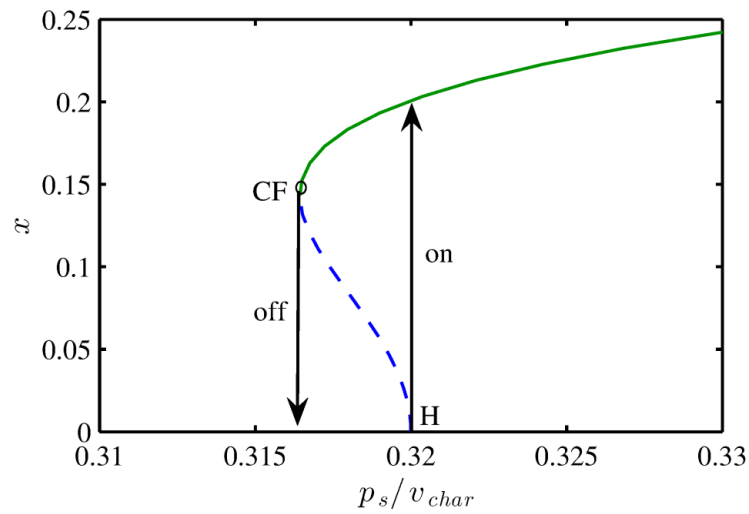


Figure 2. Bifurcation diagram for $h(x) = r + sx^4$, with $r = 0.32$, $s = 64$. The “on” arrow shows the oscillation onset at a subcritical Hopf bifurcation (H), whereas the “off” arrow shows the oscillation offset at a cyclic fold bifurcation [17].

Asymmetries between the right and left vocal folds may be simulated by considering a pair of coupled van der Pol oscillator of the form in Equation (10), e.g.,

$$\begin{aligned} \ddot{x} - \beta(\lambda - x^2)\dot{x} + (1 - \Delta)x &= \alpha(\dot{x}' - \dot{x}) \\ \ddot{x}' - \beta(\lambda - x'^2)\dot{x}' + (1 + \Delta)x' &= \alpha(\dot{x} - \dot{x}') \end{aligned} \tag{12}$$

where x, x' are the displacements of the right and left vocal folds, respectively, $-1 < \Delta < 1$ is a stiffness asymmetry term, and α is an aerodynamic coupling coefficient [18].

Figure 3 shows a bifurcation diagram. In the upper region of curve ABE, a stable oscillation exists in which both the right and left vocal folds oscillate with the same amplitude and a phase difference. The threshold for this regime (curve ABE) increases with $|\Delta|$, which implies that a larger subglottal pressure is required at larger asymmetries in order to achieve synchronization. In region EBD, both folds also oscillate, but at an $n:m$ phase entrainment regime other than 1:1.

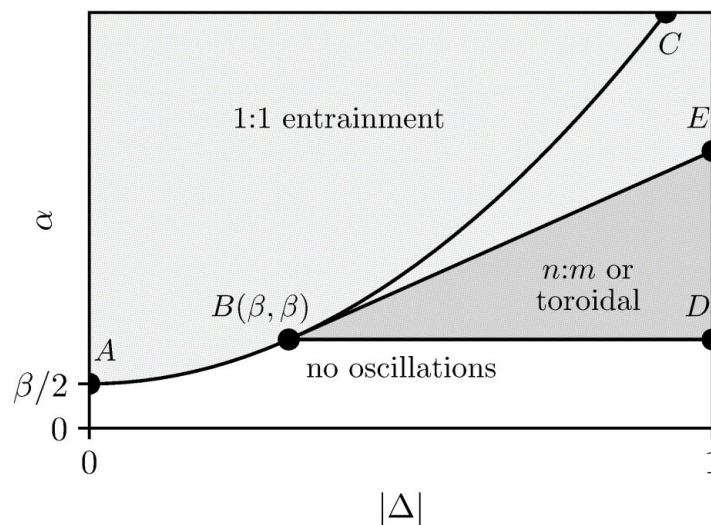


Figure 3. Bifurcation diagram and regions of oscillation. Curve ABC: Hopf bifurcation, curve BE: a $1/4 \ 3 = 2$, saddle-node bifurcation between limit cycles, line BD: double Hopf bifurcation [18].

Figure 4 shows plots of the oscillation for the 1:1 and 3:3 synchronization regimes. It is interesting to note that the synchronized oscillation may be put into Adler’s equation for phase dynamics [19]

$$\dot{\phi} = \Delta - \alpha \sin \phi \tag{13}$$

where ϕ is the phase difference between the right and left oscillators. This is a common model for circadian clocks in biology, and it is curious that the vocal folds follow the same dynamics. When $\Delta = 0$, the right and left folds oscillate with the same phase (i.e., $\phi = 0$). When Δ increases, ϕ also increases, and as the right vocal fold becomes more relaxed (or stiffer) than the left one, the oscillation of the right fold becomes delayed (or advanced) in relation to the left one.

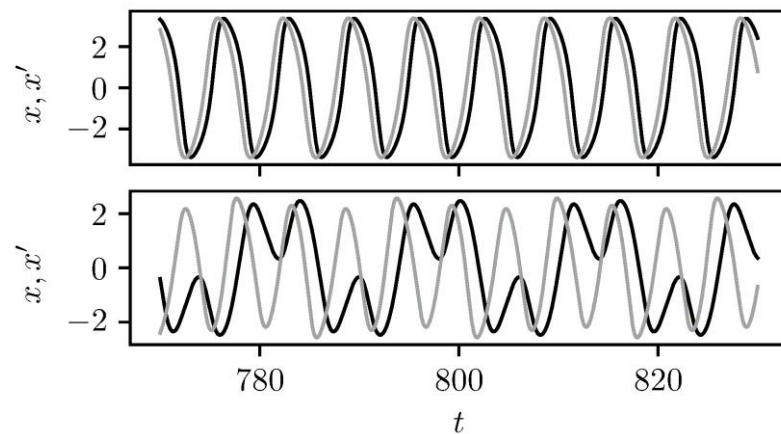


Figure 4. Right (black curve) and left (gray curve) oscillations in the 1:1 (top) and 3:3 synchronization regimes [18].

3.3. Low-Dimensional Acoustics

Finally, the acoustic subsystem may be added to the above models by regarding the vocal tract as a straight tube with resonant frequencies determined by wave propagation. In Equations (6) and (7), the supraglottal pressure p_e was expressed as a lumped-element inertance multiplied by a flow derivative. The inertance is frequency-dependent, however, and results from standing waves in the vocal tract. There is also a frequency-dependent resistance that dissipates energy. Using a partial wave approach to the solution of the wave equation, the acoustic pressure anywhere in the tube has two components: a forward traveling wave, from the glottis to the mouth, given by

$$p_e^+ = p_e^- + \left(\frac{\rho_0 c}{A}\right) u \tag{14}$$

where ρ_0 is the unperturbed air density, c is the sound speed, A is the area of the epilottis, and p_e^- is a backward travelling wave component, given by

$$p_e^- = -\gamma p_e^+(t - \zeta). \tag{15}$$

Here, γ is an attenuation factor for energy dissipation in the vocal tract, and ζ is the time delay for the acoustic wave to travel forth and back the vocal tract tube [20]. With some manipulation of the equations, we obtain

$$m\ddot{x} + h(x)\dot{x} + kx = \frac{p_s}{v_{char}}\dot{x} + \delta\bar{v} \left[x(t) + \sum_{n=1}^{\infty} (-\gamma)^n x(t - n\zeta) \right] \tag{16}$$

where δ is an acoustic coupling coefficient and \bar{v} is the mean particle velocity.

From the solution of Equation (16), the diagram shown in Figure 5 is obtained. It shows the oscillation frequency vs. the acoustic coupling coefficient and the natural frequency

$\sqrt{k/m}$. The figure becomes the classical shape of a cusp catastrophe. Suppose that the case of a large coupling (e.g., $\delta \approx 2$) is modeled, and that the vocal folds increase their natural frequency. Then, Figure 5 shows that the oscillation frequency will move from the lower part of the surface to the upper part, crossing the fold.

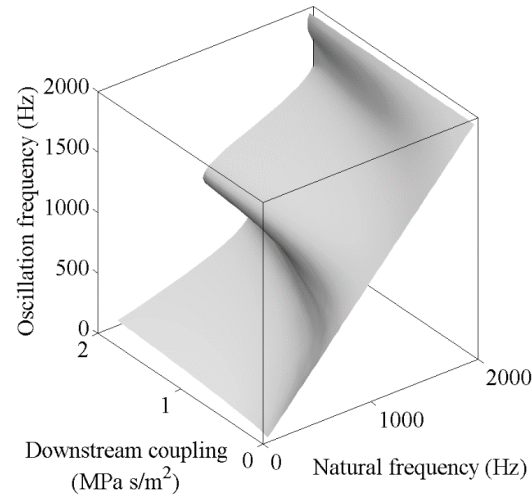


Figure 5. Cusp catastrophe from Equation (16) [20].

Figure 6 further illustrates the case. Crossing the fold in the surface will produce a frequency jump. If, instead of crossing the fold, the coupling is reduced (e.g., $\delta \approx 0$), the vocal folds will follow a path around the cusp and the jump will not occur.

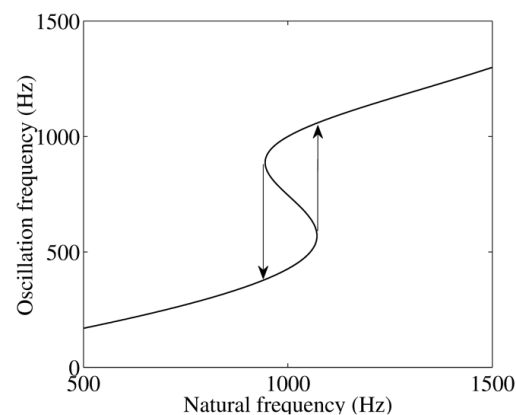


Figure 6. Oscillation frequency when moving across the fold in Figure 5. The arrows indicate frequency jumps, forming a hysteresis loop [20].

In summary, low-dimensional models are a simplified abstraction of a system which reduces it to its most fundamental mechanisms while keeping its main qualitative behavior. The vocal folds may be regarded as a flow-induced mechanical oscillator that generates an acoustical wave. Then, the overall system may be expressed in terms of three interacting sub-systems, namely, the glottal airflow, the vocal fold tissue mechanics, and the acoustic wave propagation [21].

3.4. High-Dimensional Simulations with Pre-Computations

High-dimensional finite-element simulations are becoming ever more feasible to solve the entire production-perception chain of events with few simplifying assumptions [22–25]. Hundreds to thousands of spatial elements are used to represent tissue and air continua. Execution is generally on arrays of supercomputers. If both air and tissue dynamics are three-dimensional across the entire airway and no simplifying assumptions are made, a

few cycles of oscillation may take days to compute. However, exploration with computer simulation often calls for hundreds, if not thousands, of repetitions with incremental steps of a parameter. This is especially the case if optimization of an output is the goal. Computation speed is therefore still an issue, even if arrays of supercomputers are used. Practitioners in our field may not have access to supercomputer facilities, or they prefer to work on laptops or desktops. That brings up the question of pre-computation for high-cost portions of the simulation that are repeated over and over again.

Some pre-calculations can be made in the next generation with the use of computational learning: (1) Acoustic targets can be pre-determined from perceptual targets, (2) Biomechanical parameters can be pre-determined from acoustic targets, (3) A large variety of feed-forward biomechanic-to-acoustic mappings can be pre-calculated, (4) Deep learning with acoustic and somatosensory feedback can be implemented to predict novel acoustic outputs from a training set, and (5) Vocal fold driving pressures in a complex glottis can be predicted with a neural network from a pre-calculated data base. These driving pressures are the costliest in high-dimensional simulations with high fidelity.

As an example, a large variety of normal mode surface vibrations with different contact islands and diverse flow channels (confluent, di-fluent, convergent, divergent) can be used to generate a data base. Deep learning can then be applied to predict a new pressure profile from the pre-computed data base. Figure 7 shows one of many possible cases. Each of the four rectangles represents the glottal surface, left to right being inferior to superior (direction of airflow) and bottom to top being posterior to anterior. In left to right sequence, we show glottal contact (two black areas), glottal width, calculated glottal pressure with a high-fidelity immersed boundary solution, and measured glottal pressure from a scaled-up physical model. This mapping between a glottal shape and the corresponding driving pressures can be pre-calculated for hundreds of glottal shapes, from which a neural network can then predict the pressure pattern for a novel shape. It is evident from Figure 7 that a single channel Bernoulli approach cannot capture the pressure profile on the glottal surfaces. Airflow is multi-directional and therefore the pressure gradients are also multi-directional.

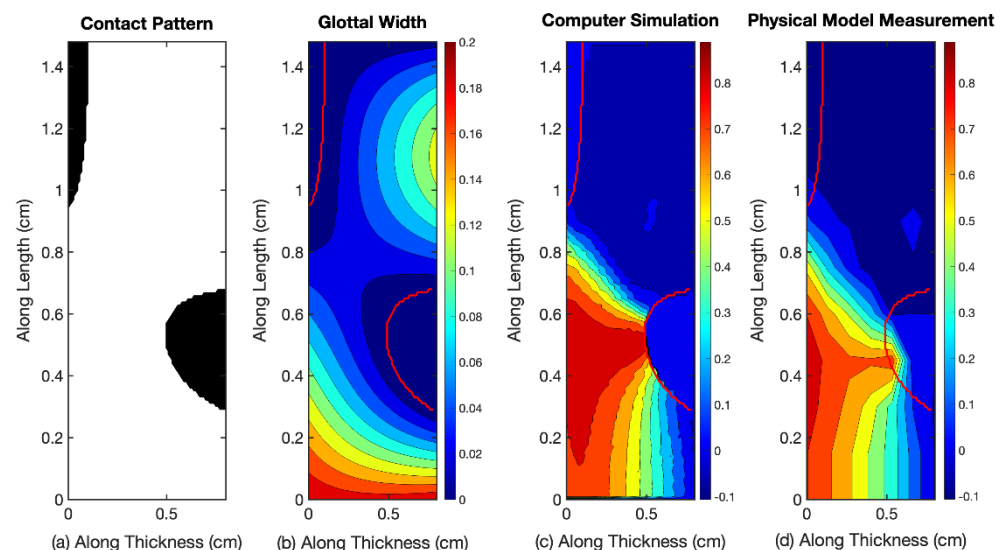


Figure 7. Four variables depicted in a rectangular glottis with inferior (left), superior (right), posterior (bottom) and anterior (top) boundaries. (a) vocal fold contact regions, (b) glottal width, (c) calculated surface pressure with high-dimensional computation, and (d) measured pressure with an up-scaled static model [26].

All machine learning approaches require training data that are likely to occur in phonation. This is a bit of a “lifting by the bootstrap” approach. A priori, we do not know which normal modes of vibration will be excited by the predicted pressures, and we do not

know how often these modes exist in self-sustained oscillation. With a large training set and assigned weights, the neural network will likely converge on the realistic cases that occur in self-sustained oscillation.

4. Discussion: Model Validation

It has been difficult to identify standard airway configurations and standard vocal fold configurations for both measurement and computation. Once test configurations are standardized, the validity of the solution is still an issue. An approach might be: (1) Compare computed formant frequencies and bandwidths to standard data bases for given set of vocal tract configurations, (2) Compare computed contact areas, glottal areas, glottal flows to standard data bases for humans and animals, (3) Use asymptotic conditions (parameters going to zero or infinity) to find agreement with lower dimensional digital or analytic solutions.

5. Conclusions

Computer simulation has delivered on much of the promise to be a major research tool in voice and speech production. Both low-dimensional mass-spring models and high-dimensional finite element continuum models have contributed to advances in our understanding of the myoelastic-aerodynamic theory of phonation. A current challenge is to use high-dimensional simulations to reduce dimensionality and complexity with shortcuts that deliver 80–90% of the 3-D effects [24]. Investigators around the world benefit from some joint, yet compartmentalized, efforts to move the field forward. New investigators require a lifetime to duplicate everything that has gone before at multiple sites. A high priority is the development of some airway configurations against which a given simulator can be validated.

Author Contributions: Both authors have contributed substantially to all phases of the manuscript, as follows: Conceptualization, I.R.T. and J.C.L.; methodology, I.R.T. and J.C.L.; software, I.R.T. and J.C.L. independently; validation, I.R.T. and J.C.L.; formal analysis, I.R.T. and J.C.L. independently; investigation, I.R.T. and J.C.L.; resources, I.R.T. and J.C.L.; writing—original draft preparation, reviewing and editing, I.R.T. and J.C.L.; project administration, I.R.T. and J.C.L.; funding acquisition, I.R.T. and J.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: I.R. Titze was supported by grant number R01 DC 017998-03 from the National Institutes on Deafness and Other Communication Disorders (USA). J. C. Lucero was supported by FAPDF grant 00193.00001208/2021-22 and CNPq grant 302718/2021-4 (Brazil).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: An expanded data set on glottal pressure profiles will be available to the public at the National Center for Voice and Speech website (ncvs.org).

Acknowledgments: The authors wish to thank Angela Keeton for manuscript preparation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ning, D.Z.; Zang, J.; Liu, S.X.; Eatok Taylor, R.; Teng, B.; Taylor, P.H. Free-surface evolution and wave kinematics for nonlinear uni-directional focused wave groups. *Ocean Eng.* **2009**, *36*, 1126–1243. [[CrossRef](#)]
2. Titze, I.R.; Palaparthi, A.; Smith, S.L. Benchmarks for time-domain simulation of sound propagation in soft-walled airways: Steady configurations. *J. Acoust. Soc. Am.* **2014**, *128*, 828–838. [[CrossRef](#)] [[PubMed](#)]
3. Pelorson, X.; Hirschberg, A.; van Hassel, R.R.; Wijnands, A.P.J. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *J. Acoust. Soc. Am.* **1994**, *96*, 3416. [[CrossRef](#)]
4. Massey, B.S. *Mechanics of Fluids*, 9th ed.; Spon Press: New York, NY, USA, 2012; pp. 162–167, 328.
5. Sundström, E.; Oren, L.; Farbos de Luzan, C.; Khosla, S. Fluid-Structure Interaction Analysis of Aerodynamic and Elasticity Forces During Vocal Fold Vibration. *J. Voice* **2022**. [[CrossRef](#)] [[PubMed](#)]

6. Titze, I.R. The physics of small-amplitude oscillation of the vocal folds. *J. Acoust. Soc. Am.* **1988**, *83*, 1536–1552. [[CrossRef](#)] [[PubMed](#)]
7. Titze, I.R. Can the vocal folds oscillate with a minimal mucosal wave? *JASA-EL* **2022**, *2*, 105201. [[CrossRef](#)] [[PubMed](#)]
8. Hirano, M.; Kakita, Y.; Kawasaki, H.; Gould, W.J.; Lambiase, A. Data from high speed motion picture studies. In *Vocal fold Physiology*; Hirano, M., Stevens, K.N., Eds.; University of Tokyo Press: Tokyo, Japan, 1981; pp. 85–91.
9. Arneodo, E.M.; Mindlin, G.B. Source-tract coupling in birdsong production. *Phys. Rev. E* **2009**, *79*, 061921. [[CrossRef](#)] [[PubMed](#)]
10. Lucero, J.C.; Schoentgen, J. Modeling vocal fold asymmetries with coupled van der Pol oscillators. *Proc. Mtgs. Acoust.* **2013**, *19*, 060165. [[CrossRef](#)]
11. Grimshaw, R. *Nonlinear Ordinary Differential Equations*; CRC Press: Boca Raton, FL, USA, 1991.
12. van der Pol, B.; van der Mark, J. The heartbeat considered as a relaxation oscillation, and an electrical model of the heart. *Philos. Mag.* **1928**, *6*, 763–775. [[CrossRef](#)]
13. Beuter, A.; Edwards, R.; Titcombe, M. Data analysis and mathematical modeling of human tremor. In *Nonlinear Dynamics in Physiology and Medicine*; Beuter, A., Glass, M., Mackey, M.C., Titcombe, M.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2003.
14. Laje, R.; Mindlin, G.B. Modeling source-source and source-filter acoustic interaction in birdsong. *Phys. Rev. E* **2005**, *72*, 036218. [[CrossRef](#)] [[PubMed](#)]
15. Titze, I.R. Phonation threshold pressure: A missing link in glottal aerodynamics. *J. Acoust. Soc. Am.* **1992**, *91*, 2926–2935. [[CrossRef](#)] [[PubMed](#)]
16. Appleton, E.V.; van der Pol, B. On a type of oscillation-hysteresis in a simple triode generator. *Philos. Mag.* **1922**, *43*, 177–193. [[CrossRef](#)]
17. Lucero, J.C. Bifurcations and limit cycles in a model for a vocal fold oscillator. *Commun. Math. Sci.* **2005**, *3*, 517–529. [[CrossRef](#)]
18. 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](#)]
19. Adler, R. A study of locking phenomena in oscillators. *Proc. IRE* **1946**, *34*, 351–357. [[CrossRef](#)]
20. Lucero, J.C.; Lourenço, K.G.; Hermant, N.; Van Hirtum, A.; Pelorson, X. Effect of source–tract acoustical coupling on the oscillation onset of the vocal folds. *J. Acoust. Soc. Am.* **2012**, *132*, 403–411. [[CrossRef](#)] [[PubMed](#)]
21. Titze, I.R. *The Myoelastic-Aerodynamic Theory of Phonation*; The National Center for Voice and Speech: Denver, CO, USA; Iowa City, IA, USA, 2006.
22. Alipour, F.; Berry, D.; Titze, I. A finite element model of vocal fold vibration. *J. Acoust. Soc. Am.* **2000**, *108*, 3003–3012. [[CrossRef](#)] [[PubMed](#)]
23. Zheng, X.; Mittal, R.; Xue, Q. Direct-numerical simulation of the glottal jet and vocal-fold dynamics in a three-dimensional laryngeal model. *J. Acoust. Soc. Am.* **2011**, *130*, 404. [[CrossRef](#)] [[PubMed](#)]
24. Švancara, P.; Horáček, J.; Hruža, V. FE Modelling of the Fluid-Structure-Acoustic Interaction for the Vocal Folds Self-Oscillation. In *Proceedings of the Vibration Problems ICOVP 2011*. Part of the Springer Proceedings in Physics Book Series; Springer Science + Business Media B.V.: Dordrecht, The Netherlands, 2011; Volume 139, pp. 801–807.
25. Zhang, Z. Characteristics of phonation onset in a two-layer vocal fold model. *J. Acoust. Soc. Am.* **2009**, *125*, 1091. [[CrossRef](#)] [[PubMed](#)]
26. Titze, I.; Maxfield, L.; Manternach, B.; Palaparathi, A.; Scherer, R.; Wang, X.; Zheng, X.; Qian, X. Comparison of Calculated and Measured Pressure Profiles in Complex Glottal Geometries. *JASA* 2022. *submitted-under review.*