Benchmarks for time-domain simulation of sound propagation in soft-walled airways: Steady configurations

Ingo R. Titze,^{a)} Anil Palaparthi,^{b)} and Simeon L. Smith^{b)}

National Center for Voice and Speech, 136 South Main Street, Suite 320, Salt Lake City, Utah 84101

(Received 7 November 2013; revised 1 October 2014; accepted 8 October 2014)

Time-domain computer simulation of sound production in airways is a widely used tool, both for research and synthetic speech production technology. Speed of computation is generally the rationale for one-dimensional approaches to sound propagation and radiation. Transmission line and wave-reflection (scattering) algorithms are used to produce formant frequencies and bandwidths for arbitrarily shaped airways. Some benchmark graphs and tables are provided for formant frequencies and bandwidth calculations based on specific mathematical terms in the one-dimensional Navier–Stokes equation. Some rules are provided here for temporal and spatial discretization in terms of desired accuracy and stability of the solution. Kinetic losses, which have been difficult to quantify in frequency-domain simulations, are quantified here on the basis of the measurements of Scherer, Torkaman, Kucinski, and Afjeh [(2010). *J. Acoust. Soc. Am.* **128**(2), 828–838].

© 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4900563>]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.72.Ja [BHS]

Pages: 3249–3261

I. INTRODUCTION

Two of the earliest time-domain vocal tract simulations (Kelly and Lochbaum, 1962; Ishizaka and Flanagan, 1972) were able to capture the transmission and resonance properties of human airways with one-dimensional acoustic wave propagation. Kelly and Lochbaum used a partial-wave reflection approach, while Ishizaka and Flanagan used a transmission line “circuit” approach. In both cases, the vocal tract was discretized into serially-abutted cylindrical sections in which the cross sectional area remained constant. Variation in cross sectional area along the vocal tract axis was then handled with boundary conditions between sections. The wave reflection (or wave scattering) approach, described only as a computational algorithm in a short research note by Kelly and Lochbaum, was developed in great detail in the dissertations of Liljencrants (1985) and Story (1995). Zhang and Espy-Wilson (2004) and Birkholz (2005) have recently revived the transmission-line approach for high quality speech simulation in the time domain.

There are major advantages and disadvantages in the use of the two approaches. The transmission line “circuit” approach involves fewer assumptions with regard to mass transport, kinetic and viscous energy losses, and energy transferred to yielding walls. In other words, fewer terms are eliminated from the fundamental Navier–Stokes momentum equation and the mass transfer continuity equation before discretization takes place. The wave reflection approach begins with a highly idealized one-dimensional wave equation, which is then modified by adding correction terms that account for the original over-simplifications in energy loss. The major advantage of the wave reflection approach,

however, is its computational speed and stability. A large number of serial sections can be used with a relatively low sampling rate and a high degree of computational stability.

A question could be raised at the outset of this study about the merit of expending resources on one-dimensional approaches to vocal tract acoustics when three-dimensional approaches with complex geometry are now routine in computational fluid mechanics and aeroacoustics (Vampola *et al.*, 2008; Milenkovic *et al.*, 2010). The simple answer is that real-time (or near real-time) simulations on ordinary home, office, or hand-held computers is still a high priority. Furthermore, it is not yet clear which perceptually relevant aspects of 3-D vocal tract acoustics (Vampola *et al.*, 2013) cannot be captured with shortcuts in one-dimensional approaches. Until such clarity exists, side-by-side developments of 3-D approaches and simplified 1-D or 2-D approaches are defensible.

In this article, we review the basic assumptions made in the transmission-line “circuit” approach and how they relate to the Navier–Stokes equation. First, we benchmark the requirements of time and space discretization (Δt and Δz) for accurate representation of four formant frequencies. Second, we benchmark bandwidths in terms of viscous losses, radiation losses, and wall vibration losses by comparison to frequency-domain approaches used by Badin and Fant (1984) and Stevens (1998). Third, we introduce a new simulation of kinetic losses based on empirical data by Scherer *et al.* (2010). A subsequent paper will deal with gross dynamical changes of the vocal tract, such as rapid length change and local airway collapse.

II. REVIEW OF THEORY

In a reference book on the myoelastic-aerodynamic theory of phonation (Titze, 2006), Chap. 6 was devoted to the derivation of transmission line components from the

^{a)}Author to whom correspondence should be addressed. Also at: The Department of Communication Sciences and Disorders, the University of Iowa, Iowa City, Iowa 52242. Electronic mail: ingo-titze@uiowa.edu

^{b)}Also at: The University of Utah, Salt Lake City, Utah 84101.

Navier–Stokes momentum equation for non-steady compressible flow,

$$\rho \frac{dv}{dt} = -\nabla p + (\lambda + \mu) \nabla(\nabla \cdot \mathbf{v}) + \mu \nabla^2 \mathbf{v} \quad (1)$$

and the corresponding continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (2)$$

In the above equations, ρ is the air density, $\mathbf{v}$ is the air particle velocity vector, p is the pressure, λ is the Lamé compressibility of air, and μ is the air viscosity. The equations contain mass transport in that the vector $\mathbf{v}$ can have a steady component and an oscillatory component. On the right side of Eq. (1), the first term is the pressure gradient, the second term is the kinetic pressure (loss or gain, depending on whether the tube expands or contracts), and the third term is the viscous pressure loss. The term on the left side of Eq. (1) is the inertial term for non-steady flow. The continuity equation [Eq. (2)] describes air compliance and flow sources (if there are any). Wall compliance is introduced as a boundary condition (described later).

For one-dimensional spatial discretization in the z direction, the vocal tract was divided into cylindrical sections (tubelets) as shown in Fig. 1. A radial and an axial component of the air particle velocity vector were defined as w and v , such that in vector form the velocity was

$$\mathbf{v} = w\hat{r} + v\hat{z}. \quad (3)$$

The axial length of the section was Δz . The yielding wall was described by the radial displacement ξ and velocity $\dot{\xi}$ at the center of a wall depth d_w . The radius to the inside of the wall was r and the radius to the effective depth of vibration was R . Thus,

$$r = a + \xi, \quad (4)$$

$$R = a + \xi + d_w/2, \quad (5)$$

where a is the unstrained radius of the wall and d_w is the depth where vibration ceases (linearly).

FIG. 1. Cylindrical section, (a) side view and (b) axial view.

The air particle velocity gradients were linearized within the section as follows:

$$\frac{\partial v}{\partial z} = \frac{v_2 - v_1}{\Delta z}, \quad (6)$$

$$\frac{\partial w}{\partial r} = \frac{w_R - w_0}{R} = \frac{\dot{\xi} - 0}{R}, \quad (7)$$

where v_1 and v_2 are entry and exit particle velocity components in the axial directions, while w_R and w_0 are radial components that vary from zero at the center of the tubelet to $\dot{\xi}$ at the center of the wall. Figure 2 shows sketches of the combined velocity vector for an expanding and contracting wall.

Pressure gradients were also linearized within a section

$$\frac{\partial p}{\partial z} = \frac{p_2 - p_1}{\Delta z}, \quad (8)$$

$$\frac{\partial p}{\partial r} = \frac{p_w - p_c}{R}, \quad (9)$$

where p_1 and p_2 were inlet and outlet pressures, respectively, p_w was the pressure at the wall, and p_c was the pressure at the center of the tube. With these and a few other linearity approximations, the momentum equation became a pressure-drop equation in a circuit (Fig. 3),

$$p_1 - p_2 = (R_{v1} + R_{k1})v_1 + I_1\dot{v}_1 + (R_{v2} + R_{k2})v_2 + I_2\dot{v}_2 \quad (10)$$

and the continuity equation became a shunt equation in the same circuit,

$$\dot{p} = \frac{1}{C_a} \left(v_1 - v_2 - 2 \frac{\Delta z}{R} \dot{\xi} \right), \quad (11)$$

where

$$R_{v1} = \text{entry viscous resistance (to be described later)}, \quad (12)$$

$$R_{k1} = \text{entry kinetic resistance} = -\frac{\rho}{2} \left(v_1 + \frac{4 \Delta z}{3 R} \dot{\xi} \right), \quad (13)$$

$$I_1 = \text{entry inductance} = \rho \Delta z / 2, \quad (14)$$

$$R_{v2} = \text{exit viscous resistance (to be described later)}, \quad (15)$$

FIG. 2. Velocity profiles, (a) expanding wall and (b) contracting wall.

FIG. 3. Acoustic “circuit” based on momentum and continuity equations.

$$R_{k2} = \text{exit kinetic resistance} = +\frac{\rho}{2} \left(v_2 - \frac{4\Delta z}{3R} \dot{\xi} \right), \quad (16)$$

$$I_2 = \text{exit inductance} = \rho \Delta z / 2, \quad (17)$$

$$C_a = \text{air compliance} = \frac{\Delta z}{\rho c^2}. \quad (18)$$

We have deliberately chosen the symbol I for inductance (rather than L for inductance) to draw closer to mechanical engineering terminology. Also, C stands for compliance rather than capacitance. Equations (10) and (11) are first-order differential equations that were shown to be solvable by Runge–Kutta methods, but wall displacement needed to be included with an additional second-order equation,

$$M \ddot{\xi} + D \dot{\xi} + K \xi + S_1(\xi - \xi^-) + S_2(\xi - \xi^+) = p = \frac{p_1 + p_2}{2}, \quad (19)$$

where lumped-element parameters were expressed in terms of continuous media parameters, namely, tissue density ρ_w , tissue viscosity η_w , and tissue Young’s modulus E_w ,

$$M = \rho_w d_w \text{ (mass per unit area)}, \quad (20)$$

$$D = 2\eta_w / d_w \text{ (damping per unit area)}, \quad (21)$$

$$K = E_w / d_w \text{ (stiffness per unit area)}. \quad (22)$$

The superscript “ $-$ ” and “ $+$ ” in Eq. (19) refer to nearest neighbor (left and right) quantities in the spatial discretization. The same superscripts are used for the shear stiffnesses,

$$S_1 = (\mu_w^- + \mu_w)(d_w^- + d_w) / (\Delta z^- + \Delta z)^2, \quad (23)$$

$$S_2 = (\mu_w^+ + \mu_w)(d_w^+ + d_w) / (\Delta z^+ + \Delta z)^2. \quad (24)$$

Figure 4 shows the wall elements defined in the equations above. Springs perpendicular to the wall represent K , springs between sections represent S , and M represents a one-layer mass. The wall parameter values used here were taken from Ishizaka *et al.* (1975) and are listed in Table I.

The circuit of Fig. 3 was used by Ishizaka and Flanagan (1972) and more recently by Birkholz *et al.* (2006). Heat conduction losses did not fall out of the continuity equation [Eq. (2)], but can be added as another shunt resistance. They are, in general, the least important of the energy losses and are not explicitly formalized here.

Table I shows a complete list of physical constants used in the simulations to be described. We have used the centimeter-gram-second system of units because the geometry is

FIG. 4. A vocal tract with yielding walls, represented by masses and springs.

more suitable to this system. The numbers have uneven degrees of accuracy and confidence. Some are listed with many significant figures, but only because of the importance of benchmark comparison. Densities of air and tissue, viscosity of air, and velocity of sound are accurate to three significant figures, but other parameters related to wall properties and kinetic pressure losses (to be discussed later) have only single-digit significance or may vary over an order of magnitude across age, gender, and location along the airways.

A noteworthy difference between the equations written here and those used traditionally in circuit acoustics is that particle velocity is the field variable in our simulation rather than volume velocity (flow). There are two minor advantages in the use of particle velocity as a field variable. First, it is the variable in the Navier–Stokes solution, which makes it easier to compare simplified 1-D simulations to more complete 2-D and 3-D computations. Second, a potential mathematical singularity is removed by not dividing all circuit elements by A , the cross-sectional area of the tubelet. For example, the inertive elements I_1 and I_2 in Fig. 3, which appear in the denominator when the differential equations are solved by Runge–Kutta methods, are $\rho \Delta z$ instead of $\rho \Delta z / A$. Also, kinetic losses vary inversely with A instead of A^2 when particle velocity is used as the dependent variable. Division by cross sectional area A in the vocal tract becomes a serious computational problem when the airway becomes occluded. As a further example, the radiation resistance and inductance vary inversely with lip radius when written as a

TABLE I. Parameters used for the NCVS simulations.

Parameter	Value	Unit
Min. allowed radius	0.008	cm
Density of Air	1.14×10^{-3}	g/cm ³
Viscosity of Air	1.86×10^{-4}	dyn-s/cm ³
Velocity of Sound	35 400	cm/s
Pressure recovery coefficient (α_m)	0.5	
Entry loss coefficient (γ_m)	1.0	
Complete detachment Area ratio (β_d)	1.2	
Maximum entry loss Area ratio (β_c)	0.6	
Mass/wall area, M	1.5	g/cm ²
Stiffness/wall area, K	66 667	dyn/cm ³
Damping/wall area, D	800	dyn-s/cm ³
Tissue density of wall	1.04	g/cm ³
Depth of the wall, d_w	1.44	cm
Tissue viscosity, η_w	577	dyn-s/cm ²
Tissue Young’s modulus, E_w	9.6×10^4	dyn/cm ²

pressure-flow ratio (Flanagan, 1965, pp. 32–36). As a pressure-velocity ratio, they are

$$R_r = 128\rho c/9\pi^2 \text{ Pa s/m}, \quad (25)$$

$$I_r = 8\rho r/3\pi \text{ Pa s}^2/\text{m}, \quad (26)$$

where r is the mouth radius. These quantities do not contain a division by zero when mouth area goes to zero. A disadvantage, however, is that typical acoustic power and acoustic impedance calculations (pressure $\times$ flow or pressure/flow) have to be reformulated in terms of mechanical force (pressure $\times$ area) and particle acceleration $\dot{v}$. For details of how the differential equations reviewed above are solved, the reader is referred to Titze (2006), Chap. 6, pp. 311–313.

III. BENCHMARK FOR TIME AND SPACE DISCRETIZATION IN A UNIFORM CLOSED-OPEN TUBE

Consider a rigid uniform tube, closed at one end and open on the other. The resonant frequencies are known analytically from basic acoustic theory (Fant, 1960, p. 39),

$$F_n = (2n - 1) \frac{c}{4L}, \quad (27)$$

where c is the sound velocity and L is the length of the tube. Choosing $c = 35\,400$ cm/s for warm, humid air in human airways (Stevens, 1998, p. 138) and $L = 15$ cm as an approximate mean supraglottal length between males and females (Stevens claims 14.1 cm for females and 16.9 cm for males on average), the first four formant frequencies are 590, 1770, 2950, and 4130 Hz. Equation (27) assumes no energy losses in the air or the wall of the tube, nor from sound radiation out of the tube. In other words, the resonances have zero bandwidth. As a first benchmark for time-domain simulation, lossless wave propagation for the uniform tube was simulated with variable number of N sections, so that $\Delta z = 15/N$ cm. All resistances in the foregoing equations were set to zero and the wall displacement ξ was also set to zero. A step function for ν_1 at the glottal end was used as an excitation and four formant frequencies were extracted from the natural response. The results are shown in Fig. 5. This first benchmark establishes the time and space sampling necessary for computational stability and accuracy. Note that formant frequencies reach their theoretical values asymptotically with 15–40 sections. It is clear that more sections are required for higher formant frequencies to reach their theoretical values.

A. Stability rule of thumb

A stability rule of thumb (Taflove and Hagness, 2005, p. 49) is

$$\Delta t \leq \frac{\Delta z}{2c}, \quad (28)$$

$$\leq \frac{L}{2cN}. \quad (29)$$

For the transmission line approach outlined in Sec. II and detailed in Titze (2006), Chap. 6, we chose a sampling

FIG. 5. (Color online) The first four formant frequencies of a lossless closed-open tube of 15 cm length, extracted from a step response of a digital transmission line “circuit” of N -sections.

period $\Delta t = 1/(176\,400)$, which meant that the lower limit for Δz was 0.4 cm (corresponding to an upper limit for N of 37 sections). A fourth order Runge–Kutta method effectively increased the sampling rate by a factor of 2, which predicts a conservative upper limit of $N = 74$ and a lower limit of $\Delta z = 0.2$ cm. Our simulation was stable up to $N = 96$ ($\Delta z = 0.16$ cm) for this simple uniform tube.

With the wave reflection approach [Liljencrants, 1985; Story, 1995; Titze, 2006 (Chap. 6)], the choice of Δt is much simpler. The exact relation $\Delta t = \Delta z/c$ can be used, which provides perfect computational stability because arrival and departure of incident, reflected, and transmitted waves are in lock-step with the computation interval. Furthermore, the computation can be down-sampled by a factor of 2 ($\Delta t = 2\Delta z/c$) because odd and even section interface calculations occur only at odd or even time steps [Titze, 2006 (Chap. 6)]. Thus, speed of computation can generally be four times faster with the wave reflection algorithm than the transmission line algorithm when no losses are included. When losses are included in the wave reflection algorithm (in the form of correction factors), the computation time increases.

B. Accuracy rule of thumb

A 0.5% accuracy rule of thumb (Taflove and Hagness, 2005, p. 45) for simulation of traveling and standing waves is

$$\Delta z < \frac{\lambda}{20}, \quad (30)$$

where λ is the wavelength in the tube. For the first formant frequency of a 15 cm quarter-wave tube, the wavelength is $4L$ (60 cm), which means that Δz should not be greater than 3.0 cm, or N should not be less than 5 to capture the standing wave pattern accurately. Figure 5 shows that, with $N=5$, F_1 is underestimated by 3 Hz (587 instead of 590), or about 0.5%. Higher formants require more sections for similar accuracy. For F_4 , for example, the wavelength of a seven quarter wavelength tube is $4L/7$ (8.57 cm), which means that Δz should be less than 0.43 cm, or N should be greater than 35. In Fig. 5, the analytic value of $F_4=4130$ Hz is reached within 17 Hz with 35 sections (calculated value 4113), an error a little less than 0.5%.

IV. VISCOUS, RADIATION, AND WALL LOSSES IN A UNIFORM TUBE AND THREE VOWEL SHAPES

A. Viscous losses

It has long been known that viscous losses in the vocal tract are frequency-dependent (Fant, 1960; Flanagan, 1965). An effective boundary layer is formed around the perimeter S of the wall whose depth (sometimes referred to as skin depth) gets smaller with increasing frequency. The relation

$$R_v = \frac{S\Delta z}{A} \sqrt{\frac{\omega\rho\mu}{2}} \quad (31)$$

applies to a ratio between driving pressure and mean particle velocity across the tube [Fant, 1960 (p. 32); Flanagan, 1965 (p. 27)]. In the above, Δz is the section length, A is the above-mentioned cross-sectional area, ρ is the warm air density, and μ is the air viscosity. Unfortunately, this relation does not asymptote to any steady-flow (zero frequency) relation obtained by solving the Laplace equation [Eq. (1), first and last term on the right],

$$\nabla^2 v = \frac{1}{\mu} \frac{\partial p}{\partial z}. \quad (32)$$

For an elliptical cross-section with a minor radius a and a major radius b , the Laplace equation yields the following ratio between pressure and mean particle velocity:

$$R_v = 4\mu\Delta z \frac{a^2 + b^2}{a^2 b^2}. \quad (33)$$

Many attempts have been made to use some version of Eq. (33) for viscous losses in the vocal tract, but the Laplace equation grossly underestimates the viscous resistance in most cases of practical interest in speech.

If Eq. (31) is to be used in time domain simulations, the frequency ω is not explicitly known. A reasonable

approximation to viscous losses is obtainable by choosing a constant frequency of 1000 Hz in Eq. (31) (Wakita and Fant, 1978). As a logarithmic mean between 100 Hz and 10 000 Hz, the value 1000 Hz overestimates low frequency losses and underestimates high frequency losses, but symmetrically in a logarithmic speech spectrum. Formant bandwidths are affected by only a few Hz with this approximation, as will be shown.

To make Eq. (31) applicable to cross-sections that vary all the way from a circle to a very flat ellipse (nearly rectangular), the perimeter S and the area A are written as

$$S \approx \pi[3(a+b) - \sqrt{(3a+b)(a+3b)}], \quad (34)$$

$$A = \pi ab. \quad (35)$$

This transition from circular to flat is especially applicable in the subglottal entry and supraglottal exit regions in the larynx, as well as in lip apertures.

B. Losses in a uniform tube

A three-way comparison between formant frequencies and bandwidths of a 15 cm long tube (uniform 3 cm² cross section) was conducted to obtain a benchmark. The tube was closed on the glottal end and allowed to radiate at the lip end. The sources for bandwidths were Badin and Fant (1984), Stevens (1998), and Ishizaka *et al.* (1975). To gain confidence in our custom-developed time-domain software developed at the National Center for Voice and Speech (NCVS), an identical lumped element circuit was created using the Simulink Simscape library under MATLAB (The Mathworks, 2014) for the uniform tube (vowel shapes were too tedious to program under Simulink). Whenever possible, formant frequencies and bandwidths were quantified for each loss separately, and for all losses combined. The transmission-line circuits were created with 44 serial sections. This number of sections falls within the stability and accuracy criteria suggested above for four formants.

1. Influence of equation solvers

Figure 6(a) shows an $N=44$ section circuit built using the Simulink Simscape library. Air particle velocity was made equivalent to current and pressure equivalent to voltage in the lumped element circuit model. A step current source connected to the first section was used as an air particle velocity input to obtain the step response of the vocal tract system. Figure 6(b) shows the current source system configuration, and Fig. 6(c) shows the output pressure (voltage) measured across the radiation section of the vocal tract model. The I_r and R_r in Fig. 6(a) corresponds to the shunt radiation inductor (inertance) and resistor respectively. The output voltage was passed through an anti-aliasing elliptic low pass filter of order 8, pass band edge frequency of 177.6 kHz (4×44.1 kHz), pass band ripple of 2 dB and stop band attenuation of 60 dB. The signal was then sampled using a zero-order hold block with a sampling frequency of 352.8 kHz (double the pass band edge) before porting the pressure (voltage) data into a MATLAB workspace for spectral analysis and display.

FIG. 6. (a) Simulink lumped element circuit solver, (b) current source configuration, (c) output port pressure (voltage) for display in MATLAB workspace.

Figure 7 shows the frequency spectrum of the step response computed for combined losses using the Simulink solver (solid) and our custom-written NCVS solver (dot-dashed). The curves are normalized to the peak of the first formant. The comparison is for the purpose of assessing the influence of different equation solvers, not the physics. Simulink uses an ordinary differential equation solver (ODE-15s) for stiff problems. It employs implicit backward difference equations, whereas our Runge–Kutta method is an explicit ODE 45 equation solver. Our preference for the explicit solver hinges on the ability to embed conditional statements (if...then...) into the simulation, which make the solutions nonlinear and difficult to handle with implicit difference equations. For benchmarking, it is important to show that for similar space and time discretization, the algorithms for finite-difference approximations do not affect the

FIG. 7. Magnitude spectrum of the step response of the Simulink model and NCVS simulator with all losses included.

solutions (Boersma, 1998). Various approaches to implicit time and space discretization have been described by Maeda (1982) and by Birkholz (2005).

2. Formants and bandwidths

Table II lists the formant frequencies and corresponding bandwidths for each loss separately and all losses combined for this 15 cm long tube with a 3 cm² uniform cross section. Comparisons are shown between analytical calculations reported by Stevens (1998), frequency-domain simulations reported by Badin and Fant (1984) (BF), NCVS time-domain simulation, and Simulink time-domain simulation. Dashes in the table indicate cases not reported. Note that all formant frequencies and bandwidths calculated are in close agreement between NCVS and Simulink computations, suggesting consistency in programming. The mean absolute error between the NCVS and Simulink computed formant frequencies and bandwidths is 8.2 and 0.6 Hz respectively, and the standard deviation of error is 5.1 and 0.7 Hz, respectively. Comparing the time-domain simulations to Stevens' analytical calculations and the frequency-domain calculations of BF, however, shows the mean overestimation of viscous losses by 89% at low frequencies (e.g., F_1) as described earlier. Also seen is a mean underestimation of radiation losses by 11.7% at high frequencies (e.g., F_4). Birkholz and Jackel (2004) proposed a corrective technique that utilizes computational damping to offset the declining bandwidth trend with frequency, but looking at the combined losses (last column in Table II), it appears that a correction may not be necessary because some errors are offsetting.

Wall losses are also within 1.5 Hz agreement between analytical calculations and simulations, but their contributions to overall bandwidth are important only for F_1 . For F_4 ,

TABLE II. Formant frequencies and bandwidth calculations in Hz for a 15 cm long closed-open tube, 3 cm² in cross-section.

Formant	Model	No losses		Radiation losses		Viscous losses		Wall losses		All losses	
		F	BW	F	BW	F	BW	F	BW	F	BW
First	Stevens	590.0	0.0	—	3.0	—	6.0	—	8.0	592.0	20.0
	BF	—	—	502.2	2.6	—	4.8	548.1	10.5	542.5	18.3
	NCVS	591.7	0.7	561.7	5.0	591.7	14.0	615.0	7.0	585.0	23.5
	Simulink	593.5	0.8	561.5	3.9	593.5	14.0	629.0	9.5	598.5	25.5
Second	Stevens	1770.0	0.0	—	24.0	—	10.0	—	1.0	1682.0	39.0
	BF	—	—	1507.9	26.9	—	8.3	1533.5	1.0	1518.0	39.2
	NCVS	1772.0	0.6	1682.0	29.7	1772.0	14.5	1778.0	0.5	1688.0	43.5
	Simulink	1775.0	0.5	1683.0	29.0	1775.0	15.0	1788.0	0.8	1696.0	45.0
Third	Stevens	2950.0	0.0	—	67.0	—	12.0	—	0.0	2804.0	84.0
	BF	—	—	2521.9	80.5	—	10.7	2554.2	0.2	2525.3	91.6
	NCVS	2948.0	0.4	2805.0	72.0	2948.0	15	2952.0	0.6	2812.0	85.5
	Simulink	2955.0	0.3	2812.0	73.0	2955.0	14.8	2962.0	0.6	2820.0	86.0
Fourth	Stevens	4130.0	0.0	—	131.0	—	15.0	—	0.0	3927.0	152.0
	BF	—	—	3545.9	141.1	—	12.6	3596.8	0.1	3549.4	155.3
	NCVS	4122.0	0.2	3935.0	123.0	4122.0	14.5	4125.0	0.5	3942.0	134.5
	Simulink	4135.0	0.2	3950.0	123.0	4135.0	15.0	4140.0	0.4	3956.0	136.0

for example, wall losses are an order of magnitude smaller than viscous losses and two orders of magnitude smaller than radiation losses. When all losses are combined, there are no major discrepancies between formant frequency calculations by Stevens and NCVS, suggesting that an average of the two values can be used as a benchmark with less than 0.5% error. BF formant frequency calculations are categorically lower and would not be included in the average. Bandwidth calculations would favor an average between BF and Stevens, given that they include frequency-dependence in the viscous losses.

Between the two time-domain simulations (NCVS and Simulink), bandwidth error (± 2 Hz) is due to the finite frequency resolution used to compute the fast Fourier transform (FFT). We used linear interpolation to improve the resolution when measuring the bandwidths. There can also be spectral leakage error when computing FFT of non-periodic signals (impulse responses) which is more difficult to quantify. Spectral leakage will spread energy into adjacent bins leading to increase in bandwidths and shift in the location of formant peaks. We took a relatively large time sample duration (0.3 s) of the impulse response to reduce the effect of spectral leakage.

3. Influence of random section lengths

A final test with a uniform tube was to vary the section lengths randomly. A random number generator was used to vary Δz across the 44 sections. Results for our NCVS simulator are shown in Fig. 8. Dot-dashed lines are for equal section lengths of 0.35 cm, whereas solid lines are for randomized length with a mean of 0.35 cm and a standard deviation of 0.1 cm. According to the outliers produced with a Gaussian distribution, the minimum length was 0.2 cm and the maximum length was 0.55 cm. There is no visible difference between constant and variable lengths, except for a small discrepancy of about 0.5 dB in the troughs of the spectrum (35 dB below the F_1 peak). We conclude that section

length can be arbitrary with the transmission line approach, as long as the stability rule [Eq. (28)] and the accuracy rule [Eq. (30)] are not compromised.

C. Losses for three vowel shapes

Table III shows formant frequencies and bandwidths calculations for the x-ray derived configurations of the vowels /a/, /i/, and /u/ (Fant, 1960). The comparison is between frequency-domain computations by Badin and Fant (1984) and time-domain simulations with the NCVS simulator. For the purpose of comparison, we chose area-dependent wall losses, piston in sphere (PIS) radiation impedance losses, and combined viscosity and heat conduction losses from BF. The mean absolute difference is 27.7 Hz and standard deviation is 47.7 Hz between the two formant frequency computations, with mean percentage difference of 2%. The

FIG. 8. Step function response for a uniform tube with losses. Solid lines are for randomized section lengths Δz and dashed lines are for constant section lengths.

TABLE III. Formant frequencies and bandwidth calculations in Hz from all losses excluding kinetic losses for three vowel shapes.

	Formant	Model	Radiation losses		Viscous losses		Wall losses		All losses excluding kinetic losses	
			F	BW	F	BW	F	BW	F	BW
/a/	First	BF	642.3	3.9	—	10.4	705.8	12.3	705.5	25.9
		NCVS	640.5	6.5	678.5	18.3	700.5	5.6	660.5	26.5
	Second	BF	1085.3	16.7	—	11.3	1109.6	3.4	1108.6	30.2
		NCVS	1081.0	23.5	1191.0	16.8	1201.0	1.8	1093.0	42.0
	Third	BF	2469.1	40.4	—	19.0	2473.0	0.2	2471.5	58.3
		NCVS	2468.0	45.7	2557.0	16.6	2562.0	1.3	2475.0	61.7
	Fourth	BF	3620.0	135.0	—	20.6	3626.3	0.2	3625.1	152.1
		NCVS	3610.0	142.0	3785.0	14.9	3789.0	0.8	3617.0	157.8
	Fifth	BF	4134.5	29.2	—	26.1	4150.9	0.5	4146.2	56.2
		NCVS	4123.0	28.3	4174.0	20.1	4178.0	0.4	4135.0	50.4
/i/	First	BF	227.0	0.1	—	6.4	304.6	31.8	301.6	36.9
		NCVS	227.5	0.9	230.5	27.9	270.5	23.2	267.5	43.3
	Second	BF	2276.0	5.2	—	12.5	2285.3	0.6	2285.6	18.4
		NCVS	2279.0	6.5	2288.0	9.9	2293.0	1.3	2278.0	16.7
	Third	BF	3104.6	162.0	—	29.2	3107.5	0.5	3118.3	187.6
		NCVS	3095.0	213.0	3289.0	20.0	3294.0	0.4	3103.0	231.6
	Fourth	BF	3729.5	56.5	—	21.7	3735.1	0.1	3732.0	76.4
		NCVS	3721.0	59.8	3794.0	16.0	3798.0	0.3	3727.0	75.4
	Fifth	BF	4748.6	384.9	—	22.0	4740.0	0.2	4756.1	402.9
		NCVS	4744.0	—	4945.0	13.4	4948.0	0.7	4725.0	—
/u/	First	BF	237.3	0.1	—	8.1	318.1	30.4	316.3	36.8
		NCVS	237.5	0.6	249.5	34.5	289.5	22.1	280.5	46.9
	Second	BF	600.2	0.2	—	9.5	636.9	9.4	636.0	18.7
		NCVS	601.5	1.3	609.5	25.1	627.5	4.6	622.5	27.3
	Third	BF	2383.0	0.2	—	19.3	2384.8	0.1	2384.8	19.6
		NCVS	2386.0	0.6	2388.0	16.5	2394.0	0.5	2390.0	16.8
	Fourth	BF	3710.2	0.4	—	20.1	3710.9	0.0	3711.2	20.5
		NCVS	3704.0	1.5	3706.0	13.9	3710.0	1.6	3706.0	14.2
	Fifth	BF	4055.9	1.1	—	21.3	4058.2	0.1	4057.9	22.3
		NCVS	4046.0	1.6	4049.0	14.1	4053.0	0.3	4052.0	15.3

bandwidths, however, have various levels of agreement. Radiation losses are larger and wall losses are smaller for the NCVS simulator. This might be due to the difference in approach used to model the radiation and wall losses by BF. We used a parallel RL circuit whereas BF used a parallel RLC circuit in series with an inductor to model radiation impedance. Similarly, we used a series RLC circuit to model the wall losses while BF used series RL circuit in parallel with a resistor. The mean percentage increase in bandwidth for radiation losses is 29.2% and decrease in bandwidth for wall losses is 41.3%. The bandwidths due to viscous losses follow a similar trend as was seen for the uniform tube, with a mean overestimation at low frequencies by 190.1% and underestimation at high frequencies by 26%. Of particular interest are the narrow F_3 , F_4 , and F_5 bandwidths for the /u/ vowel, an order of magnitude smaller than those of /i/. In these calculations, this is entirely related to the small radiation losses with small mouth opening. Bandwidths that are less than 1 Hz are not included in the computation of the statistics as they are within the frequency resolution error of 2 Hz and also skew the percentages by large extent. Yet to be included, however, are the kinetic losses, which tend to be largest for /u/.

V. BENCHMARK FOR KINETIC LOSSES IN EXPANSIONS AND CONTRACTIONS

Kinetic losses can occur within a given section, as Eqs. (13) and (16) show, but they are usually small inside a section with constant cross-sectional area unless wall movement or air compliance is large. If conical sections are used (Maeda, 1982), kinetic losses play a greater role within the section and less of a role at the junctions. For cylindrical sections used here, $v_1 \approx v_2$, and the entry gain in kinetic pressure generally cancels the exit loss, or vice versa. Across sections, however, the kinetic gains or losses can be rather large. In previous work [Titze, 2006 (p. 309)] a loss factor α was proposed for a pressure drop due to a sudden expansion,

$$p_{2,n} - p_{1,n+1} = -\frac{1}{2}\rho v_{2,n}^2 \left[\alpha + (2-\alpha) \frac{A_n}{A_{n+1}} \right] \left(1 - \frac{A_n}{A_{n+1}} \right). \quad (36)$$

For $\alpha = 1$, there is full pressure recovery, as predicted by Bernoulli's energy conservation law. For $\alpha = 0$, Eq. (36) is the expression for pressure recovery at glottal exit suggested by Ishizaka and Matsudaira (1972) for fully detached jet

flow in an expanding pipe. It was then used by [Ishizaka and Flanagan \(1972\)](#) in their two-mass vocal fold model. Some compromise between the attached flow boundary condition and the non-attached flow boundary condition must be reached for general use in the vocal tract. For an area expansion ratio $\beta = A_{n+1}/A_n$, we propose the piece-wise continuous function,

$$\alpha = 1 - .5 \alpha_m [1 - \cos \pi(\beta - 1)/(\beta_d - 1)],$$

$$1 < \beta < \beta_d, \quad (37)$$

$$\alpha = \alpha_m, \quad \beta \geq \beta_d, \quad (38)$$

where α_m is a maximum exit pressure recovery loss and β_d is the area ratio for which this maximum recovery loss occurs due to flow detachment. [Liljencrants \(1991\)](#) and [Pelorson et al. \(1994\)](#) predicted β_d to be on the order of 1.2. In other words, the flow detaches from the wall when the area suddenly expands by 20% or more. The value α_m varies anywhere from 0.2 to 0.8, depending on how the jet spreads after detachment. The half-cosine transition region chosen in Eq. (37) guarantees smooth endpoints for zero and maximum detachment.

For a sudden contraction, a vena contracta may form that makes the flow lines contract more than the wall boundary. [White \(1979\)](#) suggests a kinetic pressure loss coefficient ranging linearly from 1.0 to 1.42 with decreasing area ratio. [Beavers et al. \(1970\)](#) and [Li et al. \(2012\)](#) cast doubt on the formation of a vena contracta, however, regardless of how sharp the entry corner is. In either case, additional viscous or vorticity losses do occur at a sudden contraction. Here we propose an entry loss coefficient $(1 + \gamma)$, where γ is dependent on the area ratio β and a limiting value β_c ,

$$\gamma = 0.5 \gamma_m [1 - \cos \pi(1 - \beta)/(1 - \beta_c)], \quad 1 > \beta > \beta_c, \quad (39)$$

$$\gamma = \gamma_m, \quad \beta \leq \beta_c. \quad (40)$$

The junction pressure is then

$$p_{2,n} - p_{1,n+1} = -(1 + \gamma) \left(\frac{A_n^2}{A_{n+1}^2} - 1 \right) \frac{1}{2} \rho v_{2,n}^2. \quad (41)$$

For $\gamma = 0$, the expression yields the Bernoulli junction pressure drop [the negative of Eq. (36) for $\alpha = 0$]. For a severe contraction, $A_n \ll A_{n+1}$, the junction pressure drop is $-(1 + \gamma_m) \frac{1}{2} \rho v_{2,n}^2$, and for $A_n = A_{n+1}$ there is no junction pressure drop.

Assuming that kinetic losses are the same for oscillatory flow as for steady flow (the quasi-steady flow assumption to be discussed later), the data for the M5 model by [Scherer et al. \(2010\)](#) can be used to determine the empirical coefficients α_m , β_d , γ_m , and β_c from pressure distributions in physical models of expansions and contractions. These data do not separate viscous losses from kinetic losses, however. The viscous loss from Eq. (31) is therefore maintained in conjunction with the kinetic losses of Eqs. (36)–(41) when benchmarks are produced. We used nearly the entire data set of [Scherer et al. \(2010\)](#) (nine convergence angles: -40° ,

-20° , -10° , -5° , 0° , $+5^\circ$, $+10^\circ$, $+20^\circ$, $+40^\circ$; seven minimum diameters: 0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32 cm; four transglottal pressures: 3, 5, 10, 15 cm H₂O) to optimize the values α_m , γ_m , β_d , and β_c . The optimization was done using the simulated annealing algorithm under MATLAB. The best overall fit was for $\alpha_m = 0.5$, $\gamma_m = 1.0$, $\beta_d = 1.2$, $\beta_c = 0.6$.

A. Evaluating the optimization performance

Figure 9 (top left) shows an area function of our model that approximates the area function of [Scherer et al. \(2010\)](#) for a contraction into a glottal constriction of length 1.2 cm, minimum diameter 0.04 cm, and 0° divergence in the glottis, one of the many cases investigated by [Scherer et al. \(2010\)](#). A sudden expansion follows. The transglottal pressure was 10 cm H₂O. The top right shows the minor radius a and the major radius b as the vocal tract cross section changes from circular to highly elliptical to approximate the rectangular cross sections of [Scherer et al. \(2010\)](#). The middle left panel shows the empirical functions $(1 + \gamma)$ in dotted lines and α in solid lines as they vary along the glottis. Note that $1 + \gamma$ rises from the value 1.0 ($\gamma = 0$) to 2.0 ($\gamma = \gamma_m = 1.0$) during the transition into the glottis. It falls sharply and remains 1.0 in the glottis and during expansion. The function α is 1.0 during the transition into and through the glottis, but drops to $\alpha_m = 0.5$ in the sudden expansion at glottal exit. If Bernoulli (energy conservation) conditions were imposed, α and $1 + \gamma$ would both remain constant at the value 1.0 over the entire length of the contraction and expansion. The middle right panel shows the measured volume flow of [Scherer et al. \(2010\)](#) in cm³/s (dashed line) and our calculated flow (solid line). The trace near the bottom in the middle panel is the calculated air particle velocity in m/s. The bottom left panel

FIG. 9. (Color online) Combined viscous and kinetic loss test for a transition into a rectangular constriction with 0° angle, 0.04 cm diameter, and 10 cm H₂O transglottal pressure. (Top left) area function, (top right) radii a and b of an elliptical cross section, (middle left) functions $1 + \gamma$ and α , (middle right) flow and particle velocity, (bottom left) pressure profile, (bottom right) change in kinetic pressure. Dashed lines are the measurements of [Scherer et al. \(2010\)](#) and solid lines are calculations ($\alpha_m = 0.5$, $\gamma_m = 1.0$, $\beta_d = 1.2$, $\beta_c = 0.8$).

shows the pressure profile, dashed lines for the data of Scherer *et al.* (2010) and solid line for our computation [Eqs. (36)–(41)].

The bottom right shows the change in kinetic pressure P_k . Note that the increase at glottal entry is not equal to the decrease at glottal exit, as would be the case for Bernoulli energy conservation.

Figure 10 shows a similar set of curves for a constriction with a 20° glottal convergence angle. All other parameters are the same. The pressure drop at entry is a bit overestimated, but the flow matches. The M5 model of Scherer *et al.* (2010) had entry and exit roundings. At this stage, our area discretization is not fine enough to model the exact nature of the entry and exit roundings, which may account for some of the differences (Scherer *et al.*, 2001). Also, the sampling of eight pressure taps of Scherer *et al.* (2010) may not be adequate to capture the exact peak. We used 100 samples to plot the continuous curves.

Figure 11 shows similar curves for a constriction with a 20° glottal divergence angle. All other parameters are the same. Again, the pressure drop at entry is overestimated, but the flow is underestimated slightly. Considering that we targeted the best combined values of α_m , γ_m , β_d , and β_c for 180 glottal shapes of Scherer *et al.* (2010) (convergent, rectangular, divergent), it is not surprising that the match for any individual shape is not ideal with only four parameters.

Percentage of error was computed to find the differences between pressures and flows calculated by NCVS simulation and the data of Scherer *et al.* (2010). The eight pressure taps and the flow inside the glottis are considered separately in the error determination. The error for each pressure tap was normalized to the transglottal pressure and flow error was

FIG. 10. (Color online) Combined viscous and kinetic loss test for a transition into a convergent constriction with 20° angle, 0.04 cm diameter, and 10 cm H₂O transglottal pressure. (Top left) area function, (top right) radii a and b of an elliptical cross section, (middle left) functions $1 + \gamma$ and α , (middle right) flow and particle velocity, (bottom left) pressure profile, (bottom right) change in kinetic pressure. Dashed lines are the measurements of Scherer *et al.* (2010) and solid lines are calculations ($\alpha_m = 0.5$, $\gamma_m = 1.0$, $\beta_d = 1.2$, $\beta_c = 0.8$).

FIG. 11. (Color online) Combined viscous and kinetic loss test for a transition into a divergent constriction with 20° angle, 0.04 cm diameter, and 10 cm H₂O transglottal pressure. (Top left) area function, (top right) radii a and b of an elliptical cross section, (middle left) functions $1 + \gamma$ and α , (middle right) flow and particle velocity, (bottom left) pressure profile, (bottom right) change in kinetic pressure. Dashed lines are the measurements of Scherer *et al.* (2010) and solid lines are calculations ($\alpha_m = 0.5$, $\gamma_m = 1.0$, $\beta_d = 1.2$, $\beta_c = 0.6$).

normalized to the measured value of Scherer *et al.* (2010) to obtain percent errors. The results across the 180 combinations of the glottal shapes are presented in Fig. 12 using box-and-whisker plots. (The box-and-whisker plots depict minimum, 25th percentile, median, 75th percentile and maximum errors). Outliers (above 2.7σ) are represented by “+” signs. Overall, pressures and flows are matched with a median error of less than 20% (center line in boxes).

B. Kinetic losses for three vowel shapes

Table IV shows the predicted bandwidth contributions from kinetic losses computed for the three Fant vowels /a/, /i/, and /u/ based on the Fant (1960) area functions (see Table V). The bandwidths from kinetic losses range between 10 to 35 Hz for the three vowels, except for the first and

FIG. 12. (Color online) Box and whisker plot for percent errors in kinetic pressure profile and flow calculations. T1 to T8 represent pressure taps across the glottis from entry to exit.

TABLE IV. Formant frequencies and bandwidth calculations in Hz from kinetic losses for three vowel shapes.

Vowel	Formant	Kinetic losses		All losses including kinetic losses	
		F	BW	F	BW
/a/	First	678.5	21.8	660.5	42.9
	Second	1191.0	20.8	1091.0	64.5
	Third	2557.0	26.2	2475.0	84.8
	Fourth	3785.0	22.5	3615.0	177.8
	Fifth	4174.0	36.7	4135.0	93.0
/i/	First	230.5	33.8	267.5	64.3
	Second	2288.0	4.2	2277.0	21.3
	Third	3289.0	34.4	3105.0	300.2
	Fourth	3794.0	21.0	3726.0	88.0
	Fifth	4945.0	11.8	4740.0	—
/u/	First	241.0	217.5	273.5	143.3
	Second	595.5	67.7	613.5	81.4
	Third	2388.0	13.8	2387.0	28.5
	Fourth	3706.0	14.0	3703.0	26.0
	Fifth	4049.0	12.7	4053.0	26.5

second formant of /u/, where they are significantly larger. The vowel /u/ has a major contraction-expansion at the lips, which would be expected to dissipate kinetic energy. The overall (combined) losses are also re-computed with kinetic losses included. Large bandwidths are noted for the fourth formant of /a/, the third formant of /i/, and the first formant of /u/. By far the most important is the first formant bandwidth of /u/, where kinetic losses dominate all other losses. The overall bandwidth rises from 47 to 144 Hz in our calculations.

While an exhaustive and definitive investigation of bandwidths with severe kinetic losses in human subjects was outside the scope of this paper, some validation of the results is shown in Table VI for two male subjects. Vocal fry phonation was used to sample the formants adequately with closely spaced frequencies. While the bandwidths vary significantly for /i/ and /u/, they show that predicted bandwidths above 100 Hz can occur in human subjects. The difficulty in obtaining normative data may stem from significant variations in vocal tract configurations and their associated kinetic losses. Kinetic losses in tense vowels, semi-vowels, and consonants should be the focus of future studies.

VI. MAXIMUM VELOCITY AND PRESSURE TESTS

When airway sections are near collapse, viscous and kinetic resistances can become very large because they vary inversely with cross sectional area. Since all calculations are based on linearized acoustic pressures and flows, it is important to conduct tests for maximum air particle velocity and maximum pressure throughout the entire airway. Basic rules of thumb are that air particle velocity should be less than 10% of the speed of sound (<35 m/s) and that pressure magnitudes (positive or negative) should be less than 10% of atmospheric pressure (<10 kPa). These conditions were basically met for the simulations shown here. Particle velocity

TABLE V. Area functions of three Russian vowels given by Fant (1960) across distance x from the glottis.

x (cm)	/a/	/i/	/u/
0	2.60	3.20	2.60
0.5	2.60	3.20	2.60
1.0	1.60	2.60	2.60
1.5	1.30	2.00	2.00
2.0	1.00	2.00	2.00
2.5	4.00	8.00	10.50
3.0	2.60	8.00	10.50
3.5	1.60	10.50	10.50
4.0	1.00	10.50	8.00
4.5	0.65	10.50	8.00
5.0	0.65	10.50	5.00
5.5	0.65	10.50	3.20
6.0	1.00	10.50	1.60
6.5	1.30	10.50	1.30
7.0	1.60	10.50	1.00
7.5	2.00	10.50	1.00
8.0	2.60	8.00	1.00
8.5	2.60	8.00	1.60
9.0	1.60	6.50	2.00
9.5	3.20	4.00	1.30
10.0	4.00	2.60	1.60
10.5	5.00	1.30	2.00
11.0	6.50	0.65	2.00
11.5	8.00	0.65	2.00
12.0	8.00	0.65	2.60
12.5	8.00	0.65	3.20
13.0	8.00	0.65	5.00
13.5	8.00	0.65	6.50
14.0	8.00	0.65	8.00
14.5	8.00	1.00	10.50
15.0	8.00	1.30	13.00
15.5	6.50	1.60	13.00
16.0	5.00	3.20	13.00
16.5	5.00	4.00	13.00
17.0	5.00		10.50
17.5			5.00
18.0			2.00
18.5			0.32
19.0			0.32
19.5			0.65

did reach about 30 m/s for the divergent glottis of Scherer's (2010) M5 model. The magnitude of air particle velocity should always be tested, especially for non-steady flow conditions that involve inertial pressures.

TABLE VI. Formant frequencies and bandwidths in Hz for three vowel shapes in two male subjects.

Vowel	Subject	F1	BW1	F2	BW2
/a/	S1	667.8	38.8	1075.9	52.9
	S2	644.5	42.1	1041.5	96.1
/i/	S1	300.8	90.3	1987.2	121.4
	S2	300.4	60.7	2206.3	78.3
/u/	S1	283.8	150.3	806.2	51.4
	S2	327.1	58.4	825.4	165.8

VII. DISCUSSION AND CONCLUSIONS

Time-domain simulation of vocal tract acoustics is in wide use. For one-dimensional approaches, where computational speed is the motivating factor, several algorithms are in use that capture the essence of wave propagation and energy loss when vocal tract geometries are complex. Depending on the application, many simplifying assumptions are made to avoid solution of the complete three-dimensional Navier–Stokes equation. Some benchmarks for testing the validity and accuracy of these assumptions and algorithms have been proposed here. We provided a comparison between formant frequencies and bandwidths determined computationally in the frequency domain by [Badin and Fant \(1984\)](#), analytically in the frequency domain by [Stevens \(1998\)](#), and computationally with the time-domain transmission line algorithm. It was shown how the formant frequencies are underestimated with insufficient spatial discretization and how formant bandwidths are affected by viscous, wall vibration, kinetic, and radiation losses. A 0.5% accuracy rule of thumb was given for spatial discretization. It was also shown that arbitrary section lengths can be used as long as the stability and accuracy rules are met. Data by [Scherer et al. \(2010\)](#) were used to benchmark kinetic losses in constrictions that may occur in the vocal tract, not only in the glottis, but also for vowel production like /u/ or /o/ and consonantal semi-occlusions. Four optimized parameters were proposed for kinetic losses in a variety of configurations, but the optimization was for steady flow, not oscillatory flow. The quasi-steady flow assumption has not yet been invalidated, either by theory or by experiment. We expect that, with today's computational tools, the answer is not more than 3–5 years in the future.

The most important limitation in this study was the one-dimensional approximations to the Navier–Stokes equation. No vortex shedding and jet instabilities are included, except by empirical rule, and no wave propagation perpendicular to the axial direction in the vocal tract is currently calculated. A second limitation was the rather simple (and possibly outdated) approximation of vocal tract wall losses. We used it simply as a benchmark comparison to earlier calculations. Wall losses may vary along the airways due to different tissue compositions. Effective wall mass, stiffness, and damping are likely to be different in the epiglottal region than in the buccal or tracheal regions. Much work is still needed to characterize the wall properties along the entire airway.

A significant advancement has been made in the use of elliptical cross sections, with a major and a minor axis, which can allow the vocal tract to collapse at any portion of the airway. With circular cross sections, airway collapse is unrealistic because tissues would need to compress and expand radially. This is not possible at sonic frequencies. Rather, airway cross sections deform to produce slit openings (e.g., glottis, ventricular glottis, or lips). We have begun the task of quantifying losses when cross-sections change from circular to elliptical under wall stresses, but no benchmarks exist currently for collapse with fluid-structure interaction.

ACKNOWLEDGMENT

Support for this research comes from Grant No. 1 R01 DC008612-01A1 by the National Institute on Deafness and Other Communication Disorders.

- Badin, P., and Fant, G. (1984). "Notes on vocal tract computation," in *Quarterly Progress and Status Report* (Department for Speech, Music, and Hearing, KTH, Stockholm), pp. 53–108.
- Beavers, G. S., Sparrow, E. M., and Magnuson, R. A. (1970). "Experiments on hydrodynamically developing flow in rectangular ducts of arbitrary aspect ratio," *Int. J. Heat Mass Transf.* **13**, 689–702.
- Birkholz, P. (2005). *3D-Artikulatorische Sprachsynthese (3-D Articulatory Speech Synthesis)* (Logos Verlag, Berlin), pp. 47–102.
- Birkholz, P., and Jackel, D. (2004). "Influence of temporal discretization schemes on formant frequencies and bandwidths in time-domain simulations of the vocal tract system," in *INTERSPEECH-2004, 8th International Conference on Spoken Language Processing*, pp. 1125–1128.
- Birkholz, P., Jackel, D., and Kröger, B. J. (2006). "Construction and control of a three-dimensional vocal tract model," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006)*, Toulouse, France, pp. 873–876.
- Boersma, P. (1998). *Functional Phonology: Formalizing the Interactions Between Articulatory and Perceptual Drives* (Holland Academic Graphics/IFOTT, The Hague, Netherlands).
- Fant, G. (1960). *The Acoustic Theory of Speech Production* (Moulton, The Hague, Netherlands), pp. 32, 39, 125–128, 292.
- Flanagan, J. L. (1965). *Speech Analysis: Synthesis and Perception* (Springer-Verlag, Berlin), pp. 27, 32–36.
- Ishizaka, K., and Flanagan, J. L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* **51**, 1233–1268.
- Ishizaka, K., French, I. C., and Flanagan, J. L. (1975). "Direct determination of vocal tract wall impedance," *IEEE Trans. Acoust. Speech Sign. Process.* **23**, 370–373.
- Ishizaka, K., and Matsudaira, M. (1972). "Theory of vocal cord vibration," *Re. Univ. Electro-Comm. Sci. Tech. Sect.* **23**, 107–136.
- Kelly, J. L., and Lochbaum, C. (1962). "Speech synthesis," in *Proceedings of the 4th International Congress on Acoustics*, paper G42, pp. 1–4.
- Li, S., Scherer, R. C., Wan, M., and Wang, S. (2012). "The effect of entrance radii on intraglottal pressure distributions in the divergent glottis," *J. Acoust. Soc. Am.* **131**(2), 1371–1377.
- Liljencrants, J. (1985). "Speech synthesis with a reflection-type line analog," Ph.D. dissertation, Royal Institute of Technology, Department of Speech Communication and Music Acoustics, Stockholm, Sweden.
- Liljencrants, J. (1991). "Numerical simulations of glottal flow," in *Vocal Fold Physiology: Acoustic Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular Publishing Group, San Diego), pp. 99–104.
- Maeda, S. (1982). "A digital simulation of the vocal tract system," *Speech Commun.* **1**, 199–229.
- The Mathworks. (2014). MATLAB, version 8.3.0.532 (R2014a).
- Milenkovic, P. H., Yaddanapudi, S., Vorperian, H. K., and Kent, R. D. (2010). "Effects of a curved vocal tract with grid-generated tongue profile on low-order formants," *J. Acoust. Soc. Am.* **127**(2), 1002–1013.
- Pelorsson, X., Hirschberg, A., Van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (1994). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," *J. Acoust. Soc. Am.* **96**(6), 3416–3431.
- Scherer, R. C., De Witt, K. J., and Kucinschi, B. R. (2001). "The effect of exit radii on intraglottal pressure distributions in the convergent glottis," *J. Acoust. Soc. Am.* **110**(5), 2267–2269.
- Scherer, R. C., Torkaman, S., Kucinschi, B. R., and Afjeh, A. A. (2010). "Intraglottal pressures in a three-dimensional model with a non-rectangular glottal shape," *J. Acoust. Soc. Am.* **128**(2), 828–838.
- Stevens, K. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA), pp. 138, 162.
- Story, B. (1995). "Physiologically based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. dissertation, University of Iowa, Iowa City.

- Taflove, A., and Hagness S. C. (2005). *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed. (Artech House, Norwood, MA), pp. 45, 49.
- Titze, I. R. (2006). *The Myo-elastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Salt Lake City, Utah), pp. 297–338.
- Vampola, T., Horáček, J., Laukkanen, A.-M., and Švec, J. G. (2013). “Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modeling based on CT measurement,” *Logoped. Phoniatr. Vocol.* 1–10, doi: 10.3109/14015439.2013.775333.
- Vampola, T., Horáček, J., and Švec, J. G. (2008). “FE modeling of human vocal tract acoustics. Part I: Production of Czech vowels,” *Acta Acust. Acust.* **94**(3), 433–447.
- Wakita, H., and Fant, G. (1978). “Toward a better vocal tract model,” *STL-QPSR.* **19**(1), 9–29.
- White, F. M. (1979). *Fluid Mechanics* (McGraw-Hill, New York), p. 356.
- Zhang, Z., and Espy-Wilson, C. Y. (2004). “A vocal tract model of American English /I/,” *J. Acoust. Soc. Am.* **115**(3), 1274–1280.