9

Acoustic tubes

Wave propagation in acoustic tubes forms the basis of musical sound production in wind and brass instruments, as well as in the human voice. For various reasons, synthesis based on models of such instruments has seen a lot of activity—chief among these, perhaps, are the links with speech analysis and synthesis, the relative ease with which such systems may be analyzed and the extensive work in musical acoustics, and the use of waveguides which deal with some special cases with extreme efficiency.

The key concepts which are employed in the acoustical analysis of tube-based instruments are impedance, as well as ideas from scattering theory, such as reflectance, wave decompositions, and so forth—these are, of course, of extreme utility in model validation, given that measurements of tube responses are often best understood in the frequency domain. As in the rest of this book, this point of view is incorporated here, but to a lesser degree than is usual in pure musical acoustics. One can design time domain simulations without invoking any notion of frequency; the big bonus of a finite difference formulation is that the resulting implementation can be almost embarrassingly simple—see the code example given in Section A.10.

A general model of 1D wave propagation, Webster's equation, is presented in Section 9.1, along with a discussion of boundary conditions (particularly the radiation condition), modes, associated finite difference schemes, and the special cases of tubes of cylindrical and conical profile. As an application, such difference schemes are applied to the case of the vocal tract in Section 9.2, where one may find a direct correspondence between rudimentary difference schemes and scattering-based synthesis methods along the lines of the famous Kelly–Lochbaum model [201]. Finally, the more difficult case of the coupling of a tube to a nonlinear reed-like mechanism is covered in Section 9.3, from which follows a completely discrete wind instrument synthesis method.

9.1 Webster's equation

There are various ways of defining what is meant by the word "tube." In general, it can be described as an enclosure for which the length scale in one coordinate is significantly greater than in the others. In linear problems, the dynamics of the material filling the tube (air in musical applications) will satisfy the wave equation in 3D. If it is true that the wavelengths of interest are longer than the length scale in the two "short" coordinate directions, then it is possible to simplify the dynamics to one dimension, which will be called here, as in the case of bars and strings, x. A general tube-like object in musical acoustics, then, is characterized by the material properties of air, as well as

Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics Stefan Bilbao
 ISBN: 978-0-470-51046-9

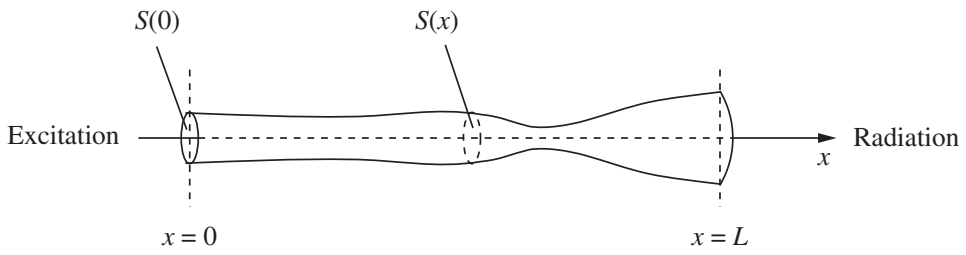


Figure 9.1 A 1D acoustic tube, with excitation at the left end, and radiated sound produced at the right end.

the function $S(x)$, representing, grossly, the cross-sectional area[1] of the tube at position x, where $x \in [0, L]$. See Figure 9.1.

A good model of the dynamics of such a tube, in terms of the variables $p(x, t)$ and $u(x, t)$, the pressure and volume velocity respectively at any point x and time t, may be derived through linearization of the equations of fluid dynamics:

$$\frac{S}{\rho c^2} p_t = -u_x \qquad \frac{\rho}{S} u_t = -p_x \tag{9.1}$$

where ρ is density, and c is wave speed. The first equation above corresponds to pointwise conservation of mass and the second to conservation of momentum. From this first-order system, one may derive what is often referred to as Webster's equation [406]:

$$S\Psi_{tt} = c^2 (S\Psi_x)_x \qquad \text{with} \qquad p = \rho\Psi_t, \quad u = -S\Psi_x \tag{9.2}$$

The assumptions which lead to this compact form are many: linearity, an equation of state of a special form relating pressure and density, and also that variations in u and p (and thus Ψ) are larger in scale than the tube width—these are all roughly valid in the case of the human voice, and in wind instruments except under high-amplitude playing conditions, or when bell flare becomes large. Such assumptions are discussed in detail by Benade and Jansson [32]. It is also worth pointing out that the first-order system (9.1), in the power conjugate variables p and u, is of the same form as the transmission line or telegrapher's equations [80], when p and u are replaced by a voltage and current, respectively, and the material constants by inductances and capacitances. The transmission line point of view is at the heart of many synthesis algorithms, including the Kelly–Lochbaum model [201], as well as digital waveguides and other scattering-based methods [344, 343]. For time domain synthesis, however, a second-order form such as (9.2) is perhaps simpler to deal with. Sometimes, as in the case of speech synthesis [280], Webster's equation is written directly in terms of the pressure variable through time differentiation of (9.2), but (9.2) is more fundamental, in that a correct energy balance is preserved—see Section 9.1.3.

Webster's equation appears in slightly different (equivalent) forms in the literature; one variant of particular interest is

$$\Phi_{tt} = c^2 (\Phi_{xx} - a(x)\Phi) \qquad \text{with} \qquad \Phi = \sqrt{S}\Psi, \quad a(x) = \frac{S_{xx} S - \frac{1}{2} S_x^2}{2S^2} \tag{9.3}$$

Notice that the second spatial derivative term appears with a constant coefficient in this case; the quantity $a(x)$ is an expression for the curvature of the tube at x—see Problem 9.1.

[1] In fact, it represents the area of an isophase surface of the pressure distribution in the tube [38, 32].

Scaled form

It is again useful to scale variables, through the introduction of the dimensionless coordinate $x' = x/L$; another simplification is the scaling of the surface area function S, through a variable $S' = S(x)/S_0$, where S_0 is a reference surface area. A good choice of S_0 is $S_0 = S(0)$, so that the scaled area function always has $S'(0) = 1$. The dependent variables p and u in system (9.1) may also be scaled as $p' = p/\rho c^2$, $u' = u/cS_0$, to yield, after removing primes,

$$Sp_t = -\gamma u_x \qquad \frac{1}{S}u_t = -\gamma p_x \qquad \text{for} \qquad x \in \mathbb{U} \tag{9.4}$$

whereas in the case of the 1D wave equation (see Section 6.1.2), $\gamma = c/L$.

The resulting systems corresponding to (9.2) and (9.3) are, after introducing the variables $\Psi' = \Psi/cL$ and $\Phi' = \Phi/cL\sqrt{S_0}$, and removing primes, of the form

$$S\Psi_{tt} = \gamma^2 (S\Psi_x)_x \qquad \Phi_{tt} = \gamma^2 (\Phi_{xx} - a(x)\Phi) \qquad \text{for} \qquad x \in \mathbb{U} \tag{9.5}$$

where, now, the non-dimensional variables p and u may be recovered from non-dimensional Ψ as

$$p = \frac{1}{\gamma}\Psi_t \qquad u = -S\Psi_x \tag{9.6}$$

9.1.1 Dispersive behavior

Webster's equation, like the equations defining the string of variable density and bar of variable cross-section (see Section 7.10), is linear and time invariant, but not more generally shift invariant. In this case the group and phase velocity are difficult to define properly, but one could attempt to examine such quantities on a "local" basis, assuming that the variation in $S(x)$ is not too large. The second form given in (9.5) is a good starting point. Suppose one examines a test solution of the form $\Phi(x, t) = e^{j\omega t + j\beta_{x_0} x}$, where the quantity β_{x_0} indicates a wavenumber in the neighborhood of x_0. One then arrives at the local dispersion relation

$$\omega^2 = \gamma^2 \left(\beta_{x_0}^2 + a(x_0)\right)$$

from which one can derive local phase and group velocities v_{ϕ,x_0} and v_{g,x_0}:

$$v_{\phi,x_0} = \gamma \frac{\sqrt{\beta_{x_0}^2 + a(x_0)}}{\beta_{x_0}} \qquad v_{g,x_0} = \gamma \frac{\beta_{x_0}}{\sqrt{\beta_{x_0}^2 + a(x_0)}}$$

Obviously, then, just as in the case of the ideal bar (see Section 7.1.1), wave propagation is dispersive to an extent which depends on the size of a. In contrast with the case of the bar, however, one may remark that, in the limit of high wavenumbers (or high frequencies), one has $v_{\phi,x_0}, v_{g,x_0} \to \gamma$, regardless of position x_0. Also, it is evident that dispersion will be greatest for low frequencies.

A graphical example is in order at this point—see Figure 9.2. The phenomenon of wave propagation in an acoustic tube is indeed dispersive—notice, however, that compared to the similar case of a pulse in a vibrating ideal bar, as illustrated in Figure 7.1, the transit of the pulse is relatively coherent. There is no cascade of ripples which precedes the main body of the pulse, but there is a pronounced low-frequency backscatter as the pulse passes through a constriction (i.e., a region of high tube curvature). This reflects the limiting behavior of the group and phase velocities, as discussed above. In addition, even as the pulse passes through the constriction, the overall speed of the pulse remains constant.

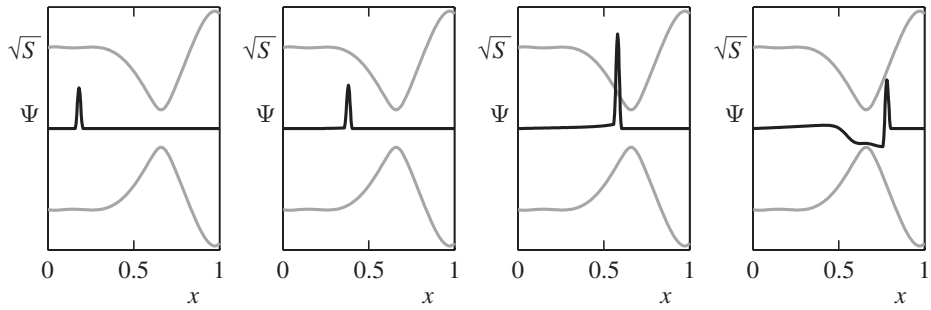


Figure 9.2 Dispersion of a pulse in an acoustic tube described by Webster's equation (9.5). The pulse travels from left to right, and the tube profile is shown as grey contours.

The distinctions among various types of spatial variation will have consequences on the numerical methods which result. One might expect that in this case, because the group and phase velocities approach a constant value, regardless of the extent of the spatial variation of the tube cross-section, stability analysis will be greatly simplified with respect to the case of the bar, and that computational complexity will also be independent of the tube cross-section[2]—these hunches are indeed borne out, and will be discussed in Section 9.1.5.

9.1.2 Cylinders and cones

Two special cases of the acoustic tube, often exploited in physical modeling sound synthesis for wind instruments, are the cylindrical tube and the conical tube. In the first case, one has a cross-sectional area function of the form $S(x) = 1$ (recall that the surface area function has been non-dimensionalized), and in the second, one has $S(x) = (1 + \epsilon x)^2$, for some constant $\epsilon > -1$, but normally assumed positive, so that the tube opens outward toward the bell. In both cases, Webster's equation can be reduced to the 1D wave equation, allowing the use of efficient digital waveguide structures.

For the cylinder, this is obvious from inspection of the first form given in (9.5). Because the digital waveguide in this case has been treated previously in Section 6.2.11, there is little more to be said here at the fundamental level, though there is, of course, a large body of techniques which has grown up around more sophisticated variations on the basic waveguide structure—see [334] and the references therein. A waveguide cylindrical tube model is a good first approximation to an instrument such as the clarinet, and was indeed the very first application of the technique in musical sound synthesis [327].

For the cone, the second form given in (9.5) is a better starting point. When $S(x) = (1 + \epsilon x)^2$, the curvature $a(x)$ vanishes, so that the 1D wave equation holds in the variable $\Phi = \sqrt{S}\Psi$. Thus, one is free to use a digital waveguide in order to propagate wavelike solutions which combine to form a physical variable Φ, and in an additional step, one must scale the output by $1/\sqrt{S}$ in order to obtain perceptually meaningful variables such as pressure and velocity. The existence of a digital waveguide form in the case of conical tubes expands considerably the utility of the waveguide technique in physical modeling synthesis, allowing for efficient solutions to be computed in the case of instruments such as the saxophone [314] and brass instruments with a conical bore. See [334] for much more on this topic.

[2] The astute reader here might argue that any conclusions reached regarding high-wavenumber limiting values of the phase and group velocity are in conflict with the very assumptions which lead to Webster's equation in the first place—and would be right! Here, as elsewhere, one must be very clear with oneself about which system one is taking to be "correct." This is a necessary choice if one wishes to arrive at any conclusions at all about the resulting numerical methods, which is, of course, the main goal here.

Much attention has been paid in the literature to tubes composed of cylindrical and conical segments, especially when each such segment is analyzed from an input/output (reflectance/transmission) point of view [150, 232]—this has carried over, in a natural way, to synthesis algorithms based on scattering structures [314, 368]. Here, however, where the emphasis is on direct simulation, it is entirely sufficient to work with an arbitrary cross-section $S(x)$. Indeed, for finite difference schemes, once one has made a choice of $S(x)$, the resulting algorithm is, from a programming standpoint, completely insensitive to this choice. For this reason, and also because cylindrical and conical waveguides are covered in great detail elsewhere, the present investigation of these special cases will end here, though some comments will appear in Section 9.2.4, in which finite difference schemes are related back to classic scattering structures used in articulatory speech synthesis.

9.1.3 Energy and boundary conditions

An energy balance follows for Webster's equation. Proceeding immediately to the case of Webster's equation defined over the unit interval $\mathcal{D} = \mathbb{U}$, one has, after taking an inner product with u_t,

$$\frac{d\mathfrak{H}}{dt} = \mathfrak{B} \qquad \text{with} \qquad \mathfrak{H} = \mathfrak{T} + \mathfrak{V} \qquad \mathfrak{T} = \frac{1}{2}\|\sqrt{S}\Psi_t\|_{\mathbb{U}}^2 \qquad \mathfrak{V} = \frac{\gamma^2}{2}\|\sqrt{S}\Psi_x\|_{\mathbb{U}}^2 \tag{9.7}$$

The energy $\mathfrak{H}$ is identical to that of the 1D wave equation, but in a norm weighted by S, and the boundary term $\mathfrak{B}$ is given by

$$\mathfrak{B} = \gamma^2 \left(S(1)\Psi_t(1,t)\Psi_x(1,t) - S(0)\Psi_t(0,t)\Psi_x(0,t)\right)$$

The boundary terms represent the power supplied to the tube at the boundaries. (Recalling the relationship between the derivatives of Ψ and the state variables p and u, from (9.6), both terms are of the form pressure $\times$ volume velocity. Note that if one were to work with Webster's equation written in terms of pressure instead of the potential function Ψ, this interpretation of the energy balance would be lost.)

In musical instrument models, it is conventional to deal with the left boundary condition, at $x = 0$, through an excitation mechanism—see Figure 9.1. For preliminary analysis of modes (see the next section), it is useful to leave the excitation condition at $x = 0$, and assume a closed tube end—in terms of the conditions discussed in Section 6.1.9, this amounts to $\Psi_x(0,t) = 0$, or a Neumann (zero-velocity) condition. At the radiating end, there are various levels of approximation. The simplest condition is of Dirichlet type (zero pressure), i.e.,

$$\Psi_t(1,t) = 0 \qquad \text{open end} \tag{9.8}$$

Such a condition is, clearly, lossless. A more realistic condition which includes the effects of loss and inertial mass has been introduced briefly in the case of the 1D wave equation in Section 6.1.9. In the present case of Webster's equation in the variable Ψ, a simple form may be written as

$$\Psi_x(1,t) = -\alpha_1\Psi_t(1,t) - \alpha_2\Psi(1,t) \qquad \text{open end with inertia and loss} \tag{9.9}$$

where the constants α_1 and α_2 depend on the tube parameters. In the case of a tube terminating on an infinite plane, for instance, the constants may be calculated [14] as

$$\alpha_1 = \frac{1}{2(0.8216)^2\gamma} \qquad \alpha_2 = \frac{L}{0.8216\sqrt{S_0 S(1)/\pi}} \tag{9.10}$$

which is useful in the case of, for example, vocal synthesis (these constants vary slightly in the literature—see [280] for different values). Another case of interest is the unflanged tube, for which the constants α_1 and α_2 take on moderately different values:

$$\alpha_1 = \frac{1}{4(0.6133)^2\gamma} \qquad \alpha_2 = \frac{L}{0.6133\sqrt{S_0 S(1)/\pi}} \tag{9.11}$$

Normally, in the literature, this condition is expressed as an impedance relationship between pressure and volume velocity. The above condition corresponds, in the frequency domain, to a first-order rational approximation [280] to a more general complex function, which is a little different from the usual polynomial form [14, 136]—in simulation, one must take great care when approximating such an impedance, which must remain positive real in order for it to correspond to a passive termination (which a radiating end of a tube certainly is!). This is a fact well known to electrical network specialists [30, 407], but less so to musical acousticians. See Problems 9.2 and 9.3. Notice that in the low-frequency limit (i.e., as γ becomes small), the simpler boundary condition (9.8) is recovered. This condition also marks the appearance of another necessary dimensionless parameter, namely $L/\sqrt{S_0 S(1)/\pi}$, which is precisely the ratio of tube length to the radius at the radiating end.

Under condition (9.9) at the right end, and the Neumann condition at the left end, the energy balance for the tube becomes

$$\frac{d}{dt}\left(\mathfrak{H} + \mathfrak{H}_\mathrm{b}\right) = -\mathfrak{Q}_\mathrm{b} \leq 0 \qquad \mathfrak{H}_\mathrm{b} = \frac{\gamma^2 \alpha_2 S(1)}{2}\left(\Psi(1,t)\right)^2 \geq 0 \qquad \mathfrak{Q}_\mathrm{b} = \gamma^2 \alpha_1 S(1)\left(\Psi_t(1,t)\right)^2 \geq 0$$

which again shows strictly dissipative behavior of a non-negative energy function, once the stored inertial energy for the termination is taken into account.

When the tube is driven at the left end (somehow!), the above energy balance is generalized to

$$\frac{d}{dt}\left(\mathfrak{H} + \mathfrak{H}_\mathrm{b}\right) = -\mathfrak{Q}_\mathrm{b} - \gamma^2 \Psi_t(0,t)\Psi_x(0,t) = -\mathfrak{Q}_\mathrm{b} + \gamma^3 p_\mathrm{in} u_\mathrm{in} \tag{9.12}$$

where $p_\mathrm{in}(t)$ and $u_\mathrm{in}(t)$ are the non-dimensional pressure and volume velocity at the driving end.

9.1.4 Modes

As the tube cross-section $S(x)$ is arbitrary, there is, in general, no closed-form expression for the modes of a 1D acoustic tube. Still, before continuing to the more complete scenario involving radiation and connections to an excitation mechanism, it is useful to gain some insights about how the modes vary with the area function $S(x)$. One important observation is that, regardless of the tube cross-section $S(x)$, or of the type of boundary condition, the number of degrees of freedom N_m (twice the number of modes whose frequencies occur over a band $f \in [0, f_\mathrm{s}/2]$, for a given f_s) will always be approximately

$$N_\mathrm{m} \approxeq 2 f_\mathrm{s}/\gamma$$

which is unchanged from the case of the 1D wave equation, as discussed in Section 6.1.11. See Figure 9.3. It also remains true that computational complexity for numerical methods for Webster's equation is independent of the bore profile as well, as will be shown shortly.

This behavior of the modes of the acoustic tube in the limit of high frequencies follows from the analysis of phase and group velocity in Section 9.1.1. As noted earlier, however, dispersion can be a large effect at low frequencies, and one might expect that, as a result, there will be a considerable variation in the positions of the low modal frequencies depending on $S(x)$, which is indeed true—see Figure 9.3. This variation is of considerable timbral importance in musical acoustics, particularly in the case of the human voice, in which case the frequencies of the low modes become associated with so-called formants. See Section 9.2.2.

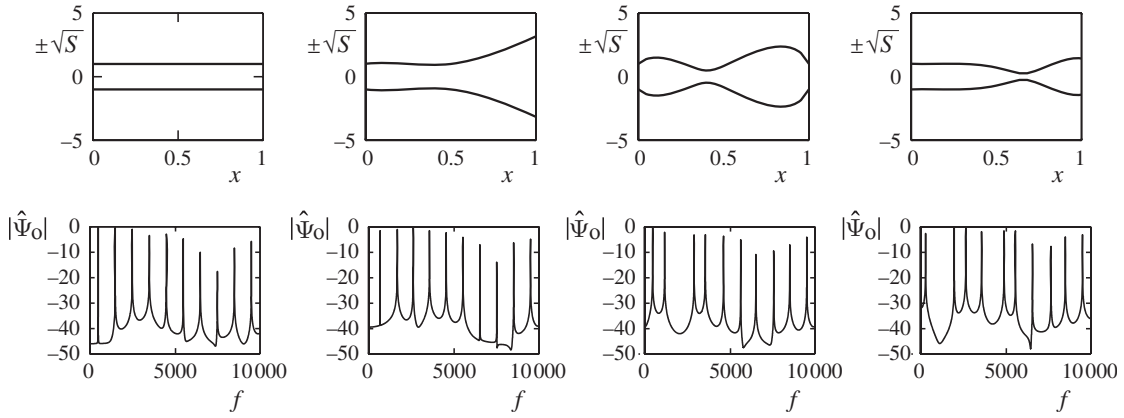


Figure 9.3 A variety of tube profiles (shown in terms of $\sqrt{S(x)}$, which scales with radius), at top, and typical distributions of harmonics for Webster's equation, under a Neumann (closed end) condition at $x = 0$ and a Dirichlet (open end) condition at $x = 1$. In this case, $\gamma = 2000$, and the average distance between harmonics is 1000 Hz, regardless of the tube profile, though a high degree of variation in the frequencies of the lowest modes is apparent. In this case, the tube is excited with a volume velocity impulse at the left end, and output is read at a point 0.2 of the way along the tube.

9.1.5 Finite difference schemes

As in the case of the bar of variable cross-section and the string of variable density (see Section 7.10), the number of choices of even simple varieties of finite difference schemes increases enormously, due to the spatial variation of the area function $S(x)$. It is again important to distinguish between the grid function S_l, which consists of sampled values of the continuous function $S(x)$ at locations $x = lh$, and a grid function $[S]_l$, which is some approximation to the continuous function $S(x)$, of an exact form yet to be determined—it is assumed that $[S]_l$ is a second-order accurate approximation to $S(x)$. A finite difference scheme for Webster's equation can then be written as

$$[S]\delta_{tt}\Psi = \gamma^2 \delta_{x+}\left((\mu_{x-}S)\,(\delta_{x-}\Psi)\right) \tag{9.13}$$

which may be expanded to give

$$\Psi_l^{n+1} = \frac{\lambda^2\,(S_{l+1}+S_l)}{2[S]_l}\Psi_{l+1}^n + \frac{\lambda^2\,(S_l+S_{l-1})}{2[S]_l}\Psi_{l-1}^n + \left(2 - \frac{\lambda^2\,(S_{l+1}+2S_l+S_{l-1})}{2[S]_l}\right)\Psi_l^n - \Psi_l^{n-1}$$

where $\lambda = \gamma k/h$ is again the Courant number for the scheme. It is straightforward to show that (9.13) is a second-order accurate approximation to Webster's equation. See Problem 9.4. The scheme above is similar to that presented in [381].

Energy analysis and boundary conditions

Energy analysis is direct, and similar to that carried out for the spatially varying systems in Section 7.10. Again, moving directly to the case $\mathcal{D} = \mathbb{U}_N$, the numerical energy balance is obtained, using familiar techniques, as

$$\delta_{t+}\mathfrak{h} = \mathfrak{b} \quad \text{with} \quad \mathfrak{h} = \mathfrak{t} + \mathfrak{v} \quad \mathfrak{t} = \frac{1}{2}\|\sqrt{[S]}\delta_{t-}\Psi\|_{\mathbb{U}_N}^2 \quad \mathfrak{v} = \frac{\gamma^2}{2}\langle(\mu_{x+}S)\delta_{x+}\Psi,\, e_{t-}\delta_{x+}\Psi\rangle_{\underline{\mathbb{U}_N}} \tag{9.14}$$

and where the boundary term is

$$\mathfrak{b} = \gamma^2\,(\delta_{t\cdot}\Psi_N)\,(\mu_{x+}S_N)\,(\delta_{x+}\Psi_N) - \gamma^2\,(\delta_{t\cdot}\Psi_0)\,(\mu_{x-}S_0)\,(\delta_{x-}\Psi_0)$$

These expressions are completely analogous to those which occur in the energy balance for the continuous system, in (9.7). Note, however, that the boundary term depends on values of the grid function S_l which are outside of $\mathbb{U}_N$—these values must be set judiciously.

First, consider boundary conditions corresponding to a closed tube end (left) and a radiating boundary (right):

$$\delta_{x-}\Psi_0 = 0 \qquad \delta_{x+}\Psi_N = -\alpha_1\delta_{t\cdot}\Psi_N - \alpha_2\mu_{t\cdot}\Psi_N \tag{9.15}$$

Under these conditions,

$$\mathfrak{b} = -\gamma^2\alpha_1(\delta_{t\cdot}\Psi_N)^2 - \frac{\gamma^2\alpha_2}{2}\delta_{t+}(\mu_{t-}\Psi_N)^2$$

which are strictly dissipative, such that

$$\delta_{t+}(\mathfrak{h} + \mathfrak{h}_\mathrm{b}) = -\mathfrak{q}_\mathrm{b} \qquad \mathfrak{h}_\mathrm{b} = \frac{\gamma^2(\mu_{x+}S_N)\alpha_2}{2}\mu_{t-}(\Psi_N)^2 \geq 0 \quad \mathfrak{q}_\mathrm{b} = \gamma^2(\mu_{x+}S_N)\alpha_1(\delta_{t\cdot}\Psi_N)^2 \geq 0$$

Again, as in the context of lumped element/string modeling as described in Section 7.7, and following Rule of Thumb #3 given on page 192, a semi-implicit approximation to the radiating boundary condition has been employed and leads immediately to the non-negativity guarantees above.

Given that the scheme is strictly dissipative under boundary conditions (9.15), in order to arrive at a stability condition, it is as before necessary to find a condition under which the numerical energy is non-negative. The next step is to bound the potential energy term using various identities given in Chapter 5, and especially (5.32):

$$\mathfrak{v} = \frac{\gamma^2}{2}\langle(\mu_{x+}S)\delta_{x+}\Psi, e_{t-}\delta_{x+}\Psi\rangle_{\underline{\mathbb{U}_N}} \geq -\frac{\gamma^2k^2}{8}\|\sqrt{\mu_{x+}S}\delta_{x+}\delta_{t-}\Psi\|^2_{\underline{\mathbb{U}_N}} \geq -\frac{\lambda^2}{2}\|\sqrt{\mu_{xx}S}\delta_{t-}\Psi\|^2_{\mathbb{U}_N}$$

Now, a bound for the total energy $\mathfrak{h}$ is

$$\mathfrak{h} \geq \frac{1}{2}\|\sqrt{[S]}\delta_{t-}\Psi\|^2_{\mathbb{U}_N} - \frac{\lambda^2}{2}\|\sqrt{\mu_{xx}S}\delta_{t-}\Psi\|^2_{\mathbb{U}_N}$$

which is a quadratic form in the grid function $\delta_{t-}\Psi$. One arrives immediately at the following condition:

$$\lambda \leq \min\left(\sqrt{\frac{[S]}{\mu_{xx}S}}\right) \approxeq 1$$

Note that the stability condition is approximately independent of the area function $S(x)$; this is quite distinct from the cases of the string of variable density, or the bar of variable cross-sectional area. The reason for this distinction is that the limiting wave speed at high wavenumbers remains γ everywhere within the tube, though dispersion does intervene—see the comments in Section 9.1.1.

The numerical boundary conditions given in (9.15) are one possibility; through more extensive analysis (see Problem 9.5), centered, provably stable conditions may be derived:

$$\delta_{x\cdot}\Psi_0 = 0 \qquad \delta_{x\cdot}\Psi_N = -\alpha_1\delta_{t\cdot}\Psi_N - \alpha_2\mu_{t\cdot}\Psi_N \tag{9.16}$$

A special form

The approximate stability condition above may be made exact under a particular choice of $[S]$, i.e.,

$$[S] = \mu_{xx}S \qquad \longrightarrow \qquad \lambda \leq 1$$

This setting for $[S]$ will thus be used in the remainder of this chapter. In this case, it is also worth looking at the special form of the recursion when $\lambda = 1$:

$$\Psi_l^{n+1} = \frac{2\,(S_{l+1} + S_l)}{S_{l-1} + 2S_l + S_{l+1}} \Psi_{l+1}^n + \frac{2\,(S_l + S_{l-1})}{S_{l-1} + 2S_l + S_{l+1}} \Psi_{l-1}^n - \Psi_l^{n-1} \tag{9.17}$$

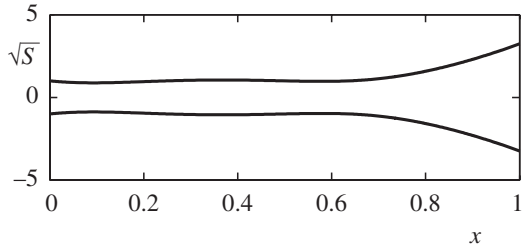
This form, for which the value of Ψ at the grid point l, n has disappeared from the recursion, hints at the special case of the digital waveguide, as discussed in Section 6.2.11. And indeed it should, for this scheme may be viewed in terms of the Kelly–Lochbaum speech synthesis model, probably the first example of a scattering-based numerical method. See Section 9.2.4.

Numerical modes and accuracy

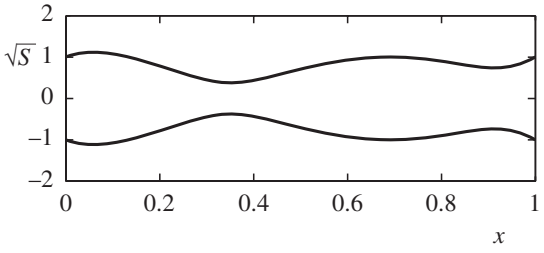
One of the interesting things about scheme (9.13) is that, in contrast with schemes for other spatially varying systems such as the arched bar (see Section 7.10), which are naturally very dispersive and thus require a good deal of improvement at the algorithmic level, it performs remarkably well.

As an example, consider the representative tube cross-sections shown in Figure 9.4. For a tube of relatively mild curvature, representing a wind instrument body, and for a typical value of γ, the modal frequencies calculated by scheme (9.13) are nearly exact; the deviation is far below the threshold of pitch discrimination. For larger values of γ, and for a more extreme cross-section, the scheme still performs very well. So good are the results that, in fact, it makes little sense to look for a better scheme, given that the inaccuracy in the assumptions underlying Webster's equation to begin with is probably much larger than the numerical inaccuracy introduced by the scheme.

Most interesting is the observation that the dispersive effects are evenly distributed over the audio spectrum—modal detuning does not deteriorate as one moves into the upper frequency range,



γ = 500



γ = 2000

Mode #	Exact Freq.	Scheme (9.13) Freq.	Cent Dev.
1	161.8	161.8	0.2
2	459.1	459.0	−0.4
3	656.7	656.6	−0.4
4	890.6	890.5	−0.2
12	2880.3	2879.9	−0.2
13	3129.9	3129.4	−0.3

Mode #	Exact Freq.	Scheme (9.13) Freq.	Cent Dev.
1	428.9	436.5	30.3
2	1072.0	1092.1	32.2
3	2870.0	2851.7	−11.1
4	3829.9	3785.5	−20.1
12	11597.9	11483.5	−17.2
13	12586.8	12456.4	−18.0

Figure 9.4 Tube profiles, and comparisons between exact modal frequencies and numerical modes resulting from the use of scheme (9.13), running at 44.1 kHz, and for a value of λ chosen as close to 1 as possible. At left, a tube profile grossly characteristic of wind and brass instruments with a relatively low value of γ, and, at right, a more extreme profile resembling a vocal tract configuration, with a higher value of γ. In each case, a zero-velocity (Neumann) condition is used at the tube's left end and a zero-pressure (Dirichlet) condition at its right end. ("Exact" values for the modal frequencies are obtained by examining the same scheme for a very small time step, typically $k \leq 10^{-6}$.)

as it does in, for example, schemes for stiff systems such as bars (see Section 7.1.4). Qualitatively, the reason for this behavior may be viewed in terms of the limiting behavior of the phase and group velocities, as described in Section 9.1.1—physical dispersion (and thus mode detuning) is greatest at low frequencies, and in the high-frequency limit the solution behaves essentially like that of the 1D wave equation. But the difference scheme (9.13), from the standpoint of conventional accuracy analysis, performs well at low frequencies, and, at high frequencies, becomes equivalent to scheme (6.34) for the 1D wave equation, which is numerically exact, at least when $\lambda = 1$. These competing tendencies tend to cancel dispersion in a way that they do not in the case of a stiff system such as the ideal bar, where physical dispersion and numerical dispersion reinforce one another.

9.2 The vocal tract and speech synthesis

Vocal synthesis was, in some respects, the starting point for all digital sound synthesis; indeed, the first physical models of the vocal tract [201] were devised at roughly the same time that Max Mathews was performing his early experiments with computerized sound. From a modern perspective, one of the nice things about vocal synthesis is that it is very cheap, computationally; in fact, it is perhaps the least costly application that appears in this book. Even in the sluggish Matlab programming language, vocal synthesis using a 1D model is now far faster than real time—see the code example in Section A.10.

The body of work in speech synthesis is many times larger than that concerned with the totality of musical sound synthesis applications, and there is no hope of doing justice to this huge topic in a short section. The focus here is on methods derived from physical considerations of the dynamics of the vocal tract. The electrical transmission line analogy to the acoustic tube (see, e.g., [123]) has been and still is a popular means of representing the vocal tract, and has had a profound influence on the types of simulation strategies which have been employed—see, e.g., [280, 231]. The scattering interpretation, first introduced by Kelly and Lochbaum in 1962 [201], fits in very easily to the digital waveguide framework (indeed, it is a direct antecedent), and was one of the early applications of the technique to singing voice synthesis by Cook [90]. Waveguide meshes have also been used in 2D models of the vocal tract [247, 246].

Of course, it is always possible to view a scattering method in terms of finite difference methods, without employing wave concepts explicitly—see, e.g., [41], and in the present context of vocal synthesis, the article by Strube [345]—though relatively little work has been done on direct simulation in the absence of a circuit/scattering/transmission line point of view. Some exceptions are, for instance, [226], the recent work of van den Doel and Ascher [381], and, especially, the early treatment of Portnoff [273], far ahead of its time in 1973, which deals with Webster's equation directly, including many features discussed here, such as time variation and loss, leading to a finite difference approximation. On the other hand, modern articulatory speech systems (see, e.g., [125]) require a more detailed model of the vocal tract in 2D or 3D, and the transmission line analogy breaks down. Direct methods such as FDTD [96] or finite elements are really the only option—see [164] and the references therein.

9.2.1 Glottal excitation

In the discussion of Webster's equation in the previous section, the excitation is left unspecified. In the synthesis of vocal sounds, there are various ways of generating input. One, probably the most scientifically rigorous, is to make use of a model of the glottis itself. This is a large open research area, especially in the speech community—see, e.g., the early paper by Ishizaka and Flanagan [178] which set the stage for a large body of subsequent work. Usually, the glottis mechanism is modeled as a low-order mass–spring network, necessarily involving collisions; as such, most models are not extremely different from typical models of the reed or lips in woodwind and brass

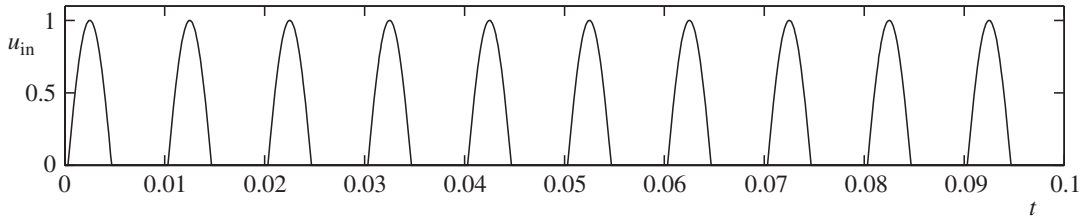


Figure 9.5 A crude synthetic glottal input waveform u_{in}, here with a period of 0.01s, and with a 45% duty cycle. This particular example has been generated by clipping of a sine wave.

instruments, which will appear shortly in Section 9.3. A standard approach in speech synthesis, and one which will be adopted here, is to assume minimal coupling between the glottis and the vocal tract [280], and to simply drive the tube with a specified waveform, normally expressed as an input volume velocity. There are many waveform types which pepper the literature—see, e.g., [124] for an overview; for the present purposes, the exact form of the input is not extremely important, but it should behave as a periodic series of pulses, with each pulse occupying a given fraction of the period (the duty cycle), corresponding to intervals during which the vocal folds are open. See Figure 9.5. Generally, the shorter the duty cycle, the more wide band the input spectrum. This is a prime example of the source-filter approach to sound synthesis—see Section 1.1.2.

Such a signal is useful in reproducing voiced (i.e., pitched) sounds; for unvoiced sounds, a noise signal can be employed, as a qualitative approximation to turbulence effects at the glottal opening. Such choices, however important they may be in practice, have no bearing on the operation or analysis of a source-filter physical model of the vocal tract, and will not be explored further here—for more information, see any standard text on speech synthesis [280], and refer to Programming Exercise 9.1, which explores some of the variations in source signal models, including vibrato effects, and the introduction of an amplitude envelope; these features are crucial, especially in singing voice synthesis.

In order to input any such waveform u_{in} into scheme (9.13), which is written in terms of Ψ, one may use (9.6) and then set, at grid point $l = 0$, using a time series u_{in},

$$\delta_{x\cdot}\Psi_0 = -u_{\text{in}} \qquad \longrightarrow \qquad \delta_{tt}\Psi_0 = \frac{2\gamma^2}{h}\left(\delta_{x+}\Psi_0 + \frac{\mu_{x-}S_0 u_{\text{in}}}{\mu_{xx}S_0}\right)$$

9.2.2 Formants

The well-known effect of the tube on the input excitation is to impose a formant structure in the frequency domain—see Figure 9.6. The formants are the tube resonances—due to the radiation condition, they possess a finite bandwidth and serve to emphasize spectral regions, allowing the listener to distinguish among various vowel sounds. For more on formant structure, and various vowel configurations, see Programming Exercise 9.2.

9.2.3 Wall vibration and loss

Webster's equation is somewhat more of an idealization in the case of the vocal tract than when applied to wind instruments; loss is an important consideration, especially when it comes to accurately reproducing formant bandwidths. The major source of loss, namely radiation, is already taken care of by the boundary condition (9.9). There are two other contributions, however, which will have a noticeable effect [339]. First, there is a viscous boundary layer at the walls of the tube; this leads to a root–frequency loss dependence, and is quite difficult to model properly using a time domain model—fractional derivatives intervene in an extension of Webster's equation. Such loss effects have indeed been approached in the context of wind instrument synthesis, in the scattering

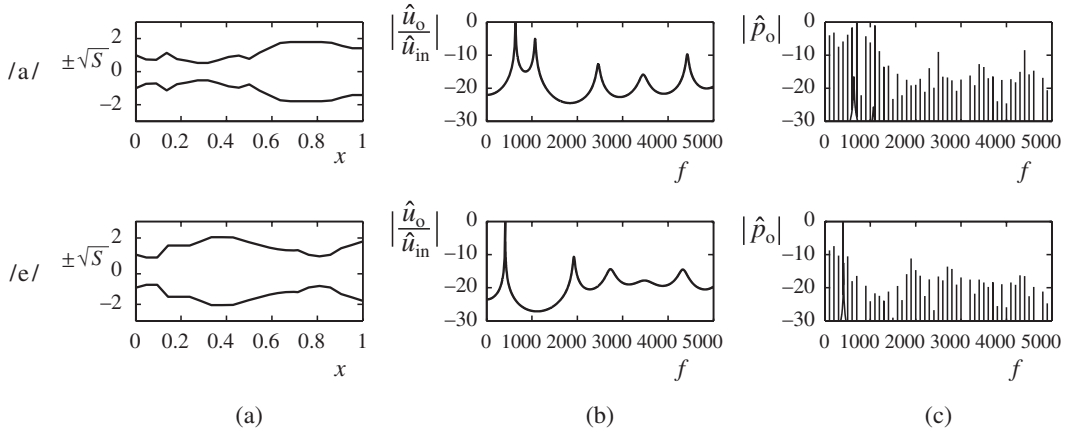


Figure 9.6 Formant structure. (a) Tube cross-sections corresponding to vocal tract configurations for the vowels /a/ and /e/ (Russian, adapted from measured data given in [280].). (b) Formant structure, here shown as a plot of the acoustic output velocity/input velocity transfer function $|\hat{u}_o/\hat{u}_{in}|$, and (c) spectrum of the output pressure waveform $|\hat{p}_o|$. In both cases, output is calculated using scheme (9.13) operating at 44.1 kHz. In (c), the input waveform is of the type shown in Figure 9.5, with a fundamental frequency of 100 Hz and with a 50% duty cycle.

framework [166, 2], but, as the topic is a rather large one, it will not be discussed further here. The other effect, that of vibration and damping in the tube walls themselves, is much more amenable to a finite difference treatment, as was carried out early on by Portnoff [273] and more recently by others [381].

A simple model of locally reacting walls [133] involves Webster's equation, coupled to another equation in a separate dimensionless function $w(x,t)$, which scales with the fractional change in local tube radius. In scaled form, and for a tube of cylindrical cross-section, the system may be written as

$$S\Psi_{tt} = \gamma^2 (S\Psi_x)_x - \epsilon S^{1/4} w_t \qquad w_{tt} + 2\sigma_0 w_t + \omega_0^2 w = \epsilon S^{1/4} \Psi_t \tag{9.18}$$

Here, σ_0 and ω_0 are damping and fundamental frequency parameters for the vocal tract walls, and ϵ, a coupling coefficient, may be written as

$$\epsilon = c\sqrt{\frac{2\rho}{M}} \left(\frac{\pi}{S_0}\right)^{1/4}$$

where ρ is the density of air, c the speed of sound, and M the mass per unit area of the vocal tract walls. Notice in particular that the equation in w involves time derivatives only, simplifying analysis enormously. System (9.18) may easily be shown to be strictly dissipative—see Problem 9.6.

A finite difference scheme for this coupled system follows immediately:

$$\mu_{xx} S \delta_{tt} \Psi = \gamma^2 \delta_{x+} ((\mu_{x-} S) \delta_{x-} \Psi) - \epsilon S^{1/4} \delta_{t\cdot} w \qquad \delta_{tt} w + 2\sigma_0 \delta_{t\cdot} w + \omega_0^2 w = \epsilon S^{1/4} \delta_{t\cdot} \Psi \tag{9.19}$$

This scheme, though apparently implicit, may be easily resolved into an explicit update in the grid functions w and Ψ—see Problem 9.7.

The main effect of using such a model is easily seen by examining the change to the formant structure, as shown in Figure 9.7. The lower formants are broadened, leading to a significant perceptual difference in resulting sound output, especially when the input pitch is changing, or under time-varying vocal tract conditions—in either case, if the lower formants are too sharply peaked, the resulting synthetic sound will ring unnaturally as frequency components approach a peak (or vice versa). See Programming Exercise 9.3.

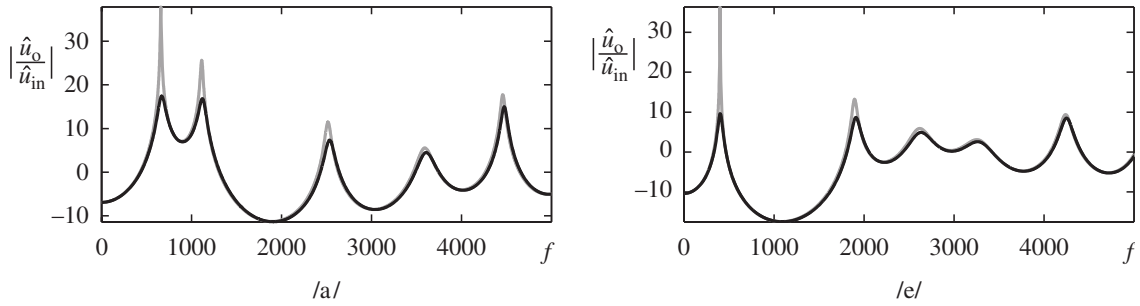


Figure 9.7 Formant structure, in decibels, for two vocal tract configurations (as illustrated in Figure 9.6), with a wall damping model (in black) and without (in grey). For the case of the wall model, curves are generated using scheme (9.19) for system (9.18), with parameters $\omega_0 = 0$, $\sigma_0 = 8125$, and $\epsilon = 2718$, and the sample rate is $f_s = 44.1\,\text{kHz}$.

9.2.4 Scattering methods and finite difference schemes

As mentioned earlier (see Section 1.2), the Kelly–Lochbaum speech synthesis algorithm was perhaps the first instance of a sound-producing physical model. What is more, it is a scattering method, allying it to other methods such as digital waveguides and wave digital filters (see Section 1.2.4) which are heavily used in modern physical modeling synthesis for a great variety of musical instrument types. As has been shown in Section 6.2.11, however, digital waveguides may be rewritten as finite difference methods, and the same is true of scattering methods for voice synthesis.

A typical scattering structure may be arrived at in the following way. First, assume that the tube profile is approximated by a series of cylindrical segments, of length h, such that $N = 1/h$ is an integer (again, the tube is assumed to be of length 1). The left and right ends of the lth tube are located at $x = lh$ and $x = (l+1)h$, for $l = 0, \ldots, N-1$. See Figure 9.8. In each such tube, the system (9.4) is satisfied, with a constant surface area $[S]_{l+\frac{1}{2}}$, which is some approximation to $S(x)$ at the tube center. In this constant-coefficient case, the system may be reduced to the 1D wave equation, which admits traveling wave solutions, in both pressure and volume velocity. Thus,

$$p(x,t) = p^{(+)} + p^{(-)} \qquad \text{and} \qquad u(x,t) = u^{(+)} + u^{(-)} \tag{9.20}$$

where, as before, the superscripts $(+)$ and $(-)$ indicate solutions traveling to the right and left, respectively, with speed γ, according to

$$p^{(+)}(x,t) = p^{(+)}(x - \gamma\xi, t - \xi) \quad p^{(-)}(x,t) = p^{(-)}(x + \gamma\xi, t - \xi)$$
$$u^{(+)}(x,t) = u^{(+)}(x - \gamma\xi, t - \xi) \quad u^{(-)}(x,t) = u^{(-)}(x + \gamma\xi, t - \xi)$$

for some time-like parameter ξ. But the two sets of wavelike solutions are not independent—they are related by

$$u^{(\pm)} = [S]_{l+\frac{1}{2}}\, p^{(\pm)} \qquad \text{in } l\text{th tube}$$

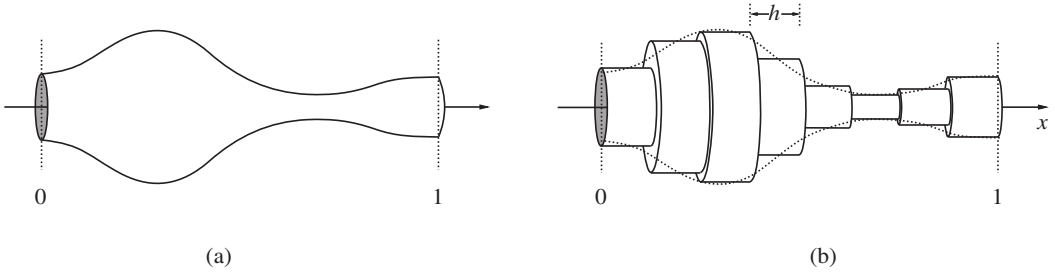


Figure 9.8 (a) An acoustic tube, and (b) a piecewise cylindrical approximation, where the length of each tube segment is $h = 1/N$, for some integer N (here, $N = 8$).

Here, $[S]_{l+\frac{1}{2}}$ serves as the admittance of the lth tube. Thus, one may choose to propagate pressure variables alone (though velocity variables are often used [280]).

Moving to the discrete setting, because the tube length is h, and the wave speed is uniformly γ over all tubes, one may take the unit delay to be $k = h/\gamma$. The wave variables, at times which are integer multiples n of k at the left and right ends of the lth tube, may then be labeled according to the junction (l) at which they are either impinging or scattered, and the side of the junction where they lie, as $p^{(+),n}_{l,\text{left}}$, $p^{(-),n}_{l,\text{left}}$, $p^{(+),n}_{l,\text{right}}$, and $p^{(-),n}_{l,\text{right}}$. A diagram is in order at this point—see Figure 9.9(a). In discrete form, wave propagation may be written as

$$p^{(+),n+1}_{l,\text{left}} = p^{(+),n}_{l-1,\text{right}} \qquad p^{(-),n+1}_{l,\text{right}} = p^{(+),n}_{l+1,\text{left}} \tag{9.21}$$

purely in terms of what are now grid functions indexed by n and l.

The scattering operation follows, as in the case of wave digital filters, from equal pressures and velocities at the junctions between tubes. Between the lth and $(l+1)$th tubes, at junction index l, and at time step n, one must have

$$p^{(+),n}_{l,\text{left}} + p^{(-),n}_{l,\text{left}} = p^{(+),n}_{l,\text{right}} + p^{(-),n}_{l,\text{right}} \qquad [S]_{l-\frac{1}{2}} \left(p^{(+),n}_{l,\text{left}} - p^{(-),n}_{l,\text{left}} \right) = [S]_{l+\frac{1}{2}} \left(p^{(+),n}_{l,\text{right}} - p^{(-),n}_{l,\text{right}} \right)$$

This pair of equations may be rewritten in scattering form, relating junction outputs to inputs in terms of a single reflection parameter r_l:

$$\begin{bmatrix} p^{(-),n}_{l,\text{left}} \\ p^{(+),n}_{l,\text{right}} \end{bmatrix} = \underbrace{\begin{bmatrix} -r_l & 1+r_l \\ 1-r_l & r_l \end{bmatrix}}_{\mathbf{S}_l} \begin{bmatrix} p^{(+),n}_{l,\text{left}} \\ p^{(-),n}_{l,\text{right}} \end{bmatrix} \qquad r_l = \frac{[S]_{l+\frac{1}{2}} - [S]_{l-\frac{1}{2}}}{[S]_{l+\frac{1}{2}} + [S]_{l-\frac{1}{2}}} \tag{9.22}$$

Here, the scattering operation at junction l, again a 2×2 matrix multiplication, is indicated by $\mathbf{S}_l$. The complete structure, involving the scattering operation above as well as the delays implied by (9.21), is illustrated in Figure 9.9(b).

Sometimes, in the setting of vocal synthesis, this scattering structure is extended to the time-varying case (see Section 9.2.5) by allowing the reflection coefficients r_l to depend on time, i.e., one uses r^n_l. It is worth noting that the nice stability properties of scattering networks disappear in this case [345], although if the rate of time variation is slow, there is little risk of such instability appearing. One idea is to make use of power-normalized scattering, as described in

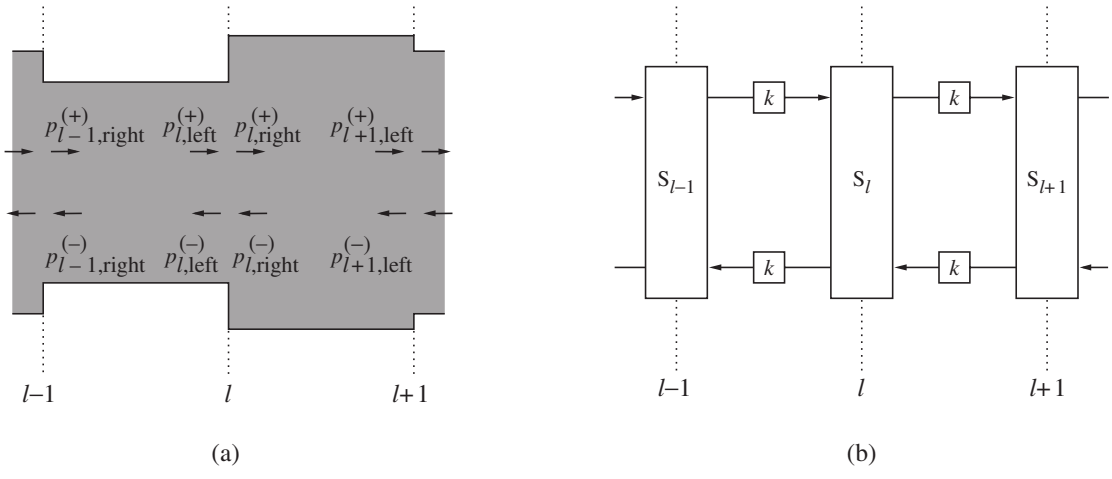


Figure 9.9 (a) A sequence of cylindrical tube sections, illustrating the traveling wave decomposition of the solution, and (b) the discrete scattering method, consisting of scattering junctions (labeled $\mathbf{S}_l$) interleaved with pairs of digital unit sample (duration k) delay lines.

Section 3.3.3 in the context of wave digital filters, rendering each scattering matrix orthogonal, regardless of time variation in r_l. The problem here, however, is that such a lossless structure cannot possibly solve Webster's equation, for the simple reason that Webster's equation is *not* lossless in the time-varying case. The deeper issue is that, while scattering methods allow for very robust algorithm design (indeed, numerical energy conservation, described at great length in this book, is an in-built property of such methods), it may not always be clear what system they solve when one begins to fiddle with them in the absence of an underlying physical model!

Equivalence with a difference scheme

Just as in the case of digital waveguides (see Section 6.2.11), it is not difficult to show that this scattering structure is indeed equivalent to a finite difference scheme. Beginning from the definition of the pressure at the junction at location l (say on the left side), one may proceed as follows:

$$
\begin{aligned}
p_l^{n+1} &\overset{(9.20)}{=} p_{l,\text{left}}^{(+),n+1} + p_{l,\text{left}}^{(-),n+1} \\
&\overset{(9.22)}{=} (1-r_l)\, p_{l,\text{left}}^{(+),n+1} + (1+r_l)\, p_{l,\text{right}}^{(-),n+1} \\
&\overset{(9.21)}{=} (1-r_l)\, p_{l-1,\text{right}}^{(+),n} + (1+r_l)\, p_{l+1,\text{left}}^{(-),n} \\
&\overset{(9.20)}{=} (1-r_l)\, p_{l-1}^{n} + (1+r_l)\, p_{l+1}^{n} - (1-r_l)\, p_{l-1,\text{right}}^{(-),n} - (1+r_l)\, p_{l+1,\text{left}}^{(+),n} \\
&\overset{(9.21)}{=} (1-r_l)\, p_{l-1}^{n} + (1+r_l)\, p_{l+1}^{n} - (1-r_l)\, p_{l,\text{left}}^{(-),n-1} - (1+r_l)\, p_{l,\text{right}}^{(+),n-1} \\
&\overset{(9.20)}{=} (1-r_l)\, p_{l-1}^{n} + (1+r_l)\, p_{l+1}^{n} - p_l^{n-1}
\end{aligned}
$$

Given the definition of r_l, from (9.22), this may be rewritten in terms of the area function $[S]$ as

$$
p_l^{n+1} = \frac{2[S]_{l+\frac{1}{2}}}{[S]_{l-\frac{1}{2}} + [S]_{l+\frac{1}{2}}}\, p_{l+1}^{n} + \frac{2[S]_{l-\frac{1}{2}}}{[S]_{l-\frac{1}{2}} + [S]_{l+\frac{1}{2}}}\, p_{l-1}^{n} - p_l^{n-1}
$$

and, when $[S]$ is defined as $[S]_{l+\frac{1}{2}} = (S_l + S_{l+1})/2$, this is exactly the special form of scheme (9.13), when $\lambda = 1$, as given in (9.17). It is, however, expressed here in terms of a pressure variable instead of Ψ, the potential.

Computational requirements

There is very little difference, in terms of computational complexity, between the Kelly–Lochbaum structure and difference scheme (9.17). Both require $2N$ units of memory—in the difference scheme, to hold the previous two grid functions in Ψ, and for the scattering structure, to hold the values in the delay elements as per Figure 9.9(b). The operation count is also very similar, depending on the implementation—see Problem 9.8.

It is important to point out, however, that the Kelly–Lochbaum structure requires that an integer number of segments of length h subdivide the unit interval, whereas scheme (9.13) in general does not; this allows for a much better tuning of formant frequencies for the difference scheme. It is possible to extend a scattering structure, using extra self-terminated delay elements at each junction, in order to account for such effects (see [41]), but at the expense of introducing three-port scattering. A "signal processing" approach, popular in digital waveguide modeling, is to make use of fractional delay terminations to the scattering structure, which are essentially all-pass structures—see, e.g., [372, 215].

A compromise between performance and efficiency, for scattering methods, is to use a more crude approximation to the vocal tract shape, employing multiple sample delay waveguides and

fewer scattering junctions, where all calculation is consolidated—see, e.g., [88], where such a technique was employed in one of the first instances of real-time singing voice synthesis; though a useful structure, it should be kept in mind that, even at high audio sample rates such as 44.1 kHz, the total number of tube segments necessary even for a full simulation is roughly 20; this leads to very low computational costs by today's standards. The need for such efficient structures in vocal synthesis is thus much less urgent than it once was. When applied to wind instrument synthesis, which involves much longer (and less dispersive) tubes, it is a much more justifiable simplification, and a very powerful one indeed.

9.2.5 Time-varying vocal tract models

The production of any interesting vocal sounds necessarily requires some temporal variation of the properties of the vocal tract—in standard models, the variation is consolidated in the surface area function S, which becomes a function $S(x, t)$ of space and time. (Another type of variation, and one which will not be explored here, is that of change in length of the vocal tract itself [341, 371, 381].) This is the first example in this book of a time-varying system, though it remains linear—see the classification of musical systems given in Section 5.1.1. In some ways, the analysis of numerical methods for linear time-varying systems can be more difficult than for fully nonlinear systems (at least those of autonomous type, for which time dependence only intervenes through the dependent variable itself). The reason for this is that the conservation properties are obscured by the assumption of time-varying coefficients—in reality, such a system is being driven by an (unmodeled) source. As such, energy is injected into the system. In practice, however, the rate of change of the area function is quite slow, and numerical schemes do perform as expected. Because the system remains linear, there is no issue of uniqueness/existence for difference schemes as occurs in the nonlinear setting.

A time-varying generalization of the first-order system (9.4) has been presented by various authors [280, 344]. In non-dimensionalized form, it is

$$(Sp)_t = -\gamma u_x \qquad \left(\frac{u}{S}\right)_t = -\gamma p_x$$

where scaling has been carried out exactly as in the time-invariant case. (One must be careful here in the choice of S_0, the reference area, which is not $S(0, t)$, but a constant such as $S(0, 0)$.) A time-varying version of Webster's equation follows as

$$(S\Psi_t)_t = \gamma^2 (S\Psi_x)_x \tag{9.23}$$

See [381] for a different form of Webster's equation in the time-varying case, as well as a description of finite difference strategies. An energy analysis of this equation is useful in order to see the effect of the time variation, which may be viewed as a driving term—see Problem 9.9. There are, predictably, many possible discretizations of (9.23). Here is one rather well-behaved explicit form:

$$\delta_{t+} \left((\mu_{t-} \mu_{xx} S) \delta_{t-} \Psi \right) = \gamma^2 \delta_{x+} \left((\mu_{tt} \mu_{x-} S) \delta_{x-} \Psi \right)$$

Stability, for this scheme, is difficult to show, at least using the present energy framework—frozen coefficient analysis [342] is one other option. There are other, safer, and provably stable schemes (see Problem 9.10), but this one suffices for a demonstration.

Formant transitions

The interest in a time-varying model is obvious—one may begin to approximate intelligible speech, or singing. Loosely speaking, one may consider the resonances of the vocal tract, or formants, to

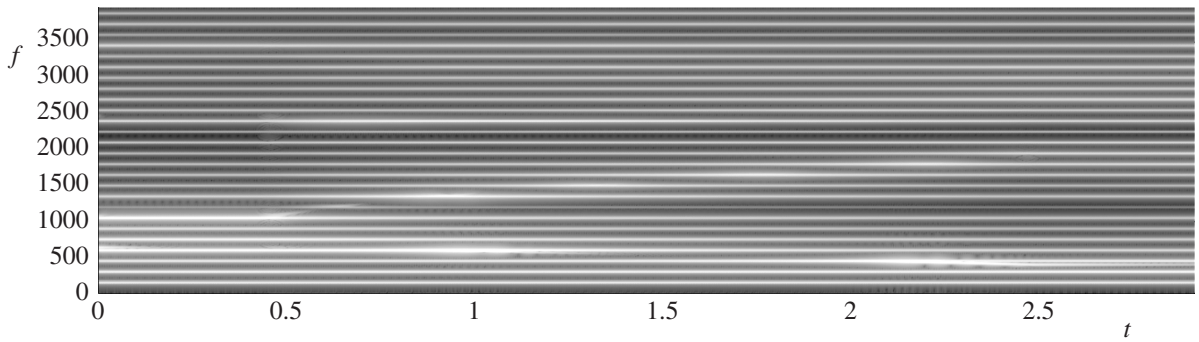


Figure 9.10 Spectrogram of output from the time-varying scheme (9.23), under a transition from an /a/ configuration to an /e/, effected between $t = 0.5$ and $t = 2.5$ s. The tube profiles shown in Figure 9.6 are interpolated linearly. The sample rate is 44.1 kHz, and the input is a waveform of the type shown in Figure 9.5, with a fundamental frequency of 120 Hz.

move when the tract configuration is changed—see Figure 9.10 for an example. Much has been written on this topic—see, e.g., [280]. While, from a numerical standpoint, there is not a lot to say here, it should be stressed that the precise trajectory of the change in vocal tract configurations is quite important in determining its perceptual correlate, the formant transitions. In the example above, the transition is effected in the most crude manner imaginable, i.e., by linearly interpolating between two configurations over the duration of the transition, and the result is only marginally natural sounding. The interested reader may wish to experiment with this—see Programming Exercise 9.4.

9.3 Reed wind instruments

Physical modeling synthesis for reed instruments has seen a relatively large amount of activity, and there are a great variety of methods—all, however, owe a great deal to the theoretical treatment of the musical self-sustained oscillator due to McIntyre et al. [236], which cast the standard model of the reed/bore interaction in a form which suggested efficient time domain digital realizations. The reed is modeled as a lumped oscillator and the bore in terms of its impulse response (or Green's function). Such time domain techniques (to be distinguished from the time–space methods employed here—see the footnote on page 17), written as a lumped nonlinear system coupled to an input/output model of the bore characterized by an impulse response, have been widely investigated—see, e.g., [25], and, for applications in synthesis, [158, 159].

Slightly later, and probably independently, Smith realized that such an impulse response could be efficiently generated for certain bore profiles by using delay lines—thus the first application of digital waveguides to synthesis [327]. For a good introduction to wave-based methods such as wave digital filters and digital waveguides for wind instruments, see [390, 388] and the references which appear in Section 1.2.3.

More closely allied to the finite difference exposition which follows here, adapted from [45], is the "K"-variable formulation of Avanzini and Rocchesso [15], as well as purely time domain characterizations of the reed mechanism using wave digital filters [41].

9.3.1 The single reed

The reed mechanism has been very briefly introduced, in an abstract form, in Section 4.3.2. A relatively standard approach, for the single reed, is to treat it as a lumped linear oscillator,

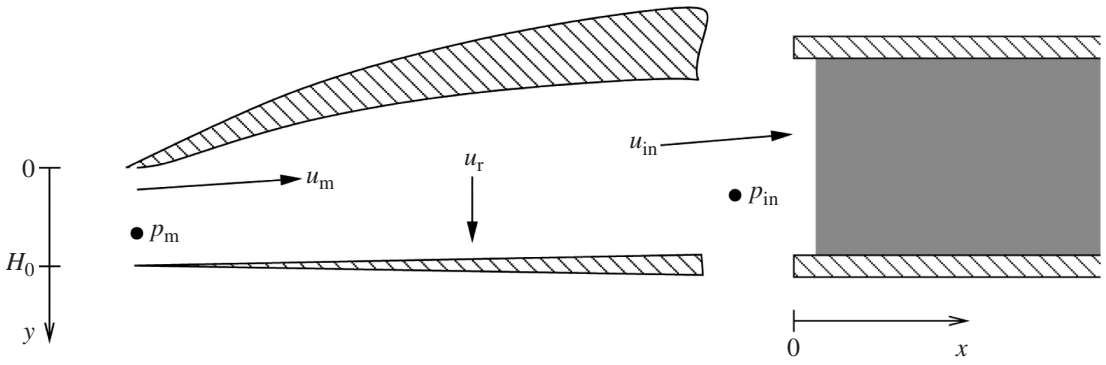


Figure 9.11 Illustration of a single reed model.

driven by a pressure drop across the mouthpiece—see Figure 9.11. The equation of motion, in dimensional form, is

$$\frac{d^2y}{dt^2} + 2\sigma_0\frac{dy}{dt} + \omega_0^2(y - H_0) - \frac{\omega_1^{\alpha+1}}{H_0^{\alpha-1}}\left(|[y]^-|\right)^\alpha = -\frac{S_r p_\Delta}{M_r} \tag{9.24}$$

Here, $y(t)$ is the displacement of the reed relative to an equilibrium position H_0, M_r is the reed mass, S_r an effective surface area of the reed, ω_0 the resonant frequency, and σ_0 a damping parameter. The nonlinear term involving the coefficient ω_1 models the collision of the reed with the mouthpiece. It becomes active when $y < 0$, and acts as a one-sided repelling force, modeled as a power law nonlinearity of exponent α—see Section 4.2.2. Here, $[y]^- = (y - |y|)/2$. The reed displacement y is thus permitted to be negative. This term, inspired by collision models used in hammer–string dynamics [75], is the sole distinguishing feature of the model, which is otherwise identical to that which appears in the literature [159, 202, 350, 100, 389, 16].

The more important nonlinearity involves the pressure difference p_Δ across the mouthpiece, defined as

$$p_\Delta = p_m - p_{in}$$

where $p_m(t)$ is the mouth pressure (which is external control data, supplied by the player) and $p_{in}(t)$ the pressure at the entrance to the acoustic tube. This pressure difference is related to the flow in the mouthpiece u_m through Bernoulli's law,

$$u_m = w[y]^+\sqrt{\frac{2|p_\Delta|}{\rho}}\text{sign}(p_\Delta) \tag{9.25}$$

where, here, w is the width of the reed channel. The flow is non-zero only when the reed is not in contact with the mouthpiece, or when $y > 0$. As such, the quantity y^+ is given by $[y]^+ = (y + |y|)/2$. Neglected here is an inertia term—see, e.g., [136]. The square root dependence of flow on velocity could be generalized to a power law [17] with few resulting complications in the discretization procedure to be outlined below.

The flow variables themselves are related by a conservation law

$$u_{in} = u_m - u_r$$

where u_{in} is the flow entering the acoustic tube, and where u_r is related to reed displacement y by

$$u_r = S_r\frac{dy}{dt}$$

In order to facilitate the eventual coupling with an acoustic tube, defined by Webster's equation, it is useful to introduce scaled or non-dimensional variables as follows:

$$y' = \frac{y}{H_0} - 1 \qquad p'_\cdot = \frac{p_\cdot}{\rho c^2} \qquad u'_\cdot = \frac{u_\cdot}{cS_0}$$

for any pressure variable $p_\cdot$ or velocity variable $u_\cdot$, where c, ρ, and S_0 are the wave speed, air density, and left-hand cross-sectional area for the tube. When inserted in the above equations (and primes subsequently removed) the following system results:

$$\frac{d^2y}{dt^2} + 2\sigma_0\frac{dy}{dt} + \omega_0^2 y - \omega_1^{\alpha+1}\left(|[y+1]^-|\right)^\alpha = -\mathcal{Q}p_\Delta \tag{9.26a}$$

$$p_\Delta = p_\mathrm{m} - p_\mathrm{in} \tag{9.26b}$$

$$u_\mathrm{m} = \mathcal{R}[y+1]^+\sqrt{|p_\Delta|}\mathrm{sign}(p_\Delta) \tag{9.26c}$$

$$u_\mathrm{in} = u_\mathrm{m} - u_\mathrm{r} \tag{9.26d}$$

$$u_\mathrm{r} = \mathcal{S}\frac{dy}{dt} \tag{9.26e}$$

where

$$\mathcal{Q} = \frac{\rho c^2 S_\mathrm{r}}{M_\mathrm{r} H_0} \qquad \mathcal{R} = \sqrt{2}\frac{wH_0}{S_0} \qquad \mathcal{S} = \frac{S_\mathrm{r}H_0}{cS_0} \tag{9.27}$$

9.3.2 Energy balance

An energy balance follows immediately from system (9.26). Multiplying (9.26a) by dy/dt leads to

$$\frac{d\mathfrak{H}_\mathrm{r}}{dt} = -\mathfrak{Q}_\mathrm{r} - \mathcal{Q}\frac{dy}{dt}p_\Delta$$

where $\mathfrak{H}_\mathrm{r}$, the stored energy of the reed (including collision effects), and the reed dissipation $\mathfrak{Q}_\mathrm{r}$ are

$$\mathfrak{H}_\mathrm{r} = \frac{1}{2}\left(\frac{dy}{dt}\right)^2 + \frac{\omega_0^2}{2}y^2 + \frac{(\omega_1|[y+1]^-|)^{\alpha+1}}{\alpha+1} \geq 0 \qquad \mathfrak{Q}_\mathrm{r} = 2\sigma_0\left(\frac{dy}{dt}\right)^2 \geq 0$$

But, using the other members of system (9.26), one may continue as

$$\begin{aligned}\frac{d\mathfrak{H}_\mathrm{r}}{dt} &\overset{(9.26e)}{=} -\mathfrak{Q}_\mathrm{r} - \frac{\mathcal{Q}}{\mathcal{S}}u_\mathrm{r}p_\Delta \overset{(9.26d)}{=} -\mathfrak{Q}_\mathrm{r} - \frac{\mathcal{Q}}{\mathcal{S}}u_\mathrm{m}p_\Delta + \frac{\mathcal{Q}}{\mathcal{S}}u_\mathrm{in}p_\Delta \\ &\overset{(9.26c)}{=} -\mathfrak{Q}_\mathrm{r} - \frac{\mathcal{Q}\mathcal{R}}{\mathcal{S}}[y+1]^+|p_\Delta|^{3/2} + \frac{\mathcal{Q}}{\mathcal{S}}u_\mathrm{in}p_\Delta \\ &\overset{(9.26b)}{=} -\mathfrak{Q}_\mathrm{r} - \mathfrak{Q}_\mathrm{m} + \frac{\mathcal{Q}}{\mathcal{S}}u_\mathrm{in}p_\mathrm{m} - \frac{\mathcal{Q}}{\mathcal{S}}u_\mathrm{in}p_\mathrm{in}\end{aligned}$$

where

$$\mathfrak{Q}_\mathrm{m} = \frac{\mathcal{Q}\mathcal{R}}{\mathcal{S}}[y+1]^+|p_\Delta|^{3/2} \geq 0$$

is the loss corresponding to the pressure drop in the mouthpiece.

At this point, one may note that the term $u_\mathrm{in}p_\mathrm{in}$ corresponds, in the coupled tube system with a radiating boundary condition, from (9.12) to

$$u_\mathrm{in}p_\mathrm{in} = \frac{1}{\gamma^3}\frac{d(\mathfrak{H}+\mathfrak{H}_\mathrm{b})}{dt} + \frac{1}{\gamma^3}\mathfrak{Q}_\mathrm{b}$$

Thus, one arrives at

$$\frac{d}{dt}\left(\mathfrak{H}_{\mathrm{r}} + \frac{\mathcal{Q}}{\mathcal{S}\gamma^3}\left(\mathfrak{H} + \mathfrak{H}_{\mathrm{b}}\right)\right) = -\mathfrak{Q}_{\mathrm{r}} - \mathfrak{Q}_{\mathrm{m}} - \frac{\mathcal{Q}}{\mathcal{S}\gamma^3}\mathfrak{Q}_{\mathrm{b}} + \frac{\mathcal{Q}}{\mathcal{S}} u_{\mathrm{in}} p_{\mathrm{m}} \tag{9.28}$$

which relates the rate of change of the total energy stored in the reed, the acoustic tube, and at the radiating end of the tube to the various loss mechanisms and the supplied power. Here, it is clear that under zero-input (or transient) conditions, or when $p_{\mathrm{m}} = 0$, the system is strictly dissipative—one may bound the various state variables corresponding to the tube and reed accordingly. Interestingly, under driven conditions, it is not clear that one may find a corresponding bound on the state in terms of p_{m}; this is in direct contrast to the situation for linear systems—see, e.g., the case of the driven SHO in Section 3.6. See Problem 9.11 for some exploration of this.

9.3.3 Finite difference schemes

A difference scheme for Webster's equation, including the radiating boundary condition, has been presented in Section 9.1.5. What remains is to devise a scheme for the reed system (9.26), and to couple it to the scheme for Webster's equation (9.13) at its left boundary. A simple scheme for the harmonic oscillator-like system (9.26a) is rather similar to those which are employed in Chapters 3 and 4:

$$\delta_{tt} y + 2\sigma_0 \delta_{t\cdot} y + \omega_0^2 \mu_{t\cdot} y - \omega_1^{\alpha+1} \left(\mu_{t\cdot}(y+1)\right) \left(\left|[y+1]^-\right|\right)^{\alpha-1} = -\mathcal{Q} p_\Delta \tag{9.29}$$

Notice the use of semi-implicit approximations for both the linear reed stiffness term and the collision term—such an approach has been described at length in Chapter 4, and leads to very robust numerical behavior. The other members of system (9.26) involve no time differentiation, and thus may be left as they are, with the exception of (9.26e), which may be discretized as

$$u_{\mathrm{r}} = \mathcal{S}\delta_{t\cdot} y \tag{9.30}$$

In order to connect the difference scheme above to scheme (9.13) for Webster's equation, at $x = 0$, one may use

$$p_{\mathrm{in}} = \frac{1}{\gamma}\delta_{t\cdot}\Psi_0 \qquad\qquad u_{\mathrm{in}} = -\delta_{x\cdot}\Psi_0 \tag{9.31}$$

Implementation

The finite difference update (9.29), along with the other time-independent members of system (9.26), as well as (9.30), coupled with scheme (9.13) for Webster's equation, is at first sight rather complex. Notice that updating of y according to (9.29) depends on p_Δ, and thus eventually on y itself; it also depends on (as yet unknown) values of Ψ, which themselves depend on y! Furthermore, the coupling is nonlinear. Interestingly, in this case, a fully explicit solution is available, as will be detailed below.

Beginning from (9.29) at time step n, one may employ identities (2.7g) and (2.7h) in order to isolate $\delta_{t\cdot} y$ and p_Δ as

$$a_1^n + a_2^n \delta_{t\cdot} y^n = -\mathcal{Q} p_\Delta^n$$

where the time series a_1^n and a_2^n depend only on previously computed values of y. In addition, by virtue of the use of semi-implicit discretizations of the stiffness and collision terms, one also has $a_2^n \geq 0$. Then, using (9.30), (9.26d), and (9.26c), one ends up with a relationship between p_Δ and u_{in}:

$$p_\Delta^n + b_1^n \sqrt{|p_\Delta^n|}\,\mathrm{sign}(p_\Delta^n) + b_2^n = b_3^n u_{\mathrm{in}}^n \tag{9.32}$$

where $b_1^n \geq 0$ and $b_3^n \geq 0$ again depend only on known values of the grid function y.

Turning now to scheme (9.13) at the connection point with the reed model, using identity (2.7g), the scheme may be rewritten as

$$\frac{2(\mu_{xx}S_0)}{k}\left(\delta_{t\cdot}\Psi_0^n - \delta_{t-}\Psi_0^n\right) = \frac{2\gamma^2}{h}\left(\mu_{xx}S_0\delta_{x+}\Psi_0^n - \mu_{x-}S_0\delta_{x\cdot}\Psi_0^n\right)$$

Employing (9.31), as well as (9.26b), this may be written as

$$u_{\text{in}}^n = c_1^n p_\Delta^n + c_2^n \tag{9.33}$$

where c_1^n and $c_2^n \geq 0$ again depend on previously computed values of the grid function Ψ, as well as the external control signal p_{m}^n. Finally, (9.32) and (9.33) may be combined to given a single equation in p_Δ:

$$|p_\Delta| + d_1^n\sqrt{|p_\Delta|} + \frac{d_2^n}{\text{sign}(p_\Delta)} = 0 \tag{9.34}$$

where $d_1^n \geq 0$ and d_2^n are known. In order to have a real solution, one must have $\text{sign}(p_\Delta^n) = -\text{sign}(d_2^n)$—the above equation may then be solved uniquely in $\sqrt{|p_\Delta^n|}$.

Given that values of the grid function Ψ and the time series y are known up through time step n, the following order of operations is thus implied:

- Calculate p_Δ^n through the solution of (9.34).
- Calculate y^{n+1} through (9.29).
- Calculate p_{in}^n using (9.26b) and the known value of p_{m}^n.
- Calculate $\delta_{t\cdot}\Psi_0^n$, and thus Ψ_0^{n+1}, using the first of the boundary conditions (9.31).
- Calculate the remaining values of Ψ_l^{n+1}, $l > 0$, using scheme (9.13).

At this point, the calculation cycle is repeated. It is important to note that this explicit computability of the solution depends on the use of the semi-implicit approximations in (9.29); such computability (or modularity) is often reported to be a benefit of scattering methods such as wave digital filters [279, 190] and other wave-based methods [389]. Here, however, it has been carried out without any explicit reference to wave variables. See Problem 9.12 and Programming Exercise 9.5.

9.3.4 Self-sustained oscillation

The reed–bore system exhibits self-sustained oscillations, just as in the case of the bow–string interaction, as discussed in Section 7.4. See Figure 9.12 which shows the envelope of the output pressure waveform from the reed–bore model in a clarinet configuration, under various constant input pressures p_{m}, as indicated.

9.3.5 Reed beating effects

A very important effect, perceptually, is that of the collision, or beating, of the reed against the mouthpiece lay. In this lumped model, when the reed is in contact with the lay, it undergoes an elastic collision, and the flow u_{m} is identically zero. It is instructive to examine both the reed displacement and the output pressure spectrum under such conditions—see Figure 9.13. The reed displacement waveform becomes more square, with a Gibbs-like oscillatory effect at the leading edge of each period—note that the non-dimensional reed displacement takes on values < -1; the extent of such "penetration" may be controlled through the choice of ω_1 and α, but the general results are in agreement with other published simulation results (see, e.g., [200]). The output spectrum exhibits a marked relative increase in the strength of high-frequency components,

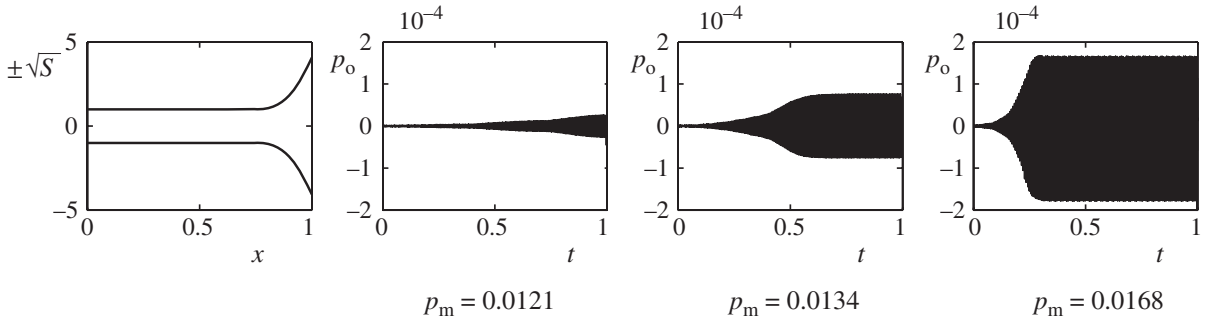


Figure 9.12 Self-sustained oscillation in the reed–bore model. At left, a non-dimensional clarinet-like bore profile, and in the three subsequent plots the output pressure waveform under increasing values of non-dimensional input pressure p_m, as indicated. At $p_m = 0.0121$, the model is at the threshold of oscillation, and as the pressure becomes higher, the note onset time becomes shorter. The reed model parameters, in this case, are, referring to system (9.26), $M_r = 3.37 \times 10^{-6}$ kg, $\omega_0 = 23\,250$ rad/s, $\sigma_0 = 1500$ s^{-1}, $H_0 = 4 \times 10^{-4}$ m, $\omega_1 = 316$, $\alpha = 3$, $S_r = 1.46 \times 10^{-4}$ m^2, and, for the bore, $L = 0.664$ m and $S_0 = 1.72 \times 10^{-4}$ m^2.

leading to a brighter sound [389]. Indeed, the realism of the resulting sound is greatly enhanced by such a beating model.

On the other hand, the lumped reed model is not wholly sufficient to capture the full behavior of such a system—in reality, the reed may not collide with the lay per se, but can curl up progressively against it. Clearly a one-mass model cannot represent such behavior, and a more refined model of the reed might allow for more degrees of freedom, or even distributed (i.e., bar-like) behavior. See, e.g., [16], where finite difference methods are used to simulate a distributed model of a reed.

9.3.6 Time-varying reed equilibrium displacement

In the model given in (9.26), there is a single external control signal, the mouth pressure p_m. In more elaborate models, there are often others—one perceptually salient one is the time variation

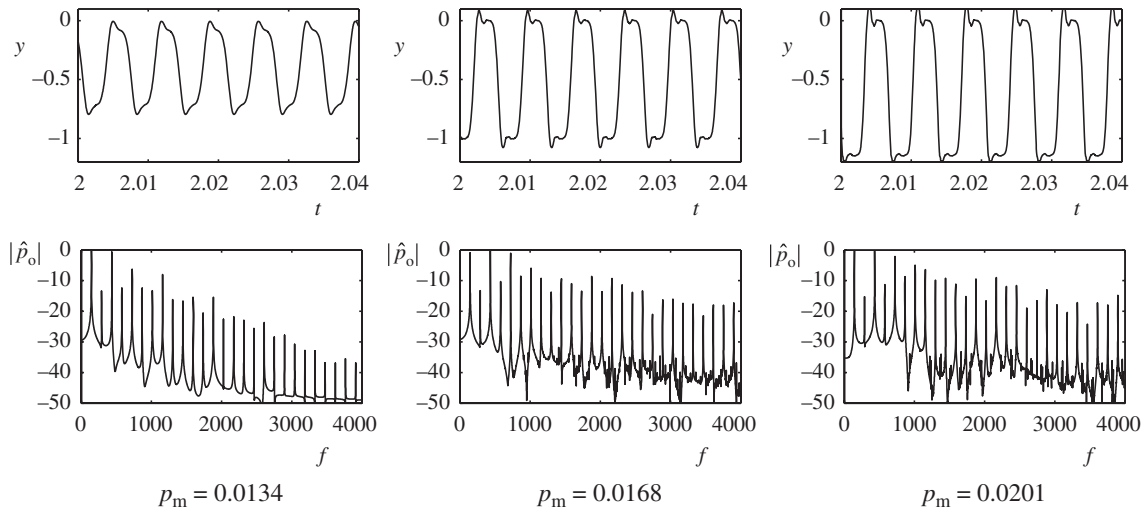
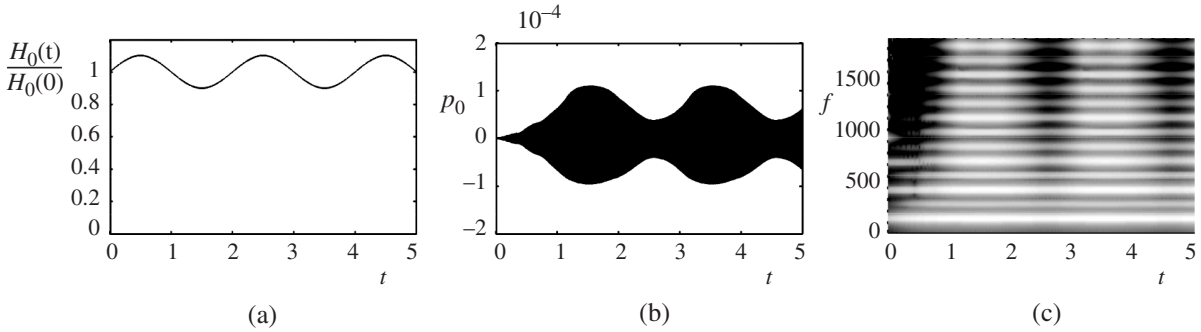


Figure 9.13 Beating effects in the reed model. Top, non-dimensional reed displacement y; bottom, non-dimensional output pressure spectrum $|\hat{p}_o|$, in decibels. Here, the bore and reed parameters are for a clarinet-like model, as in the caption to Figure 9.12, and scheme (9.29) is used, at a sample rate of 44.1 kHz. The input pressure waveform p_m is a constant, of a value indicated below the plots.



(a) (b) (c)

Figure 9.14 Time variation of the reed equilibrium displacement, for a reed–bore system of parameters as given in the caption to Figure 9.12: (a) the time variation of the displacement itself, plotted as $H_0(t)/H_0(0)$; (b) the resulting envelope of the non-dimensional pressure output; and (c) its spectrogram.

of the equilibrium displacement H_0 itself, due to increased embouchure force. The use of such a time-varying control $H_0(t)$ has been employed in the synthesis model of Guillemain et al. [159], and leads to important variations in timbre and loudness in the resulting sound output—see Figure 9.14.

From a systems point of view, variation of $H_0(t)$ is of a very different nature from that of the mouth pressure p_m, and is similar in many respects to time variation in the vocal tract cross-sectional area (see Section 9.2.5). As might be expected, a full analysis is very complex, in that though H_0 is treated as a parameter, variations are capable of injecting energy into the system; furthermore, it is used as a scaling parameter in arriving at the non-dimensional system (9.26). Given that the rate of change of H_0 will normally be very slow compared to audio rates, it seems reasonable to neglect such a complete analysis, and allow the parameters $\mathcal{Q}$, $\mathcal{R}$, and $\mathcal{S}$ to be slowly time varying as well. A choice such as this allows the difference scheme developed in Section 9.3.3 to continue to be used, though one must be alert to the possibility of instability if one begins varying H_0 at faster rates. See Programming Exercise 9.6.

9.3.7 Squeaks and multiphonics

Even with this very simple model of a single-reed wind instrument, it is possible to create a very wide variety of sounds, using merely a constant mouth pressure p_m and reed equilibrium displacement H_0, and crude bore profiles such as cylinders and cones. Indeed, the tones can evolve over long durations, warble, squeak, and exhibit a "multiphonic" character [18] (though this word is often used in conjunction with specialized fingerings—toneholes will be introduced in the next section). Indeed, the sounds generated, when one is first experimenting with such an algorithm, often resemble those made by beginners. This is exactly what one should expect from a true physical model—one must learn to play it.

For some typical (though somewhat artificial) examples of such complex timbres, see the spectrograms given in Figure 9.15. In the three cases, all that is altered is the bore profile (from cylindrical to conical) and the constant mouth pressure. At top is a typical clarinet-like note produced using a cylindrical bore, consisting of odd-numbered partials—note the rise time to self-sustained oscillation, as described earlier in Section 9.3.4. In the middle panel, the bore has been chosen conical, corresponding, very roughly, to that of a saxophone. At the indicated blowing pressure, rather complex behavior is observed over perhaps 3 s. Oscillatory behavior is set up initially, now with evenly spaced harmonics, followed by a stuttering tone occurring between 1.5 and 2 s, followed by a final settling down to a tone of a pitch slightly lower than that at the beginning of the sound. The warbling portion of the sound, in particular, is precisely a multiphonic.

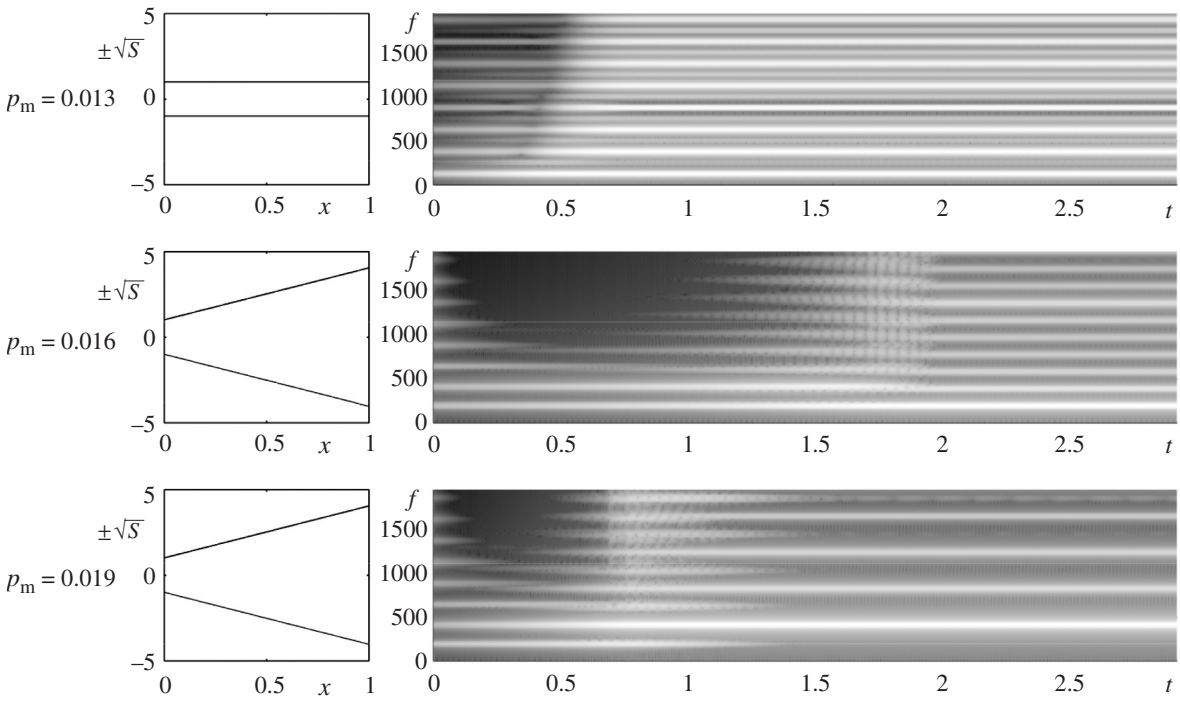


Figure 9.15 Bore profiles (left) and spectrograms of output pressure, for constant mouth pressures, as indicated. In all cases, the other system parameters are as in the caption to Figure 9.12. Results generated using scheme (9.29), at 44.1 kHz.

For the same configuration, raising the blowing pressure yields a very different behavior, as shown at bottom—the pitch of the tone makes a jump to the octave.

As one can imagine, the space of possible timbres becomes far larger when toneholes (to be discussed shortly) are introduced.

9.3.8 Toneholes

The tonehole is the primary means by which pitch changes are effected in a reed wind instrument. In musical acoustics, it is often referred to as a branched side tube, of a given radius b and height[3] τ, which may be open, closed, or partly open—see Figure 9.16(a). The obvious effect of an open hole is to change the effective tube length, and thus the sounding pitch of the wind instrument, to a degree which depends on the hole dimensions and placement, but more subtle effects result from the behavior of the mass of air enclosed in the hole volume, whether it is open or closed. The most important early technical work on the tonehole was due to Benade [31].

In the literature, the tonehole is almost invariably represented in terms of an equivalent electrical N-port, using the analogy between the acoustic bore and an electrical transmission line, and characterized in terms of an impedance or scattering matrix, linking pressures and volume velocities at the bore on either side of the branch. See, e.g., [253], and especially the pair of articles by Keefe [198, 197] (as well as subsequent corrections and refinements [199, 111, 255]), which set the stage for many later developments in experimental musical acoustics as well as synthesis algorithms for wind instruments. In particular, the circuit representation is a natural fit for scattering methods such as digital waveguides, and especially wave digital filters, which were originally devised as digital counterpart to analog circuit elements—see, e.g., [314, 336, 374, 391].

[3] The height τ, as shown in Figure 9.16(a), does not take into account the effect of curvature of the bore at the junction with the tonehole. A better definition of the tonehole height is the volume enclosed by the hole divided by the surface area πb^2—see [199] for an explicit calculation of this effective tonehole height.

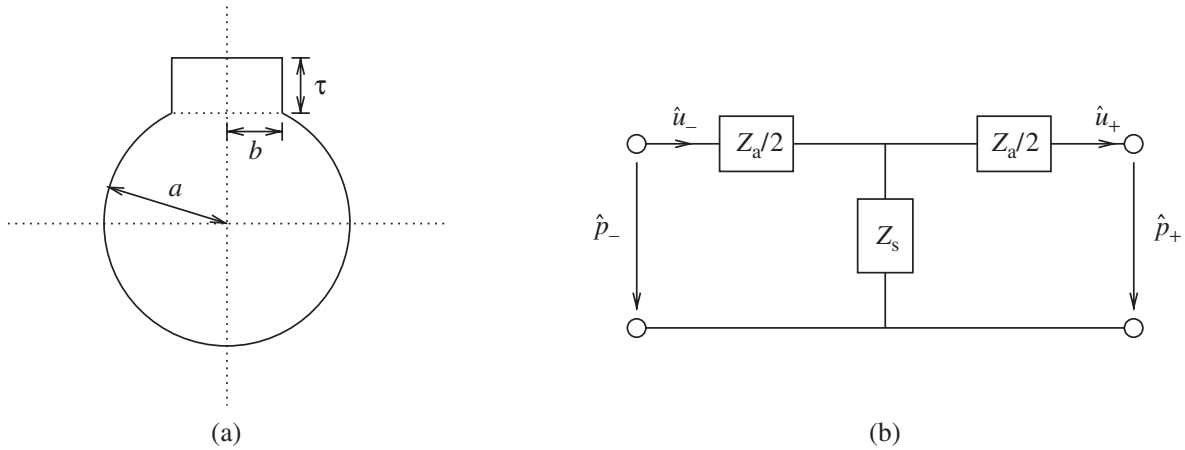


Figure 9.16 A single woodwind tonehole. (a) Bore cross-section at tonehole location. τ is the physical height of the tonehole, and b is the radius of the tonehole opening. The dimensionless parameters ξ and S_T are defined as $\xi = \tau/L$, where L is the length of the bore, and $S_T = \pi b^2/S(0)$, where $S(0)$ is the surface area of the bore at its left end. (b) Symmetric two-port circuit representation, with a series imedance Z_a and a shunt impedance Z_s.

As might be expected, however, there is no reason why the tonehole cannot be treated directly in a time–space finite difference scheme; in fact, the tonehole behaves exactly as a lumped element with mass and stiffness, in a way which is not extremely different from that of the preparation of strings using masses and springs, as discussed in Section 7.7. In fact, the case of the tonehole is easier to deal with in some respects, as in most models the lumped element is linear (though this not the case when a more refined model, taking into account the effects of nonlinear loss, is employed [112]). On the other hand, new complications arise, due to the particulars of the circuit representation and the essentially time-varying character of the tonehole in a playing situation. Again, the time domain approach to toneholes discussed here is quite distinct from, for example, those involving multiple convolutions with reflection functions from toneholes [232] with feedback to the reed mechanism [26].

Circuit representation

It is useful to take, as a starting point, the two-port representation shown in Figure 9.16(b), which includes impedances $Z_a(s)/2$ on the series branches and $Z_s(s)$ on the shunt branch. This symmetric "T" two-port is the usual form encountered in the literature, where there are slight variations in the exact forms of the shunt and series impedances. Even more general forms are possible—see, e.g., [111]. For the moment, the exact form of the impedances is left unspecified. $\hat{p}_-$ and $\hat{p}_+$ are the Laplace-transformed pressures to the left and right of the tonehole, and $\hat{u}_-$ and $\hat{u}_+$ the Laplace-transformed volume velocities—again, all dependent variables are assumed non-dimensionalized here.

In order to relate such a structure to a form of Webster's equation, it is useful to extract the pressure and volume velocity drops $\hat{p}_{\text{diff}} = \hat{p}_+ - \hat{p}_-$ and $\hat{u}_{\text{diff}} = \hat{u}_+ - \hat{u}_-$. Applying Kirchhoff's connection rules to the diagram shown in Figure 9.16(b), one arrives immediately at

$$\hat{p}_{\text{diff}} = -\epsilon(x)\frac{Z_a}{2}(\hat{u}_+ + \hat{u}_-) \qquad \hat{u}_{\text{diff}} = -\epsilon(x)\frac{1}{2Z_s + Z_a/2}(\hat{p}_+ + \hat{p}_-)$$

Here, $\epsilon(x)$ is a distribution peaked at the location of the tonehole $x = x_T$ along the bore, and is assumed to integrate to unity (it will shortly be represented by a Dirac delta function). Assuming

that the pairs $\hat{p}_+$, $\hat{p}_-$ and $\hat{u}_+$, $\hat{u}_-$ approach common values of $\hat{p}$ and $\hat{u}$ (i.e., they are spatially continuous), one arrives at

$$\hat{p}_{\text{diff}} = -\epsilon(x) Z_{\text{a}} \hat{u} \qquad \hat{u}_{\text{diff}} = -\epsilon(x) \frac{1}{Z_{\text{s}} + Z_{\text{a}}/4} \hat{p}$$

Now, consider the first-order system (9.4) in the presence of such a lumped element and after Laplace transformation; the quantities $\hat{p}_{\text{diff}}$ and $\hat{u}_{\text{diff}}$ contribute directly to the expressions for $\hat{p}_x$ and $\hat{u}_x$, respectively, to yield

$$Ss\hat{p} = -\gamma \hat{u}_x - \frac{\gamma \epsilon(x)}{Z_{\text{s}} + Z_{\text{a}}/4} \hat{p} \qquad \frac{s}{S} \hat{u} = -\gamma \hat{p}_x - \gamma \epsilon(x) Z_{\text{a}} \hat{u} \tag{9.35}$$

At this point, it is useful to introduce specific forms for the impedances Z_{a} and Z_{s}.

Keefe's tonehole model

Keefe [198] gives the following forms for the impedances Z_{a} and Z_{s}:

$$Z_{\text{a}}(s) = -\frac{s\xi_{\text{a}}^{o,c}}{\gamma S_{\text{T}}} \quad \text{(open, closed)} \qquad Z_{\text{s}} = \begin{cases} \dfrac{1}{S_{\text{T}}} \coth\left(\dfrac{s\xi}{\gamma}\right), & \text{(closed)} \\ \dfrac{s\xi_{\text{e}}}{\gamma S_{\text{T}}}, & \text{(open)} \end{cases} \tag{9.36}$$

Here, the dimensionless parameters S_{T} and various forms of ξ have appeared; S_{T} is the ratio of the tonehole cross-sectional area to the area of the bore at its left end (not at the hole!), and ξ is the ratio of the tonehole height to the tube physical tube length L. The various forms for ξ, which depend on whether the hole is opened or closed, are effective lengths, and exact expressions appear in various publications [198, 199, 391]. In fact, the effective lengths themselves are frequency dependent to a slight degree—such frequency dependence is ignored here, for the sake of simplicity.

Missing in the above expression for Z_{s}, when the hole is open, is a loss term which, for the sake of simplifying analysis, will also be ignored for the moment; this is a rather important effect, in that it allows for sound radiation from the open hole. Unfortunately, its full characterization is rather complex—frequency dependence [199] and perhaps even nonlinear effects at the edges of the hole may need to be modeled [112]. A very crude model is, however, explored in Problem 9.13. Consider first the effect of Z_{a} on the second of (9.35)—when the resulting expression is inverse transformed back to the time domain, the following PDE results:

$$\frac{1}{S^*} u_t = -\gamma p_x \qquad \text{where} \qquad S^*(x) = \frac{S(x)}{1 - \dfrac{S(x)\xi_{\text{a}}\epsilon(x)}{S_{\text{T}}}} \tag{9.37}$$

The area function $S^*(x)$ is equal to $S(x)$ except in the neighborhood surrounding the tonehole, where $\epsilon(x)$ is peaked. Notice, however, that this effective area can itself become negative—this is another interpretation of the so-called "negative length correction" which the impedance Z_{a} implies.[4] In fact, Z_{a} itself is not a positive real function, and thus does not correspond to a strictly passive circuit element! Such an feature leads, as might be expected, to some numerical difficulties when it comes to synthesis, as will be discussed shortly.

[4] One might argue that the function S^* is not well defined in the limit as $\epsilon(x)$ approaches a Dirac delta function distribution—this is the reason why a function ϵ taking on finite values has been employed here, though in the discrete setting such analytical questions can be sidestepped.

Given the above PDE relating pressure to volume velocity, it is useful to introduce a velocity potential, again called Ψ, which takes into account the effective surface area S^*, i.e., if

$$u = -S^*\Psi_x \qquad p = \frac{1}{\gamma}\Psi_t \tag{9.38}$$

then (9.37) is solved exactly. For the series impedance, it is useful, as a prelude to developing finite difference schemes, to approximate it as

$$Z_\mathrm{s} = \begin{cases} \dfrac{\gamma}{s\xi S_\mathrm{T}}, & \text{(closed)} \\ \dfrac{s\xi_\mathrm{e}}{\gamma S_\mathrm{T}}, & \text{(open)} \end{cases}$$

(Interestingly, the form involving the coth function implies a discrete-time relationship directly—see Problem 9.14.) Thus, when the hole is closed, it behaves as a capacitance, or stiffness, and when open, as an inductance, or mass. Now, the first of (9.35) may be written as

$$Ss\hat{p} = -\gamma\hat{u}_x - \frac{\gamma\epsilon(x)}{a_1 s + a_2/s}\hat{p} \quad \text{where} \quad a_1 = \begin{cases} -\dfrac{\xi_\mathrm{a}^c}{4\gamma S_\mathrm{T}}, & \text{(closed)} \\ \dfrac{\xi_\mathrm{e} - \xi_\mathrm{a}^c/4}{\gamma S_\mathrm{T}}, & \text{(open)} \end{cases} \qquad a_2 = \begin{cases} \dfrac{\gamma}{\xi S_\mathrm{T}}, & \text{(closed)} \\ 0, & \text{(open)} \end{cases}$$

or, using (9.38), and taking an inverse Laplace transformation, as the system

$$S\Psi_{tt} = \gamma^2\left(S^*\Psi_x\right)_x - \epsilon(x)\frac{dm}{dt} \qquad a_1\frac{d^2m}{dt^2} + a_2 m = \gamma\langle\Psi_t, \epsilon\rangle_{\mathbb{U}} \tag{9.39}$$

The new function $m(t)$ allows for the storage of energy in the lumped element. One can indeed work from this coupled form to arrive at a simulation routine. The difficulty, however, again relates back to the negative series impedance Z_a, which leads to a negative stored energy term in the above system, which leads, in turn, to somewhat tricky analysis of the resulting finite difference scheme. See Problem 9.15 for more on this topic. In the digital waveguide framework, it is possible [336] to "extract" the negative impedance from the lumped circuit representation, and append it as a "negative delay" to neighboring waveguides; in essence, the non-passive behavior of the element is subsumed by neighboring passive elements. This option, however, is not available in the difference setting, and is perhaps a disadvantage of a fully time–space discrete method, where it is not possible to commute the effects of a lumped element to other spatial locations.

The important point above, however, is that there is a simple coupled system (9.39) which describes the tonehole connection to the bore, which is a good starting point for the development of synthesis methods based on, perhaps, simplified models of the tonehole, such as that given below.

A simplified form and finite difference scheme

A great simplification (and one which may not be entirely justifiable) is to set the series impedance $Z_\mathrm{a} = 0$. Van Walstijn and Scavone [391], in their treatment of the wave digital tonehole, and Scavone, using a pure waveguide construction [314], do essentially this while leaving the negative inductance to be subsumed as a "length correction" by neighboring delay lines. These authors go further, and propose a model which allows more generally for the effects of open and closed holes at once, through a parallel combination of an inductance (or mass) and capacitance (or stiffness):

$$Z_\mathrm{s} = \frac{s\xi_\mathrm{e}}{\phi\gamma S_\mathrm{T}} \| \frac{\gamma}{(1-\phi)s\xi S_\mathrm{T}} = \frac{1}{\dfrac{\phi\gamma S_\mathrm{T}}{s\xi_\mathrm{e}} + \dfrac{(1-\phi)s\xi S_\mathrm{T}}{\gamma}} \tag{9.40}$$

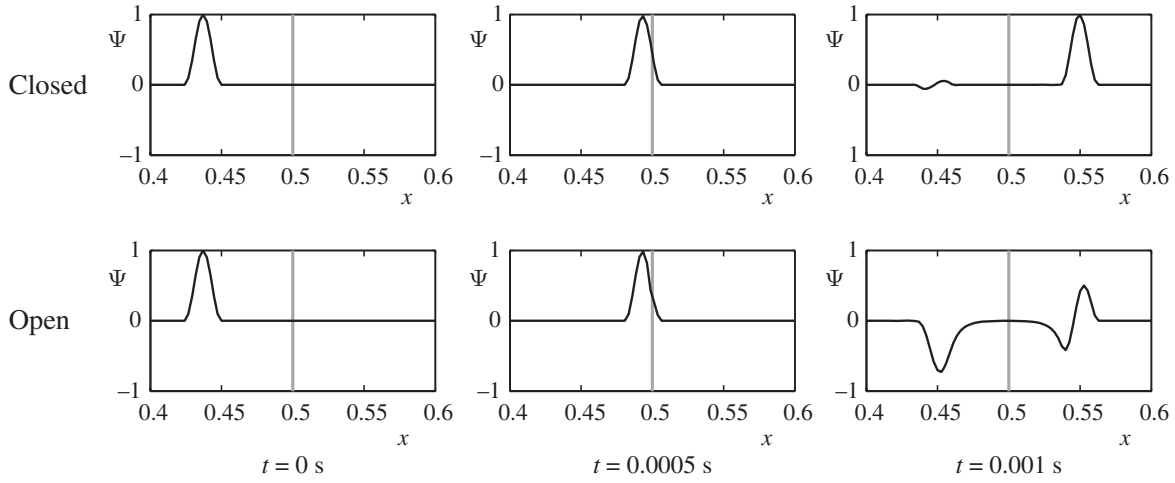


Figure 9.17 Scattering of a narrow pulse by a tonehole, when closed (top) and open (bottom), at the times indicated. The tonehole is located at $x_\mathrm{T} = 0.5$, as indicated by the grey line. The tube is cylindrical, and the tonehole parameters are $S_\mathrm{T} = 0.799$, $\xi = 0.0011$, and $\xi_\mathrm{e} = 0.0026$.

When the parameter $0 \leq \phi \leq 1$ is equal to one, the tonehole is open, and when equal to zero, it is closed. In this case, it is not difficult to show that the following generalization of Webster's equation results:

$$S\Psi_{tt} = \gamma^2 (S\Psi_x)_x - \delta(x - x_\mathrm{T})m \qquad m = \frac{\phi\gamma^2 S_\mathrm{T}}{\xi_\mathrm{e}}\Psi + (1-\phi)\xi S_\mathrm{T}\Psi_{tt} \tag{9.41}$$

Notice now that, because $Z_\mathrm{a}(s)$ has been neglected, the area function $S^* = S$, and, furthermore, the distribution $\epsilon(x)$ has been approximated as a Dirac delta function. Again, an extra variable m has been introduced, though note that the system could be combined to a single equation in the velocity potential Ψ alone—this is only possible when Z_s results from a parallel combination of elements, which is not the case, for example, when losses intervene—see Problem 9.13.

The total energy for the system will thus be $\mathfrak{H} = \mathfrak{H}_\mathrm{tube} + \mathfrak{H}_\mathrm{tonehole}$, where $\mathfrak{H}_\mathrm{tube}$ is the energy corresponding to the tube, as per (9.7), and the extra stored energy at the tonehole is

$$\mathfrak{H}_\mathrm{tonehole} = \frac{\phi\gamma^2 S_\mathrm{T}}{2\xi_\mathrm{e}}(\Psi(x_\mathrm{T}, t))^2 + \frac{(1-\phi)\xi S_\mathrm{T}}{2}(\Psi_t(x_\mathrm{T}, t))^2 \geq 0 \tag{9.42}$$

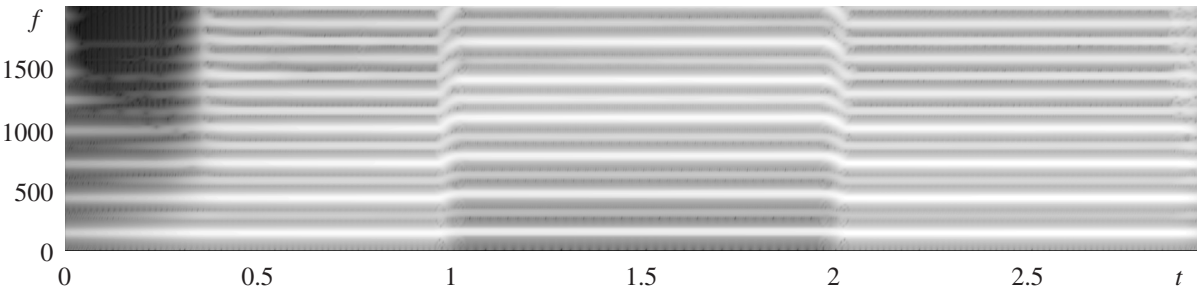


Figure 9.18 Spectrogram of sound output from a reed instrument model under a time-varying tonehole gesture. The reed model corresponds approximately to that of a clarinet, with bore profile and reed parameters as given in Figure 9.12, and with a blowing pressure which ramps linearly from $p_\mathrm{m} = 0.013$ to $p_\mathrm{m} = 0.027$ over a duration of 3 s. The tonehole is placed at $x_\mathrm{T} = 0.8$, with $S_\mathrm{T} = 0.1644$, $\xi = 0.0075$, and $\xi_\mathrm{e} = 0.0134$, and is opened over a duration of 50 ms at $t = 1$ s, and closed over the same duration at $t = 2$ s. Scheme (9.43) is used, at a sample rate of 44 100 Hz. Output pressure is read from the end of the tube, at $x = 1$.

Thus, as before in the case of string preparation, the energetic effect of the tonehole, at least under the simplified model here, is to introduce an additional non-negative component to the total energy.

A useful finite difference approximation to (9.41) is as follows:

$$\mu_{xx} S \delta_{tt} \Psi = \gamma^2 \delta_{x+} ((\mu_{x-} S) \delta_{x-} \Psi) - J_p(x_\mathrm{T}) m \tag{9.43a}$$

$$m = \frac{\phi \gamma^2 S_\mathrm{T}}{\xi_\mathrm{e}} \left(\alpha I_p(x_\mathrm{T}) \Psi + (1 - \alpha) \mu_{t\cdot} I_p(x_\mathrm{T}) \Psi \right) + (1 - \phi) \xi S_\mathrm{T} \delta_{tt} I_p(x_\mathrm{T}) \Psi \tag{9.43b}$$

Again, as in the case of the prepared string (see Section 7.7), the pth-order interpolation and spreading functions $I_p(x_\mathrm{T})$ and $J_p(x_\mathrm{T})$ have been employed—as such, the tonehole may be placed at any location along the bore. See Programming Exercises 9.7 and 9.8. Notice also the use of the free parameter α, which allows for a certain degree of tuning of the algorithm; this is an extra degree of freedom not available when, for example, methods such as wave digital filters are employed. Simple sufficient stability conditions for the above scheme follow, from energy analysis, as

$$\lambda \leq 1 \qquad\qquad \alpha \leq \frac{1}{2} \tag{9.44}$$

though it is possible to improve on this bound—see Problem 9.16.

Reflections

The gross effect of the tonehole is to induce frequency-dependent scattering; when open, the tonehole reflects low frequencies, and generally passes high-frequencies—see the example shown at bottom in Figure 9.17. In this case, the pulse is a traveling wave, and is unnaturally narrow, by musical standards, in order to exhibit the reflection and transmission characteristics. If the hole is closed, low frequencies are passed, with a slight high-frequency backscatter.

Partly open holes and time variation

The feature of interest in the model given in (9.41) is that it allows the tonehole to be partly open; furthermore, if the parameter ϕ is allowed to be time varying, it is possible to effect a change of state between open and closed. The obvious effect is, of course, to change the sounding pitch of the instrument—see Figure 9.18, which is generated using a clarinet-like model, using scheme (9.43) for the bore/tonehole connection.

More interesting, sonically, than the simple change of pitch, are the wide variety of multiphonics which may be generated though judicious placement of toneholes. If, for instance, a single tonehole is opened a good distance away from the bell, one is likely to generate a sound with a multiphonic character, exhibiting, perhaps, distinct sounding pitches and a slowly time-varying amplitude envelope—see Figure 9.19.

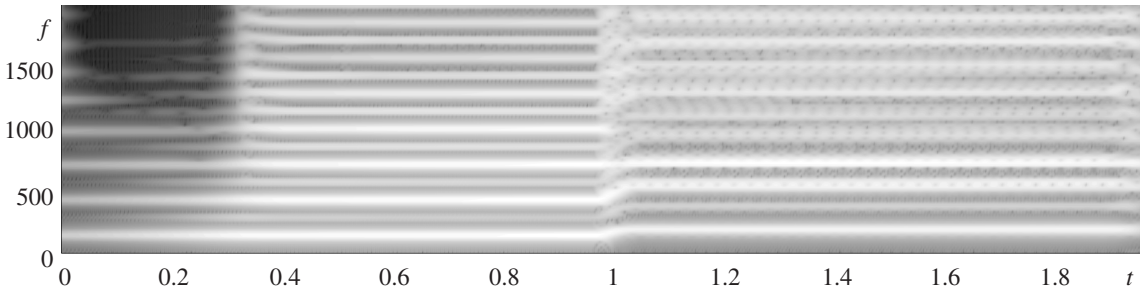


Figure 9.19 Spectrogram of sound output from a reed instrument model under a time-varying tonehole gesture. The reed, bore, and tonehole parameters are as in Figure 9.18, but the tonehole is located at $x_\mathrm{T} = 0.6$, and opened at $t = 1$ s. The resulting tone sounds at two distinct pitches, and "warbles."

9.4 Other wind instruments

Due to space considerations, only single-reed wind instruments have been described in this chapter. There are, of course, other families of instruments based on acoustic tubes, many of which have been simulated for synthesis, including double-reed instruments, the flute family, and brass instruments. Many features are shared with the reed model, as detailed below, and indeed synthesis methods based on difference schemes, or digital waveguides may be readily applied to such instruments.

Double-reed instruments, such as the oboe and bassoon, though they have seen less research at the level of musical acoustics (see, e.g., [397, 6]), differ mainly in the treatment of the reed mechanism—due to the length and geometry of the reed, finer attention to the details of the flow nonlinearity is necessary. Still, the resulting lumped representation of the reed is not extremely different from that presented in Section 9.3, and sound synthesis algorithms have resulted [7, 158].

Brass instrument synthesis has also seen a lot of activity, and again, just as in the case of the reed, the typical model is composed of a linear bore connected to a lumped valve-like excitation. In this case, however, the "reed" (the player's lips) is in a configuration such that it is blown open at higher pressures, rather than shut—see, e.g., [136] for a standard explanation of the distinction between the two types of reed behavior. Simple models used in the studies in musical acoustics and in synthesis, however, are again nearly identical to the reed model presented in Section 9.3—see, e.g., [3, 297, 398]. Real-time synthesis has been available for some time, using digital waveguides [89], and methods based on nonlinear ODEs incorporating delays [297]; perhaps the most polished result has been the BRASS physical modeling software package recently released by IRCAM [400]. One significant departure in the model of the bore has involved looking at wave propagation in the bore at high blowing pressures, in which case shock waves may be produced, leading to a brighter timbre. In this case, Webster's equation no longer suffices, and a full model of nonlinear wave propagation must be simulated. See, e.g., [172, 245, 399]. Another departure from Webster's equation results from the relatively wide flare of the bell in brass instruments relative to woodwinds, in which case a 1D model may be insufficient to describe effects at high frequencies. See, e.g., the work of Noreland [258] in which finite difference methods in higher dimensions are employed in order to simulate such behavior.

More complex still is the case of flute- and recorder-like instruments; the interaction between the jet of air and the sharp-edged mouthpiece, beyond exciting the instrument body at its resonant frequencies, introduces turbulence which is audible as a noise-like component to the resulting sound. Synthesis applications employ a greatly simplified model of this interaction, along the lines of the model proposed by McIntyre et al. [236]—see, e.g., [85]. This is also true of digital waveguide models—see, e.g., [373, 369], and especially the work of Verge [395, 394].

9.5 Problems

Problem 9.1 *Find conditions on the surface area $S(x)$ such that $a(x)$, in the transformed form of Webster's equation given in* (9.3)*, vanishes. Relate $a(x)$ to local tube curvature.*

Problem 9.2 *Recall the form of the radiating boundary condition for Webster's equation, given in* (9.9)*. Such a condition is usually encountered in an impedance form in the literature, such as*

$$\hat{p}(s) = Z(s)\hat{u}(s) \qquad \text{with} \qquad Z(s) = \frac{s}{S(1)\gamma\,(\alpha_2 + \alpha_1 s)} \tag{9.45}$$

Here, p and u are dimensionless pressure and volume velocity, respectively, and the impedance is expressed in terms of the parameters α_1 and α_2 given in (9.10) *or* (9.11)*. The hat (i.e., circumflex) notation indicates a Laplace transformation in the frequency variable s.*

(a) By replacing factors of s by partial time derivatives, show that the impedance relation above implies the radiation boundary condition (9.9) *in the variable Ψ at $x = 1$. You may use the settings $p(1,t) = \Psi_t(1,t)/\gamma$ and $u(1,t) = -S(1)\Psi_x(1,t)$.*

(b) Using a series expansion, show that the impedance $Z(j\omega)$ given above may be approximated as

$$Z(s) = \frac{s}{S(1)\gamma\alpha_2}\left(1 - \frac{\alpha_1}{\alpha_2}s\right) \tag{9.46}$$

which is the form encountered frequently in the literature. As in part (a) above, convert this impedance into a boundary condition in Ψ at $x = 1$. Can you show, using energy methods, that this condition does not correspond to a strictly dissipative termination?

Problem 9.3 *In order to represent a passive system, or one which dissipates energy, an impedance $Z(s)$ must satisfy the property of positive realness:*

$$\mathrm{Re}\,(Z(s)) \geq 0 \qquad \text{for} \qquad \mathrm{Re}(s) > 0$$

See, e.g., [30, 387].

Show that the transfer function corresponding to the radiation boundary condition given in (9.45) *above is positive real, but that the form in* (9.46) *is not. (Positive realness of an impedance corresponds exactly to strict dissipativity in the energy framework, and is used as a defining property in the design of stable wave digital filter networks.)*

Problem 9.4 *Consider scheme* (9.13) *for Webster's equation, and recall that the approximation $[S]$ is second-order accurate. Show that it may be rewritten as*

$$[S]\delta_{tt}u = \gamma^2\left(\delta_{x\cdot}S\delta_{x\cdot}u + \mu_{x+}\mu_{x-}S\delta_{xx}u\right)$$

Every approximation to either S or u in the above scheme is second-order accurate in space or time, so therefore deduce that the scheme is second-order accurate as a whole.

Problem 9.5 *Showing numerical stability for the more accurate numerical boundary conditions* (9.16) *for Webster's equation is slightly tricky—the key is to define an appropriate inner product. To this end, consider the weighted inner product defined in* (5.38)*, in which the endpoint weightings ϵ_l and ϵ_r are left unspecified for the moment.*

Show that, by applying this inner product in the energy analysis of scheme (9.13)*, the following energy balance results, generalizing* (9.14)*:*

$$\delta_{t+}\mathfrak{h} = \mathfrak{b} \quad \text{with} \quad \mathfrak{h} = \mathfrak{t} + \mathfrak{v} \quad \mathfrak{t} = \frac{1}{2}\left(\|\sqrt{[S]}\delta_{t-}\Psi\|^{\epsilon_l,\epsilon_\mathrm{r}}_{\mathbb{U}_N}\right)^2 \quad \mathfrak{v} = \frac{\gamma^2}{2}\langle(\mu_{x+}S)\delta_{x+}\Psi, e_{t-}\delta_{x+}\Psi\rangle_{\underline{\mathbb{U}}_N}$$

and where the boundary term is now

$$\mathfrak{b} = \gamma^2\delta_{t\cdot}\Psi_N\left(\frac{\epsilon_\mathrm{r}}{2}\mu_{x+}S_N\delta_{x+}\Psi_N + \left(1 - \frac{\epsilon_\mathrm{r}}{2}\right)\mu_{x-}S_N\delta_{x-}\Psi_N\right)$$

$$-\gamma^2\delta_{t\cdot}\Psi_0\left(\frac{\epsilon_l}{2}\mu_{x-}S_0\delta_{x-}\Psi_0 + \left(1 - \frac{\epsilon_l}{2}\right)\mu_{x+}S_0\delta_{x+}\Psi_0\right)$$

Show that under the special choices of $\epsilon_l = \mu_{x+}S_0/\mu_{xx}S_0$ and $\epsilon_\mathrm{r} = \mu_{x-}S_N/\mu_{xx}S_N$, the boundary conditions (9.16) *are strictly dissipative.*

Can you continue and show that under these conditions, the energy itself is non-negative? Your answer to Problem 5.10 should be of use here.

Problem 9.6 *Consider Webster's equation, with a locally reacting wall vibration condition, as described in system* (9.18).

(a) Show that this system is dissipative, i.e., that

$$\frac{d(\mathfrak{H} + \mathfrak{H}_{\mathrm{w}})}{dt} = -\mathfrak{Q}$$

where $\mathfrak{H}$ is the stored energy in the tube, given in (9.7)*, $\mathfrak{H}_{\mathrm{w}} \geq 0$ is the stored energy of the walls, and where $\mathfrak{Q} \geq 0$ is the total wall loss. Assume conservative boundary conditions.*

(b) Suppose, furthermore, that the tube is cylindrical (i.e., $S = 1$) and of infinite extent (i.e. the system is defined over $\mathcal{D} = \mathbb{R}$), and that the wall is lossless (i.e., $\sigma_0 = 0$). Insert test solutions of the form $e^{st+j\beta x}$ for both Ψ and w, and determine a characteristic equation relating frequency s and wavenumber β. Show that there will be two separate dispersion relations, and that in both cases s is purely imaginary.

Problem 9.7 *Show that the coupled difference scheme* (9.19) *may be written in an explicit update form. In order to do this, you can either expand out the operator form as given, or employ identity* (2.7g).

Problem 9.8 *Show that both the simplified difference scheme* (9.17) *for Webster's equation and the scattering structure discussed in Section 9.2.4 may be implemented in a form requiring one multiplication per grid point (or junction), where for the scattering method, all arithmetic is performed according to* (9.22)*. How does the number of additions vary?*

Problem 9.9 *Consider the time-varying generalization of Webster's equation* (9.23)*. Suppose that one defines the "time-varying" stored energy of the system as*

$$\mathfrak{H} = \frac{1}{2}\|\sqrt{S}\Psi_t\|_{\mathbb{U}}^2 + \frac{\gamma^2}{2}\|\sqrt{S}\Psi_x\|_{\mathbb{U}}^2$$

(a) Show that one may derive the following energy balance (assuming conservative boundary conditions):

$$\frac{d\mathfrak{H}}{dt} = -\frac{1}{2}\langle S_t\Psi_t, \Psi_t\rangle_{\mathbb{U}} + \frac{\gamma^2}{2}\langle S_t\Psi_x, \Psi_x\rangle_{\mathbb{U}}$$

(b) Show that

$$\frac{d\mathfrak{H}}{dt} \leq \frac{1}{2}\|\sqrt{|S_t|}\Psi_t\|_{\mathbb{U}}^2 + \frac{\gamma^2}{2}\|\sqrt{|S_t|}\Psi_x\|_{\mathbb{U}}^2$$

and find a function $\alpha(t) \geq 0$, depending on the properties of S alone, such that

$$\frac{d\mathfrak{H}}{dt} \leq \alpha\mathfrak{H}$$

(Your function $\alpha(t)$ should be zero when S is time invariant.)

(c) Finally, show that

$$\mathfrak{H}(t) \leq e^{\int_0^t \alpha(t')dt'}\mathfrak{H}(0)$$

Thus there is an bound on the rate of growth of the solution in terms of the variation in S.

(d) How loose is the bound in (c) above? Estimate the growth in $\mathfrak{H}$ using common-sense assumptions about the rate of variation of S in typical speech.

Problem 9.10 *Consider the following implicit finite difference scheme for Webster's equation under time-varying conditions:*

$$\delta_{t+}\left((\mu_{t-}\mu_{xx}S)\delta_{t-}\Psi\right) = \gamma^2\delta_{x+}\left((\mu_{tt}\mu_{x-}S)\mu_{t\cdot}\delta_{x-}\Psi\right)$$

It is possible to obtain a bound on solution growth in the same way as in the continuous case discussed above.

(a) Under conservative boundary conditions, show that the following energy balance holds:

$$\delta_{t+}\mathfrak{h} = -\frac{1}{2}\langle \delta_{t\cdot}\mu_{xx}S,\, (\delta_{t+}\Psi)(\delta_{t-}\Psi)\rangle_{\mathbb{U}_N} + \frac{\gamma^2}{2}\langle \delta_{t\cdot}\mu_{x+}S,\, \mu_{tt}(\delta_{x+}\Psi)^2\rangle_{\underline{\mathbb{U}}_N}$$

where

$$\mathfrak{h} = \frac{1}{2}\|\sqrt{\mu_{t-}\mu_{xx}S}\,\delta_{t-}\Psi\|_{\mathbb{U}_N} + \frac{\gamma^2}{2}\langle \mu_{t-}\mu_{x+}S,\, \mu_{t-}(\delta_{x+}\Psi)^2\rangle_{\underline{\mathbb{U}}_N}$$

(b) Prove that

$$\delta_{t+}\mathfrak{h} \le \alpha\mu_{t+}\mathfrak{h}$$

for some scalar time series $\alpha = \alpha^n$, which depends only on the grid function S. Show, furthermore, that

$$\delta_{t+}\mathfrak{h} \le \alpha^*\mu_{t+}\mathfrak{h}$$

where $\alpha^* = \max_n \alpha^n$.

(c) Given that $\mathfrak{h} \ge 0$ unconditionally (such is the interest in using an implicit scheme here), show that there is a constant ϵ, such that

$$\mathfrak{h}^n \le \epsilon^n \mathfrak{h}^0$$

(d) Do you think this scheme will perform well, especially in terms of numerical dispersion? Consider its behavior in the simple time- and shift-invariant case when $S = 1$.

Problem 9.11 *Consider the energy balance for the single-reed system, as given in* (9.28)*. It is not obvious that one may use this balance in order to bound the size of the state in terms of the control input p_{m}. In this problem, the idea is to show why this is difficult, and perhaps to inspire you to solve the problem!*

(a) Rewrite the energy inequality, neglecting the radiation loss term $\mathfrak{Q}_{\mathrm{b}}$, as

$$\frac{d\mathfrak{H}_{\mathrm{total}}}{dt} \le -\left(\frac{2\sigma_0 u_{\mathrm{r}}^2}{\mathcal{S}^2} + \frac{\mathcal{Q}}{\mathcal{S}}p_{\mathrm{m}}u_{\mathrm{r}}\right) - \frac{\mathcal{Q}\mathcal{R}}{\mathcal{S}}\left([y+1]^+|p_\Delta|^{3/2} - [y+1]^+|p_\Delta|^{1/2}\mathrm{sign}(p_\Delta)p_{\mathrm{m}}\right)$$

(b) Consider the following inequality [210]: for any two positive numbers A, B, and for p, q such that $1/p + 1/q = 1$, and $1 < q < \infty$,

$$AB \le \frac{A^p}{p} + \frac{B^q}{q} \tag{9.47}$$

Make use of this inequality (twice!) in order to rewrite the energy balance above as

$$\frac{d\mathfrak{H}_{\mathrm{total}}}{dt} \le a_1 p_{\mathrm{m}}^2 + a_2[y+1]^+|p_{\mathrm{m}}|^{3/2}$$

and find the corresponding values of the constants a_1 and a_2.

(c) Were it not for the factor of $[y+1]^+$, one would be able to proceed to a bound on the growth of the energy, and thus the state variables, directly (how?). Here is a less satisfying way of proceeding: again using the inequality (9.47)*, rewrite the energy balance as*

$$\frac{d\mathfrak{H}_{\mathrm{total}}}{dt} \le a_1 p_{\mathrm{m}}^2 + a_2|p_{\mathrm{m}}|^{3/2} + b_1\left([y]^+\right)^2 + b_2|p_{\mathrm{m}}|^3$$

and find the constants b_1 and b_2.

(d) Use the fact that $([y]^+)^2 \leq 2\mathfrak{H}_{\text{total}}/\omega_0^2$ *to rewrite the inequality as*

$$e^{\alpha t}\frac{d\left(e^{-\alpha t}\mathfrak{H}_{\text{total}}\right)}{dt} \leq f(p_{\text{m}})$$

and determine the value of α *and the form of* $f(p_{\text{m}})$*. Show that this leads to the bound*

$$\mathfrak{H}_{\text{total}}(t) \leq e^{\alpha t}\mathfrak{H}_{\text{total}}(0) + \int_0^t e^{\alpha(t-t')} f(p_{\text{m}})dt'$$

How big is α *for reasonable choices of the parameters for a reed–bore system? Is this a useful bound?*

Problem 9.12 *Referring to the implementation of the finite difference scheme for the reed–bore system, as discussed in Section 9.3.3, find expressions for the intermediate variables* $a_{\cdot}^n$, $b_{\cdot}^n$, $c_{\cdot}^n$, *and* $d_{\cdot}^n$*. Given that these will be used to advance the solution from time step* n *to time step* $n+1$*, they may only depend on values of* y *and the grid function* Ψ *up through time step* n*, as well as the sampled input mouth pressure* p_{m}*.*

Problem 9.13 *Consider the simplified model of the tonehole, as described on page 275, with* $Z_{\text{a}} = 0$*, but where now* $Z_{\text{s}}(s)$ *is generalized from the form given in* (9.40) *to include loss as*

$$Z_{\text{s}} = \left(\frac{s\xi_{\text{e}}}{\phi\gamma S_{\text{T}}} + R\right) || \frac{\gamma}{(1-\phi)s\xi S_{\text{T}}}$$

where, as before, $||$ *indicates a parallel combination, and where* R *is a non-negative constant.*

Show that using this impedance, a PDE system which generalizes (9.41) *may be derived. Find also an associated finite difference scheme. Will the inclusion of the extra loss term have any bearing on numerical stability?*

Problem 9.14 *Consider the impedance relationship*

$$\hat{p} = Z(s)\hat{u} \qquad \text{with} \qquad Z(s) = \coth(\alpha s/2)$$

for some positive constant α*. Show that, in the time domain, this corresponds not to a differential equation, but to a relationship among values of* $p(t)$ *and* $u(t)$ *at two time instants separated by* α *seconds. Given that one of the impedances which appears in the tonehole model described in Section 9.3.8 is of this form, do you think it is a good idea to approximate* $Z(s)$ *as* $Z(s) \approxeq 2/(s\alpha)$*? (Take sample rate considerations into account!)*

Problem 9.15 *Consider the general PDE form of the coupling between an acoustic tube and Keefe's tonehole model, as per* (9.39)*.*

(a) Show that, under conservative boundary conditions for the tube, the total stored energy is of the form $\mathfrak{H} = \mathfrak{H}_{\text{tube}} + \mathfrak{H}_{\text{tonehole}}$*, where*

$$\mathfrak{H}_{\text{tube}} = \frac{1}{2}\|S\Psi_t\|_{\mathbb{U}}^2 + \frac{\gamma^2}{2}\|S^*\Psi_x\|_{\mathbb{U}}^2 \qquad \mathfrak{H}_{\text{tonehole}} = \frac{a_1}{2}\left(\frac{dm}{dt}\right)^2 + \frac{a_2}{2}m^2$$

when $S^* \geq 0$ *(find a condition under which this is true). Notice that* a_1*, which depends on the series impedance* Z_{a}*, is negative when the hole is closed.*

(b) Consider the following finite difference scheme:

$$\mu_{xx}S\delta_{tt}\Psi = \gamma^2\delta_{x+}\left(\mu_{x-}S^*\delta_{x-}\Psi\right) - \epsilon\delta_{t\cdot}m \qquad a_1\delta_{tt}m + a_2\mu_{t\cdot}m = \gamma\langle\delta_{t\cdot}\Psi, \epsilon\rangle_{\mathbb{U}_N}$$

Here, the grid function ϵ_l is obtained through sampling of the continuous distribution $\epsilon(x)$. Assuming conservative boundary conditions, show that the scheme is conservative, with a numerical stored energy of $\mathfrak{h} = \mathfrak{h}_{\text{tube}} + \mathfrak{h}_{\text{tonehole}}$, where

$$\mathfrak{h}_{\text{tube}} = \frac{1}{2}\|\sqrt{\mu_{xx}S}\delta_{t-}\Psi\|^2_{\mathbb{U}_N} + \frac{\gamma^2}{2}\langle\mu_{x+}S^*\delta_{x+}\Psi, e_{t-}\delta_{x+}\Psi\rangle_{\underline{\mathbb{U}_N}}$$

and find the expression for $\mathfrak{h}_{\text{tonehole}}$ (which will depend only on the time series m).

(c) Show that the term $\mathfrak{h}_{\text{tube}}$ is non-negative if the following condition is satisfied:

$$\lambda \leq \sqrt{\frac{\mu_{xx}S}{\mu_{xx}S^*}}$$

and, given the expression for S^*, rewrite this as a bound $\lambda \leq r \leq 1$, and give the explicit expression for r. Can you interpret this bound in terms of a reduced effective length of the tube? Explain.

(d) From your expression for $\mathfrak{h}_{\text{tonehole}}$, and under the closed-hole condition, show that, though there is a range of values of k, the time step, over which the expression is non-negative, in the limit of small k the term must become negative. Thus, though the scheme may operate in an apparently stable manner, it is not convergent in the strict sense.

Problem 9.16 *Consider scheme* (9.43) *for the simplified tonehole model* (9.41)*, and assume that the boundary conditions at both ends of the tube are conservative.*

(a) Show that the scheme possesses a conserved energy of the form

$$\mathfrak{h} = \mathfrak{h}_{\text{tube}} + \mathfrak{t}_{\text{tonehole}} + \mathfrak{v}_{\text{tonehole}}$$

where $\mathfrak{t}_{\text{tonehole}}$ and $\mathfrak{v}_{\text{tonehole}}$ are kinetic and potential energy terms for the tonehole analogous to those which appear in (9.42)*. Recall that $\mathfrak{h}_{\text{tube}} \geq 0$ under the condition that $\lambda \leq 1$.*

(b) Show that the potential energy term $\mathfrak{v}_{\text{tonehole}}$ is non-negative under the second of conditions (9.44)*.*

(c) By considering conditions under which the sum of $\mathfrak{t}_{\text{tonehole}}$ and $\mathfrak{v}_{\text{tonehole}}$ is non-negative, determine a more general condition relating α, ϕ, the time step k, and the various material parameters for the tonehole. This condition, combined with $\lambda \leq 1$, is a stronger stability condition for the scheme.

(d) Find an explicit update for both the grid function Ψ and the time series m.

9.6 Programming exercises

Programming Exercise 9.1 *Using the vocal synthesis Matlab code provided in Section A.10 as a starting point, generalize the glottal excitation function* `uin`*. (These manipulations are not physical modeling per se, but rather abstract synthesis techniques, as discussed in Section 1.1, which are very useful in the source-filtering context.) Here are some ways you could proceed. In the code example, the glottal waveform is crudely generated as*

$$u_{\text{in}} = [y(t)]^+ \qquad \text{where} \qquad y = \sin(2\pi f_0 t)$$

Here, $[\cdot]^+$, the clipping operation, indicates the "positive part of."

(a) Generalize the waveform $y(t)$ as follows:

$$y(t) = A(t)\sin\left(2\pi\int_0^{t'}(1 + \Delta f\,\sin(2\pi f_1 t'))\,f_0(t')dt'\right)$$

Here, $f_0(t)$ is the instantaneous frequency, f_1 a vibrato frequency, normally between 3 and 5 Hz, $0 \leq \Delta f \leq 1$ the vibrato depth, and $0 \leq A(t) \leq 1$ is an amplitude envelope. You can define these, in Matlab, by breakpoint functions, then interpolate to the sampling instants nk, for integer n (learn about the function `interp1` *for this purpose). You will also need to perform the integration in the argument of the sine function above numerically. See if you can generate a sung major scale, with the vowel /a/.*

(b) In the code example, the excitation is zero mean and positively clipped—thus its "duty cycle" is 50%. Try instead a waveform of the following form, which depends on a free parameter ϵ:

$$u_{\text{in}}(t) = \frac{[y(t) - \epsilon]^+}{1 - \epsilon} \qquad \text{with} \qquad 0 \leq \epsilon < 1$$

Have a listen to the differences in timbre as you play with this parameter. Try, also, squaring the waveform above, rendering it differentiable. (Try to guess what the perceptual result of such a squaring operation will be before listening to the output.)

Programming Exercise 9.2 *In the code example, only two vocal tract forms are provided, corresponding, roughly, to the Russian vowels /a/ and /e/ given by Fant [123]; this data, as well as forms for other vowels, appears in many texts, such as, e.g., [280]. Track down some other vocal tract configurations, and extend the Matlab code to include these. In the code example, the data is specified as position/surface area pairs $(x, S(x))$, and the data is then interpolated to the finite difference grid using* `interp1`*. Note that you must scale any data that you find such that it is non-dimensionalized, i.e., $0 \leq x \leq 1$, and $S(0) = 1$.*

In addition to generating sound examples, you may also wish to generate pictures of the formant structure corresponding to various vocal tract configurations. In order to do this, you may use, as your input function `uin`*, an impulse sequence consisting of a one followed by zeros. The formant structure may then be obtained by taking the magnitude of the Fourier transform of the output sequence of volume velocities read at the right end of the tube. Referring to difference scheme* (9.13)*, you may obtain this sequence from the grid function Ψ, by taking $u_{\text{out}} = -S_N \delta_{x\cdot} \Psi_N$.*

Programming Exercise 9.3 *Introduce wall losses to the model, by using the coupled scheme* (9.19)*, in the explicit update form you derived in Problem 9.7. In order to set the parameters for the wall itself, you may wish to begin from those given in the caption to Figure 9.7; if you are more ambitious, try to track down some data from the literature—try [273] as a starting point.*

Programming Exercise 9.4 *Implement the explicit scheme* (9.23) *in the case of a time-varying vocal tract model. In order to generate a simple time-varying function $S(x, t)$, start from the two static vocal tract configurations $S_{/\text{a}/}(x)$ and $S_{/\text{e}/}(x)$ (as given in the code example), and define $S(x, t)$ as*

$$S(x, t) = \alpha(t) S_{/\text{a}/}(x) + (1 - \alpha(t)) S_{/\text{e}/}(x) \qquad with \qquad 0 \leq \alpha(t) \leq 1$$

Here, the function $\alpha(t)$ controls the degree of interpolation. Note that you should not construct $S(x, t)$ off-line, outside the main loop, as it is a huge waste of memory—rather, $\alpha(t = nk)$ should be generated in its entirety off-line, and then the function $S(x, t = nk)$ constructed in the main loop at time step n.

Programming Exercise 9.5 *Using the vocal synthesis code example provided in Section A.10 as a starting point, modify the code such that it performs reed instrument synthesis. There are various changes you need to make:*

- *Introduce, as global parameters, the reed mass M_r, effective surface area S_r, resonant frequency ω_0, collision parameters ω_1 and α, reed loss paramater σ_0, channel width w, and reed equilibrium displacement H_0, as well as air density ρ.*
- *Derive parameters $\mathcal{Q}$, $\mathcal{R}$, and $\mathcal{S}$, as per* (9.27).
- *Adjust the radiation parameters, so as to correspond to condition* (9.11)*, for an unflanged tube.*
- *Create a mouth pressure excitation p_m—this can be, in the simplest case, a constant vector, but could be generalized to a breakpoint function, interpolated to the sampling instants using the Matlab function* `interp1`.
- *Adjust the bore profile, given as* `S`*, such that it corresponds to various typical choices of interest in wind instrument synthesis, such as: a cylinder, a cone, or a cylinder with a bell-like termination. Allow the user to specify the type of profile, as well as extra parameters, when necessary. (If the profile is cylindrical, no parameters need be supplied, but if it is a cone, for example, an extra parameter, such as the ratio of surface area at the right end of the tube to that at the left end, will be necessary.)*
- *Implement the coupling of the excitation with the tube, using the scheme as outlined in Section 9.3.3, and the various intermediate variables for which you derived expressions in Problem 9.12.*

Programming Exercise 9.6 *Extend the single-reed synthesis algorithm which you programmed above to allow the reed equilibrium displacement to be time varying. As a starting point, just in order to hear the effects of such variation, you may wish to model such variation as a constant plus a sinusoidal offset, such as*

$$H_0(t) = H_0\,(1 + \epsilon \sin(2\pi f_H t))$$

where H_0 is an average value, and where $\epsilon \leq 1$ describes the depth of the variation, and f_H its frequency.

Programming Exercise 9.7 *For the single-reed instrument model you have developed over the last two exercises, introduce a static tonehole according to the simplified model* (9.41) *and the associated scheme* (9.43)*. (You should have worked out an explicit update for the scheme in Problem 9.16.) In order to do this, you will need an expression for ξ_e, the effective tonehole length in the open-hole case. A simple approximation, in terms of the physical tonehole radius b and height τ, and where the bore radius is a at the location of the tonehole, is given as*

$$\xi_e = \frac{\tau}{L} + \frac{b}{L}\left(1.4 - 0.58\frac{b^2}{a^2}\right)$$

Your code will thus require the following new parameters: ξ, S_T (the ratio of tonehole cross-sectional area to that of the bore at its left end), x_T (the position of the tonehole), ϕ, between 0 and 1, which determines the state of the hole, as well as the free scheme parameter α.

Programming Exercise 9.8 *Extend the simulation of toneholes in the following ways:*

(a) Allow the specification of a number Q of distinct toneholes, each characterized by a set of parameters as given in the previous exercise. This extension is relatively straightforward, assuming minimal interaction between holes. First, convince yourself that the dynamics of any single tonehole

may be simulated independently of those of any other, using a natural extension of scheme (9.43)*, which will now be of the form*

$$\mu_{xx} S \delta_{tt} \Psi = \gamma^2 \delta_{x+} \left((\mu_{x-} S) \delta_{x-} \Psi \right) - \sum_{q=1}^{Q} J_p(x_{\mathrm{T}}^{(q)}) m^{(q)}$$

$$m^{(q)} = \frac{\phi^{(q)} \left(\gamma^{(q)}\right)^2 S_{\mathrm{T}}^{(q)}}{\xi_{\mathrm{e}}^{(q)}} \left(\alpha^{(q)} I_p(x_{\mathrm{T}}^{(q)}) \Psi + (1 - \alpha^{(q)}) \mu_{t\cdot} I_p(x_{\mathrm{T}}) \Psi \right)$$
$$+ (1 - \phi^{(q)}) \xi^{(q)} S_{\mathrm{T}}^{(q)} \delta_{tt} I_p(x_{\mathrm{T}}^{(q)}) \Psi$$

In Matlab, it is probably best to vectorize the tonehole states, so that a loop over the Q *toneholes may be avoided.*

(b) Allow the tonehole state ϕ *for each such tonehole to be time varying. In order to make your code "playable," you might wish to begin by specifying a list of events, characterized by: start time, hole number, duration, terminal state* ϕ*.*