

SYNTHESIS OF SPEECH SOUNDS FROM A MULTI-MASS MODEL OF THE LUNGS, VOCAL TRACT, AND GLOTTIS

Paul Boersma

Abstract

Like some existing vocal-tract models, this synthesizer views the vocal apparatus as a sequence of straight tubes. The walls of all these tubes, however, are modelled as coupled mass-spring systems, which in most models is privileged to the vocal cords. This, together with the ability to continuously vary the lengths of the tubes, leads to a principled treatment of the problems of source-filter interaction, which makes our model especially suitable for the generation of speech signals that include consonants.

The model is restricted to the last step in articulatory synthesis: from a list of time-target pairs for every articulator and a table with speaker characteristics, find the resulting acoustic signal. Among the examples that we show are various brands of voicing contrasts and their articulatory correlates.

1 The model

Figure 1 shows a simplified picture of our model. As a model of the human vocal apparatus, it is a straightened one-dimensional approximation to the curved shape of the vocal tract, glottis and lungs. It consists of a sequence of straight tubes that contain air.

These tubes could represent from left to right: the lungs, the bronchi, the trachea, the glottis, the part of the larynx above the glottis, the pharynx, the volume between the tongue dorsum and the velum, the volume between the tongue dorsum and the palate, the volume between the apex of the tongue and the alveoli, and the lip opening. This resembles the subdivision that we shall use in the examples of chapter 8.

Air is forced to flow into and out of these tubes as a result of its mass inertance and its elasticity. An acoustic output is derived from the air flow at the right boundary of the rightmost tube in figure 1: it is the sound radiated from the lips into the atmosphere.

The walls of the tubes yield to pressure changes. At the same time, the equilibrium positions of the walls can be adjusted by the articulatory muscles. The walls are, therefore, modelled as mass-spring systems. The tensions of some of these springs can be adjusted, too. This reflects the ability of the vocal cords to produce tone differences, and the ability of the supralaryngeal musculature to distinguish fortis and lenis obstruents. An acoustic output is radiated from the moving masses.

The main source of energy in the tract is the variation of lung pressure. In some models, the lungs are modelled as an ideal pressure source. In our model, lung pressure is, more realistically, brought about by decreasing the lung volume, i.e., reducing the neutral width of the leftmost tube in figure 1. The modelling of the

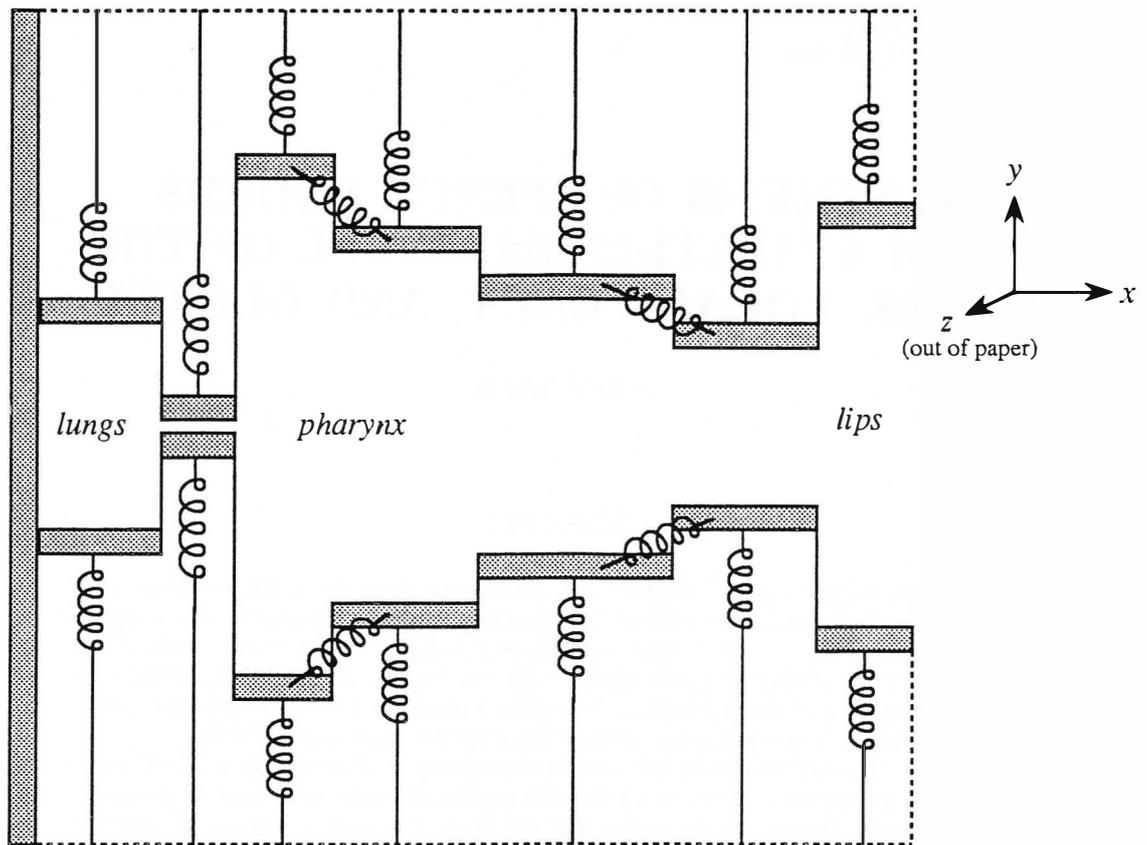


Fig. 1. Mid-sagittal view of our model of the speech apparatus. The model features a sequence of straight tubes with walls consisting of masses and springs. The leftmost of these tubes represents the lung volume and is closed at the left edge, the rightmost tube forms the opening between the lips and is open to the atmosphere at the right edge, where fluctuations in the air flow are radiated as sound. The glottis is represented by one or two tubes, which are treated exactly the same way as all other tubes. The speech muscles can alter the rest positions and the tensions of the vertical springs. Some of the masses are connected with springs to their nearest neighbours.

respiratory mechanism as lung-volume control rather than as an ideal pressure source, expresses the finite capacity of the lungs.

The walls of a tube can oscillate if they are close enough together and there is sufficient air flow along them. This follows automatically from the aerodynamic and myo-elastic equations. Thus, the vocal cords can easily vibrate in this model. Nothing withholds other articulators, though, from vibrating as well; tongue tip, velum and lips are likely candidates for producing trills.

If the particle velocity exceeds a certain threshold, noise is generated immediately downstream from the constriction; the portion of the kinetic energy that is converted into turbulence depends on the relative widths of both tubes involved.

The lengths of the tubes do not have to be equal. The upper part of the glottis, for instance, may be 1 mm thick, whereas other regions, like the pharynx, can span several centimetres (though the implementation of our model subdivides these large regions into tubes approximately 10 mm long, for reasons of computation). More important, though, is the advantage of allowing the lengths of tubes to vary with time. This permits us to model appropriately the lengthening and shortening of certain tubes by lip rounding, dorsal constriction, or up and down movements of the larynx.

The present model does not feature a nasal tract.

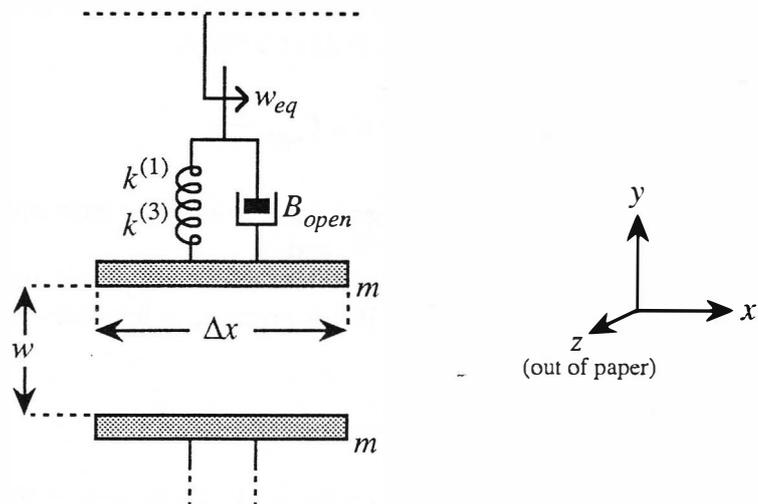


Fig. 2. Mid-sagittal view of one tube, showing its springs and masses (lower springs not shown). The articulatory muscles can directly adjust the rest position w_{eq} , the linear spring constant $k^{(1)}$, and the tube length Δx , and may also indirectly vary the mass m , the damping B_{open} , and the cubic spring constant $k^{(3)}$. All these parameters, plus the air pressure inside the tube, determine the development of the state of the tube walls, which is represented by their mutual distance w and their mutual velocity dw/dt .

2 The springs and the masses

Every tube is supposed to be enclosed along the y -axis in figure 1, by two opposing walls that consist of equal masses and springs (see figure 2). The acceleration of the walls in the y -direction is given by the total force on the two walls (which is twice the force on either wall):

$$m \frac{d^2 w}{dt^2} = \text{total force} = \text{tension force} + \text{damping force} + \text{air pressure force} \quad (1)$$

where m is the mass of either wall (in kg), and w is the distance between the two walls (in meters). The mass m need not be constant, because it is the part of the wall that actually moves; it could slowly vary in time as a function of the tension in the wall.

The *tension force* is the force in the spring that tries to bring the wall to its neutral position. It is due to the tension of the muscles inside the wall (e.g. vocalis muscle, pharyngeal constrictor muscles) and to the tension of the muscles that pull the edges of the wall (e.g. cricothyroid muscle). For small displacements from this position, the tension force is approximately proportional to the displacement. For larger displacements, a third-power term comes in. If the walls are not in contact with each other, the tension force is

$$\text{tension force} = k^{(1)} (w_{eq} - w) + k^{(3)} (w_{eq} - w)^3 \quad (2)$$

where $k^{(1)}$ is the linear spring "constant" (in N/m) of either spring, which is a function of muscle activity, w_{eq} is the equilibrium distance between the walls, which can also be adjusted by the articulatory muscles (e.g., posterior crico-arytenoid activity causes an increase of w_{eq} in the glottis, risorius does the same for the lips, and expiration is equivalent to reducing w_{eq} in the lungs), and $k^{(3)}$ is one quarter of the cubic spring constant (in N/m³) of either spring.

The *damping force* is due to internal friction in the tissue. It tries to bring the velocity of the moving wall to zero. It is proportional to this velocity:

$$\text{damping force} = -B_{open} \frac{dw}{dt} \quad (3)$$

where B_{open} is the damping (in kg/s) of either spring, which depends on the properties of the tissue and dynamically also on $k^{(1)}$, $k^{(3)}$, and m .

If the air pressure inside the tube is greater than the atmospheric pressure, the *air pressure force* tries to push the walls apart. If this pressure is less than the atmospheric pressure, it tries to pull the walls together:

$$\text{air pressure force} = 2 P \Delta x \Delta z \quad (4)$$

where P is the mean air pressure inside the tube, Δx is the length of the tube, and Δz the third dimension of the tube, making $\Delta x \Delta z$ the area of the wall. The factor 2 appears because there are two walls.

When the two masses approach one another, they collide and fold into each other. This leads us into modifying the equation of motion, which becomes

$$m \frac{d^2 w}{dt^2} = k^{(1)} (w_{eq} - w) + k^{(3)} (w_{eq} - w)^3 + F_s^{(1)} + F_s^{(3)} - (B_{open} + B_{closed}) \frac{dw}{dt} + 2 P \Delta x \Delta z \quad (5)$$

where the *stiffness forces* $F_s^{(1)}$ and $F_s^{(3)}$ (in Newtons) represent the reaction of the tissue against being pressed together, and B_{closed} is the damping inside the compressed tissue. Because the masses are not exactly parallel, the collision is not simultaneous for all points along the z -axis. Figure 3 shows a cross-sectional view of our stylization of this process; the walls smoothly close upon one another, like a zipper. The force due to the linear part of the stiffness can be computed from the average penetration depth (over the whole width Δz) and is then found to be

$$F_s^{(1)} = \begin{cases} 0 & \text{for } w \geq \Delta w \\ \frac{s^{(1)} (\Delta w - w)^2}{4 \Delta w} & \text{for } -\Delta w \leq w \leq \Delta w \\ -s^{(1)} w & \text{for } w \leq -\Delta w \end{cases} \quad (6)$$

where $s^{(1)}$ is the linear spring constant of the tissue stiffness. The cubic part is computed from the average cubed depth and is expressed as

$$F_s^{(3)} = \begin{cases} 0 & \text{for } w \geq \Delta w \\ \frac{s^{(3)} (\Delta w - w)^4}{8 \Delta w} & \text{for } -\Delta w \leq w \leq \Delta w \\ -s^{(3)} w (w^2 + \Delta w^2) & \text{for } w \leq -\Delta w \end{cases} \quad (7)$$

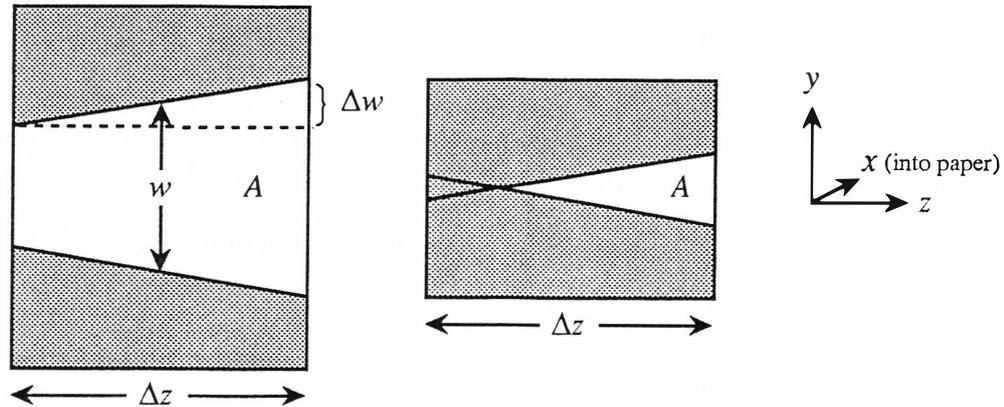


Fig. 3. Cross-sectional view of a tube, showing how the walls make contact. In the right figure, the mean distance w between the masses is $\Delta w/2$. This distance can even become negative. The cross section A stays positive as long as w is greater than $-\Delta w$.

where $s^{(3)}$ is one quarter of the cubic spring constant of the tissue stiffness. Both forces $F_s^{(1)}$ and $F_s^{(3)}$ are smooth functions of w (they are differentiable in $-\Delta w$ and in Δw).

If the walls of tube m are connected to the walls of the adjacent tubes $m-1$ and $m+1$, the coupling stiffness of the spring is expressed as

$$\begin{aligned}
 & k_{m-1,m}^{(1)} (w_{m-1} - w_m) + k_{m-1,m}^{(3)} (w_{m-1} - w_m)^3 \\
 & + k_{m,m+1}^{(1)} (w_{m+1} - w_m) + k_{m,m+1}^{(3)} (w_{m+1} - w_m)^3
 \end{aligned} \tag{8}$$

This term should be added to the tension forces.

The cross section A of a tube is computed from the distance w between the walls and is approximately w times Δz if this distance is larger than Δw . For smaller distances, the cross section is determined from figure 3. However, if we allowed very small values of A (which would appear when w comes near $-\Delta w$), the aerodynamics would show unrealistic behaviour. This is because the existence of very small values of the volume of air in a tube causes very high positive or negative pressures to arise immediately before or after the moment of contact. We shall circumvent this by allowing a very small leakage w_{min} through every tube, giving for the cross section A a smooth function of the distance w (again differentiable in $-\Delta w$ and in Δw):

$$A = \begin{cases} (w + w_{min}) \Delta z & \text{for } w \geq \Delta w \\ \left(\frac{(\Delta w + w)^2}{4 \Delta w} + w_{min} \right) \Delta z & \text{for } -\Delta w \leq w \leq \Delta w \\ w_{min} \Delta z & \text{for } w \leq -\Delta w \end{cases} \tag{9}$$

A good value for w_{min} is 0.01 mm. In this case, the relative changes in A during a sampling period are not too large (if $\Delta w \geq w_{min}$), whereas the amount of air that leaks through the orifice is negligible due to the large viscous resistance.

3 Air flows and pressures

3.1 Continuity of mass flow

The principle of the conservation of mass is expressed as follows: “The increase during a certain amount of time of the mass contained in a volume is equal to the mass that flows into that volume during that time minus the mass that leaves that volume during that time”.

In three dimensions this is formulated as the continuity equation of hydrodynamics (see e.g. Landau & Lifshitz, 1953):

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{v}) = 0$$

where $\rho(\mathbf{x}, t)$ is the mass density of the fluid in units of kilograms per cubic meter, and $\mathbf{v}(\mathbf{x}, t)$ is the particle velocity in units of meters per second. A “particle” is considered an infinitesimally small piece of the fluid; it is supposed to contain, however, infinitely many molecules.

We will now consider the one-dimensional case of a straight tube with rigid walls, where air flows only in the x -direction (the left-right direction in figure 1). If the transverse velocities v_y and v_z are zero, the continuity equation reduces to

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0$$

where v is now the signed velocity in the x -direction, averaged over all y - and z -coordinates inside the tube.

A special almost one-dimensional case is our straight tube with moving walls. Inside a straight tube with time-varying cross section $A(t)$ (in units of square meters), the velocity components v_y and v_z are not zero, but the continuity equation retains a simple form:

$$\frac{\partial(\rho A)}{\partial t} + \frac{\partial(\rho v A)}{\partial x} = 0 \quad (10)$$

It is of crucial importance here that the area A should be inside the parentheses in equation (10). Note, e.g., that in the case of an incompressible fluid

$$\frac{\partial A}{\partial t} + \frac{\partial(vA)}{\partial x} = 0$$

which states correctly that an incompressible fluid will flow out of any shrinking tube. Flanagan and Ishizaka (1977) show that this effect is negligible for vocal cord vibration. It can, however, probably not be neglected in modelling lung volume changes and variations in the tension of the supralaryngeal musculature like those that are partly responsible for voicing contrasts in obstruent consonants.

If the boundaries of the tube are moving in the x -direction, i.e., the tube becomes longer or shorter in time, equation (10) receives a change that is true to the definition of mass conservation given above, giving approximately

$$\frac{\partial(\rho A \Delta x)}{\partial t} + (\rho v A)_{\text{right}} - (\rho v A)_{\text{left}} = 0 \quad (11)$$

where Δx is the length of the tube, and v is the particle velocity relative to the velocity of the left or right boundary.

3.2 Equation of fluid motion

The second principle describes the influence of forces on the motion of air: "The particle experiences a force in the down-hill direction of the pressure gradient. At the same time, its velocity is impeded by viscous friction with other particles that have a different velocity."

In three dimensions this is expressed as

$$\rho \frac{d\mathbf{v}}{dt} = -\text{grad } P + \mu \Delta \mathbf{v}$$

where $P(\mathbf{x}, t)$ is the pressure expressed in Newtons per square meter, μ is the coefficient of shear viscosity, which is $1.86 \cdot 10^{-5}$ Ns/m² for air, and Δ is the Laplace operator. From this is derived the Navier-Stokes equation in three dimensions:

$$-\text{grad } P = \rho \frac{d\mathbf{v}}{dt} - \mu \Delta \mathbf{v} = \rho \left(\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{2} \text{grad } v^2 - \mathbf{v} \times \text{curl } \mathbf{v} \right) - \mu \Delta \mathbf{v}$$

Inside a straight tube, neglecting v_y and v_z , this reduces to

$$\rho \frac{\partial v_x}{\partial t} = -\frac{\partial P}{\partial x} - \rho v_x \frac{\partial v_x}{\partial x} + \mu \Delta v_x$$

The viscous resistance of a tube shaped as in figure 3 can be solved from the boundary condition that $v_x = 0$ on the walls. For $\Delta w \ll \Delta z$, this yields in the stationary-flow approximation

$$R_{\text{visc}} = \frac{-\mu \Delta v_x}{v} = -\frac{\mu}{v} \frac{\partial^2 v_x}{\partial y^2} = \begin{cases} \frac{12 \mu}{w^2 + \Delta w^2 + w_{\text{min}}^2} & \text{for } w \geq \Delta w \\ \frac{12 \mu}{\frac{1}{2}(\Delta w + w)^2 + w_{\text{min}}^2} & \text{for } -\Delta w \leq w \leq \Delta w \\ \frac{12 \mu}{w_{\text{min}}^2} & \text{for } w \leq -\Delta w \end{cases} \quad (12)$$

where v is v_x averaged over all y between the plates. If there are other resistances, they are subsumed under one term $R v$:

$$\rho \frac{\partial v}{\partial t} + \frac{\partial P}{\partial x} + \rho v \frac{\partial v}{\partial x} + R v = 0 \quad (13)$$

One such resistance is due to turbulent noise and will be introduced in section 3.6.

3.3 The equation of state

The third equation relates the air mass density ρ to the air pressure P . If we assume that the processes that we are interested in, are so swift that we can neglect heat conduction in the fluid, the temperatures vary with the material pressure, and no air particles exchange any heat. In this case, the relation between pressure and mass density is given by the adiabatic pressure law

$$\frac{P}{P_0} = \left(\frac{\rho}{\rho_0} \right)^\gamma$$

where P_0 and ρ_0 are a reference pressure and density, and γ is a constant of the fluid, equal to apx. 1.4 for a diatomic gas like air. If we choose for P_0 and ρ_0 the atmospheric pressure and density (no flow, no temperature gradient) P_{atm} and ρ_{atm} , a differential pressure change is given by

$$dP = P_{atm} \gamma \frac{d\rho}{\rho_{atm}} \equiv c^2 d\rho$$

for some constant c that depends only on temperature and mean pressure and that has the dimensions of a velocity. The value of this constant, which is called the *velocity of sound*, can be computed as 353 m/s, for $P_{atm} = 1.013 \cdot 10^5 \text{ N/m}^2$ and $\rho_{atm} = 1.14 \text{ kg/m}^3$.

From now on, the pressure P is taken as the difference of the real pressure and the atmospheric pressure, so that for small pressures the equation of state is approximated by the well-known equations

$$P = c^2 (\rho - \rho_{atm}) \tag{14}$$

$$dP = c^2 d\rho$$

3.4 The aerodynamic equations in terms of continuous quantities

On the boundary between two tubes of different area, most quantities show a discontinuity. The particle velocity, for example, suddenly increases when the fluid flows into a narrower tube. At the same time, the pressure suddenly drops (Bernoulli effect), and, as a consequence, the density drops as well. Some quantities are continuous, however, on tube boundaries. One of them is the *flow of mass*: the amount of air that leaves a certain tube through one of its boundaries during a certain period of time must enter the adjacent tube in the same time span, and no matter is lost. This means that the *mass flow* J , which is a vector defined as

$$J \equiv \rho v A \tag{15}$$

and which is expressed in kilograms per second, is continuous at the boundaries of tube sections. The other continuous quantity is the *flow of energy*: the energy that leaves a tube through one of its boundaries must enter the adjacent tube without delay. If the fluid flow stays laminar in crossing the boundary, the particle keeps moving in

the x -direction, and we can implicitly define the *continuous pressure*, which is denoted by Q and expressed in Newtons per square meter, by

$$dQ \equiv dP + \frac{1}{2} \rho d(v^2) \quad (16)$$

Thus, $\Delta Q = 0$ over the boundaries of tubes. Note that this equation expresses the conservation of the sum of a potential energy term dP and a kinetic energy term $\frac{1}{2} \rho d(v^2)$. Bernoulli's law states that $dQ = 0$ for stationary inviscid non-turbulent flow. To a good approximation (1% error in $\frac{1}{2} \rho v^2$, if $v < 0.2 c$), equation (16) is equivalent to stating that the continuous quantity is

$$Q \approx P + \frac{1}{2} \rho v^2 \quad (17)$$

The third quantity that is continuous in the hydrodynamics of unbounded fluids, is the flow of momentum ρv . In our case, this is not continuous: at tube boundaries, the air transfers momentum to the vertical walls, and vice versa.

As so happens, the quantities J and Q are exactly the ones that appear in the divergence and gradient parts of the aerodynamic equations (10) and (13). The three one-dimensional aerodynamic equations can thus be written as

$$\begin{aligned} \frac{\partial(\rho A)}{\partial t} + \frac{\partial J}{\partial x} &= 0 \\ \rho \frac{\partial v}{\partial t} + \frac{\partial Q}{\partial x} + R v &= 0 \\ dP &= c^2 d\rho \end{aligned} \quad (18)$$

3.5 Boundary conditions

No air flows into the lungs from the bottom. The condition at the leftmost boundary in figure 1 is therefore:

$$J = 0 \quad (19)$$

At the rightmost boundary, sound radiates from the lips. For a round orifice, the relation between Q and J can be expressed as (Sondhi & Resnick, 1983)

$$\frac{\partial Q}{\partial t} - \frac{\partial \left(\frac{cJ}{A_{lip}} \right)}{\partial t} + \frac{Qc}{a_{lip}} = 0 \quad (20)$$

where a_{lip} is the radius of the opening at the lips. This radius is taken constant at 2 centimetres in order not to attain unrealistically low damping values for small lip openings; radiation damping will still be smaller for small openings than for large ones, due to the A_{lip} factor in (20).

At the time point $t = 0$, all flows and pressures vanish.

3.6 Turbulence

Turbulent air movements are generated when air with a velocity greater than some critical velocity v_{crit} flows out of a narrower tube with cross section $A_<$ into a tube with a larger cross section $A_>$. This causes an energy loss given by the pressure drop

$$Q_{turb} = \frac{1}{2} \rho v (|v| - v_{crit}) \left(1 - \frac{A_<}{A_>}\right)^2 \quad (21)$$

In reality, this is not a pressure drop, but a failure to completely recover from the Bernoulli pressure drop $\frac{1}{2} \rho v^2$. Equation (21) is in accordance with equation (6) of Ishizaka & Flanagan (1972), which describes the pressure recovery if $v_{crit} = 0$. From Van den Berg et al. (1957), the critical velocity can be computed as 10 m/s (critical volume velocity 200 cm³/s, area 1.07 x 20 mm; for a larger opening, they found a greater critical volume velocity, which suggests a constant critical velocity).

The energy loss gives rise to an extra resistance term in equation (13):

$$R_{turb} v = \frac{Q_{turb}}{\Delta x} \quad (22)$$

where Δx is the length of the tube section.

Ishizaka & Flanagan (1972) use a similar resistance not only at the exit of the glottis, but also at the entrance of the glottis. The pressure drop there, however, is just the Bernoulli pressure for a *vena contracta*, i.e., the stream is contracted and the area of the entrance is smaller than would be expected from the distance of the walls (though Ishizaka & Flanagan acknowledge this, they do not use this smaller area in their subsequent computations). Therefore, if the flow at the inlet is laminar, the "resistance" represents no energy loss, and the pressure loss is recovered somewhere in the glottis. In this case, the effect can be neglected, and equation (21) would be approximately right in predicting, for $A_< = 0.1 A_>$, a turbulence loss of 0.81 relative to the Bernoulli pressure, comparing favourably with Van den Berg's measured value of 0.875, as opposed to the value of 1.19 predicted by Ishizaka & Flanagan.

Thus, turbulence causes a pressure drop due to the loss of kinetic energy in the x -direction. This energy is converted into chaotic particle movements with the same kinetic energy, one third of which is again in the x -direction. If we assume that this acoustic energy is translated into a low-pass Gaussian noise with a cut-off frequency f_{cutoff} , then the noise pressure to be added to the pressure at the boundary between the sections is

$$Q_{noise}(t) = (1 - 2\pi f_{cutoff} dt) Q_{noise}(t - dt) + \sqrt{\frac{1 - (1 - 2\pi f_{cutoff} dt)^2}{3}} Q_{turb}(t) N(t) \quad (23)$$

where $N(t)$ is Gaussian white noise whose power is unity. In our implementation, a frequency cut-off is automatically caused as a side-effect of our method of integration: if the longest tubes are approximately 1 cm long, the cut-off frequency of the integration is just above 6 kHz; the only noise pressure that remains to be added to the continuous pressure is the second term of equation (23).

4 Articulatory parameters and their time evolution

In our implementation of the model, the only parameters that are governed by the articulatory muscles are:

1. The equilibrium wall distances w_{eq} of the lungs, the glottis, the pharynx, the velum, the palate, the apex, and the lips.
2. The relative spring constant k_{rel} of the vocal cords. This influences the linear spring constants $k^{(1)}$, the masses m , and the dampings B_{open} in the glottis.
3. The relative spring constant k_{rel} of the supralaryngeal tubes. This influences the linear spring constants, masses and dampings of the vocal tract walls.
4. The relative tube lengths Δx_{rel} of the lips (rounding) and those of the pharynx (variations in larynx height).

These eleven quantities are, therefore, slowly varying functions of time. In our implementation, their values are interpolated linearly between the nearest target values specified. For instance, if the spring k is specified as k_1 at a time t_1 and as k_2 at a time t_2 , and there are no specifications for k at times between t_1 and t_2 , the spring at a time t between t_1 and t_2 is expressed as

$$k(t) = k_1 + \frac{t - t_1}{t_2 - t_1} (k_2 - k_1) \quad (24)$$

At least two targets have to be specified for each articulatory dimension:

1. The starting points at $t = 0$. The starting values of w_{eq} are the starting values of the distances w as well.
2. The end points at $t = T$, which is the time at which the simulation stops.

5 Speaker characteristics

The vocal apparatus is thought to consist of 10 or 11 regions: lungs, bronchi, trachea, glottis (one region for a one-mass model, or two regions for a two-mass model of the vocal cords), upper part of larynx, pharynx, velum, palate, apex, and lips. Some of these regions consist of one tube (glottis), some consist of several tubes; for reasons of correct simulation of short wavelengths, these tubes must not be much smaller than 1 cm. Within each region, the tubes have equal properties (if our model should be applied to questions of supraglottal articulation, instead of focussing on glottis-tract interaction, we should use a more sophisticated model of the transition from articulator positions to tube properties). Our model needs the following constant speaker-specific parameters for every region:

1. The number of tubes that this region is divided into.
2. The equilibrium length Δx_{eq} .
3. The third dimension ("breadth") Δz .
4. The "zipperiness" Δw .
5. The equilibrium mass of a wall m_{eq} .
6. The equilibrium linear spring constant $k_{eq}^{(1)}$.
7. The cubic spring constant $k^{(3)}$.
8. The damping ratio B_{rel} .
9. The equilibrium distance w_{eq} if it cannot be adjusted by the articulatory muscles: this goes for the bronchi, the trachea, and the upper part of the larynx.

In table 1, we define three simple model speakers (meant to sound like a woman, a man, and a young child, respectively). In the following sections, we will elaborate on the individual roles of the various parameters.

Table 1. Dimensions of the vocal tracts of three speakers (a woman, a man and a child). The tract is divided into 11 regions for a two-mass vocal-cord model (the adults), or into 10 parts for a one-mass model (the child). Each part consists of several tubes, whose lengths are not much more than 1 cm.

Name of the region	Number of tubes	Equilibrium tube length Δx_{eq} (mm)	Third dimension Δz (mm)	Zipperiness Δw (mm)	Equilibrium wall distance w_{eq} (mm)
Lungs	23	9, 10, 6	207, 230, 138	0.01	
Bronchi	6	9, 10, 6	28, 30, 19	0.01	15
Trachea	6	9, 10, 6	15, 16, 10	0.01	15
Lower glottis	1	1.4, 2, 1	10, 18, 6	0.01	
Upper glottis	1, 1, 0	0.7, 1, -	10, 18, -	0.01	
Larynx	2	9, 10, 6	15, 16, 10	0.01	15
Pharynx	6, 6, 4	9, 10, 6 \pm	28, 30, 19	0.01	
Velum	3	9, 10, 6 \pm	28, 30, 19	0.01	
Palate	4	9, 10, 6	28, 30, 19	0.01	
Apex	3	9, 10, 6	28, 30, 19	0.01	
Lips	4	4.5, 5, 3 \pm	19, 20, 12	0.01	

5.1 The dimensions of the tubes

The lengths Δx of some tubes can vary in time:

1. The length of the lip region depends on spreading and rounding. For spread lips (risorius muscle contracted), it is only half the value found in table 1. For rounded lips (orbicularis oris muscle contracted), it is twice the value of table 1. Note that the lips can be brought together or drawn apart without spreading or rounding.
2. The length of the pharynx region depends on the height of the larynx. Contraction of the stylohyoid muscle can decrease the pharynx length by 20%, whereas contraction of the sternohyoid muscle can increase it by 30%. We neglect the influences that these muscles have on the shape of what is below the glottis.
3. The length of the velum region depends on the height of the tongue body, i.e., it becomes larger if the area of this section gets smaller. This is due to the curvature of the vocal tract near the velum. In the present implementation, we neglect this.

The lengths depend on the articulatory parameter Δx_{rel} as

$$\Delta x = \Delta x_{rel} \Delta x_{eq} \quad (25)$$

Table 1 further expresses the following disproportions:

1. Male vocal cords are relatively thick (large Δx) and relatively long (large Δz), whereas those of the child are relatively thin and short.
2. The child's pharynx is relatively short (only four sections).

Not in table 1 are the following properties of the shape of the tract:

1. The leakage w_{min} , which is 0.01 mm for all regions of all three model speakers.
2. The critical velocity v_{crit} , which is 10 m/s for all regions of all three speakers.

The zipperiness Δw is taken to be the minimum needed for smooth contact.

Table 2. Properties of the walls of the vocal tracts of the same speakers as in table 1.

	Equilibrium mass of one tube m_{eq} (g)	Equilibrium tension of one tube $k_{eq}^{(1)}$ (N/m)	Damping ratio B/B_{crit}	Relative coupling between regions
Lungs	72, 80, 48	200	0.8	0
Bronchi	9, 10, 6	40	0.8	0
Trachea	4, 5, 3	160	0.8	0
Lower glottis	0.02, 0.1, 0.005	10, 12, 8	0.1	1
Upper glottis	0.01, 0.05, -	4, 4, -	0.6	0
Larynx	4, 5, 3	40	0.8	0
Pharynx	9, 10, 6	40	0.8	0
Velum	9, 10, 6	40	0.3	0
Palate	9, 10, 6	40	0.8	0
Apex	9, 10, 6	40	0.3	0
Lips	6, 6, 4	10	0.5	0

5.2 Masses and tensions

If the muscles in the walls of the tube sections contract, the masses of the vibrating parts of these walls are decreased and their tensions increased, according to

$$m = \frac{m_{eq}}{k_{rel}}, \quad k^{(1)} = k_{eq}^{(1)} k_{rel} \quad (26)$$

where the articulatory parameter k_{rel} is the dimensionless relative tension (Flanagan & Landgraf, 1968; Ishizaka & Flanagan, 1972). For the vocal cords, for instance, this relative tension is made greater than 1 by the combined efforts of the cricothyroid muscle (which pulls the cords from the outside) and by the vocalis muscle (which contracts them from within). Note that for small displacements the frequency of free oscillation is proportional to k_{rel} :

$$f \approx \frac{1}{2\pi} \sqrt{\frac{k^{(1)}}{m}} = \frac{k_{rel}}{2\pi} \sqrt{\frac{k_{eq}^{(1)}}{m_{eq}}}$$

The speakers of table 2 have equilibrium lower-glottis resonance frequencies of 112.5, 55.1, or 201.3 Hz, and equilibrium upper-glottis resonances at 100.7 Hz and 45.0 Hz.

The cubic spring constant is chosen as

$$k^{(3)} = k_{eq}^{(1)} \left(\frac{10}{\Delta z} \right)^2 \quad (27)$$

which means that for a relative tension $k_{rel} = 1$, the distance where the third-power force equals the linear force, is $\Delta z/10$. For pharynx, velum and palate, the Δz in equation (27) is doubled, due to the different attachment of the muscles in the cheeks.

The linear tissue stiffness is proportional to the equilibrium area of the wall:

$$s^{(1)} = (5 \cdot 10^6 \text{ N/m}^5) \Delta x_{eq} \Delta z \quad (28)$$

and the cubic stiffness constant is chosen to be

$$s^{(3)} = \frac{s^{(1)}}{(0.9 \text{ mm})^2} \quad (29)$$

If the tension of the tissue is isotropic, the coupling-spring constants *within* the regions of the tract are approximately equal to the average of $\frac{1}{2}(\Delta z/2\Delta x_{eq})$ times the linear spring constants of the separate masses, whereas the cubic spring constants of the coupling springs approximately equal the averages of $\frac{1}{2}(\Delta z/2\Delta x_{eq})^3$ times those of the separate masses. No couplings are found, however, *between* the regions of the tract, except between the two parts of the glottis, as seen in table 2.

5.3 Damping

Critical damping is the damping that allows a spring to reach equilibrium as quickly as possible without oscillations. As is seen from equations (26), the critical damping of the walls does not depend on k_{rel} for small displacements:

$$B_{open,crit} = 2 \sqrt{k m} \approx 2 \sqrt{k_{eq} m_{eq}}$$

However, we prefer to have damping that is constant relative to the true critical damping, which involves the cubic spring constants as well. Otherwise, the relaxation times of the oscillations would be longer in the cubic-force region than in the linear-force region, instead of the other way around. We write therefore

$$B_{open,crit} = 2 \sqrt{k_{eff} m} \quad (30)$$

where the *effective spring constant* is

$$k_{eff} = - \frac{\partial(\text{tension force})}{\partial w} = k^{(1)} + 3 k^{(3)} (w_{eq} - w)^2 \quad (31)$$

The damping ratios in table 2 are relative to this critical damping:

$$B_{open} = B_{rel} B_{open,crit} \quad (32)$$

The damping of the compressed tissue is chosen to be critical:

$$B_{closed} = B_{closed,crit} = 2 \sqrt{s_{eff} m_{eff}} \quad (33)$$

where the *effective mass* m_{eff} is the mass of the wall that is in contact with its counterpart

$$m_{eff} = \begin{cases} 0 & \text{for } w \geq \Delta w \\ m \frac{\Delta w - w}{2 \Delta w} & \text{for } -\Delta w \leq w \leq \Delta w \\ m & \text{for } w \leq -\Delta w \end{cases} \quad (34)$$

and the *effective stiffness* s_{eff} is

$$s_{eff} = \begin{cases} 0 & \text{for } w \geq \Delta w \\ \frac{\Delta w - w}{2 \Delta w} (s^{(1)} + s^{(3)} (\Delta w - w)^2) & \text{for } -\Delta w \leq w \leq \Delta w \\ s^{(1)} + s^{(3)} (3 w^2 + \Delta w^2) & \text{for } w \leq -\Delta w \end{cases} \quad (35)$$

If some tissue should be able to vibrate, the open-damping ratio must be appreciably smaller than 1. Flanagan and Landgraf (1968), for instance, use a damping factor of 0 for the vocal cords in an open glottis; Ishizaka and Flanagan (1972) use factors of 0.1 and 0.6 for the lower and upper glottis, respectively.

5.4 Viscosity

The viscous resistance is basically computed from equation (12). However, this cannot be correct for the lung region, as this region is subdivided into many "parallel" branches. For the lung region, therefore, the viscosity is multiplied by

$$1 + \frac{\Delta x_{lungs} - x}{\Delta x_{lungs}} (\text{parallel subdivision} - 1) \quad (36)$$

where Δx_{lungs} is the total length of the lung region, x is the distance of the centre of each tube to the bottom of the lungs, and *parallel subdivision* is a number that we choose to be 1000.

6 Implementation

The aerodynamic and myo-elastic differential equations are integrated by a finite-differencing method that is described in detail in the Appendix. The time step Δt of this integration is taken as the time that it takes sound to travel the smallest tube length (this is the largest time that guarantees a stable integration).

The *state* of the system at a time $n\Delta t$ is defined by the distances w_m^n between the walls, their velocities \dot{w}_m^n , and the lengths Δx_m^n , for every tube m from the first tube in the lungs ($m = 1$) to the last tube in the lips ($m = M$), and by the values of the continuous quantities J_m^n and Q_m^n at the tube boundaries $m = 0 \dots M$ (see figure A1 in the Appendix).

The changes in the state of the system between the times $n\Delta t$ and $(n+1)\Delta t$ come from three sources: the aerodynamic equations (see chapter 3), the myoelastic equations (chapter 2), and the articulation data (chapter 4).

The resulting acoustic pressure at a certain *distance* from the mouth is derived from the flow at the lips and from the vibrations of the walls:

$$\text{output pressure} = \frac{1}{4\pi \Delta t \text{ distance}} \left(J_M^{n+1} - J_M^n + \rho_{air} \Delta x \Delta z \sum_{m=1}^M (\dot{w}_m^{n+1} - \dot{w}_m^n) \right) \quad (37)$$

7 Comparison to Ishizaka and Flanagan (1972)

The purposes of both their model and ours is a simulation of the interaction between glottis and vocal tract, that accounts for the facts of speech. Some of the differences between the two models are:

1. By I&F, the pressures are computed inside the tubes, whereas the flows are computed at the boundaries; except in the glottis, where pressures as well as flows are computed inside. Our integration computes the flows at the same places as the pressures, for all tubes. This might be a more principled modelling of the glottis-tract interaction.
2. The use of different tube lengths enhances our chances of treating glottis and tract alike.
3. The use of time-varying lengths enables our system to model some extra articulatory gestures that are found in the sounds of the languages of the world.
4. The direct influence of moving walls on flows allows us to model the lungs as a finite-capacity volume, and consonants that use sucking.
5. The influence of air pressure on wall motion allows us to model the acoustic correlates of the articulatory feature "tense".
6. The smooth closing of the walls prevents spurious pressure peaks and allows walls to vibrate even if there is no complete closure.
7. Turbulence is used as an acoustic pressure source, making frication noise possible in most tubes.
8. Damping is dynamic: if the cubic spring forces play a role, there is larger damping.
9. The cubic spring constant does not depend on the linear spring constant, which is true at least to the model of a perfect string (we do not know if this is an improvement).
10. In I&F's glottis, a turbulence resistance is found above the glottis, regardless of the direction of the flow in the glottis. This is not realistic.
11. We neglect the effects of vena contracta, because we decided that our speakers have a laminar flow there.

8 Examples

Figure 5 shows the voicing of an [a] and an [u] (male speaker). The rest width of the glottis is 1 mm, the relative tension is 1.5, and the equilibrium lung width is reduced from 100 to 90 mm during the first 0.1 seconds (we used a very strong coupling here, effectively simulating a one-mass model of the vocal cords). It is seen that the frequency of oscillation is lower in [u] than it is in [a]. This could simply be due to the fact that the amplitude of vocal-cord oscillation is higher in [u] (the constriction causes a higher mean pressure in the glottis; this pushes the cords apart). By the way, the acoustic power radiated in [u] is 17 dB lower than the power radiated in [a].

Figure 6 show how the lips move if there is a closing gesture between 0.1 and 0.2 seconds (from an equilibrium width of 40 mm to -10 mm), and an opening gesture between 0.3 and 0.4 seconds. This articulation results in an [aʔa]-like utterance if it is superimposed on the [a] of figure 5. Figure 6 also shows the lung volume as a function of time. It seems to follow only reluctantly the equilibrium volume that was set by the expiration muscles after 0.1 seconds. This is because the relatively narrow glottis blocks the expiration. From 0.2 to 0.3 seconds, the glottis is open (see figure 7a), but the lung volume does not decrease at all, because no air is allowed to escape into the atmosphere. The lung pressure is also constant during this period; between 0.1 and 0.2

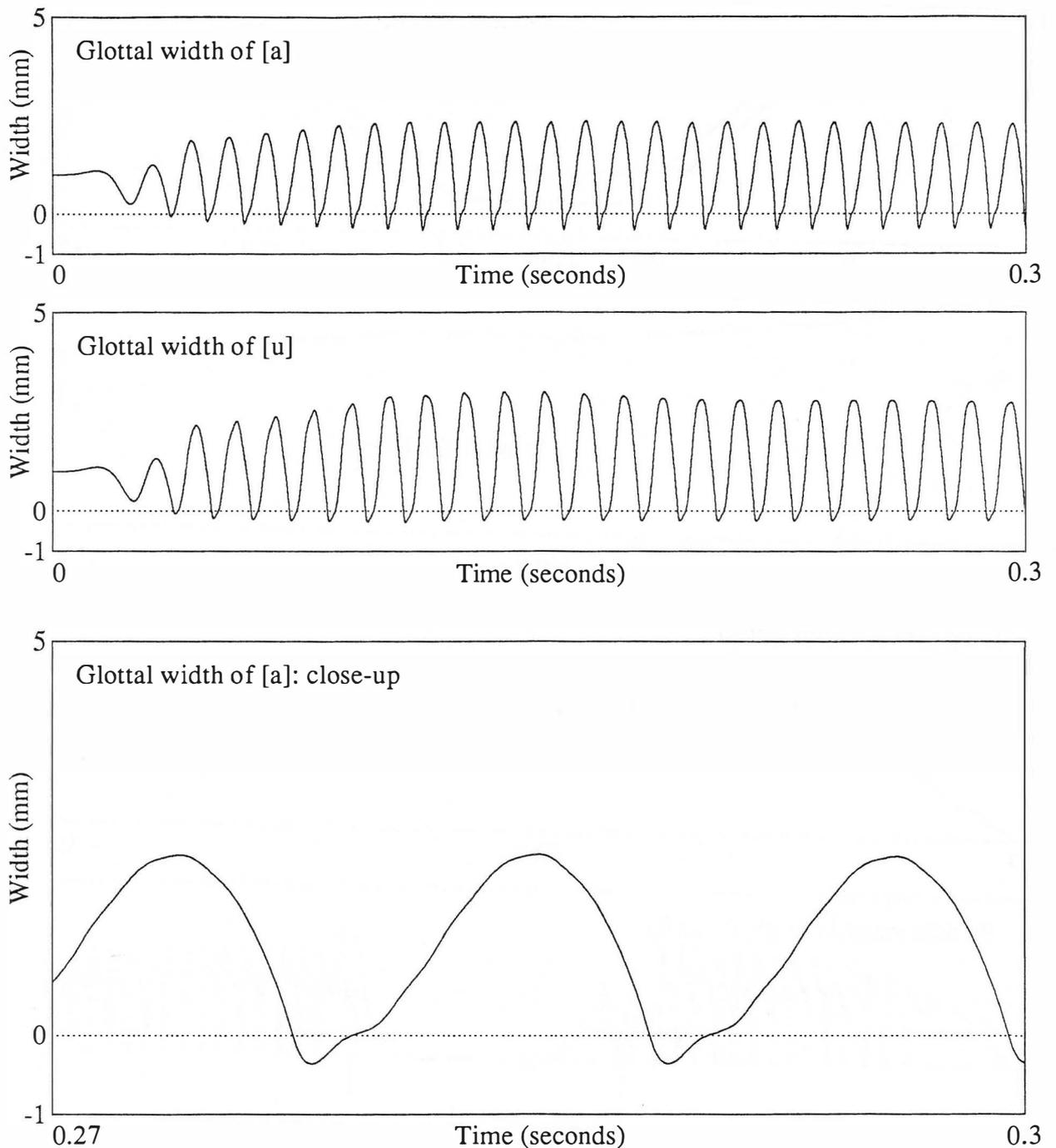


Fig. 5. Realized width of the glottis as a function of time, for an [a]-like and an [u]-like shape of the vocal tract.

seconds, and between 0.3 and 0.5 seconds, the pressure is seen to fall from 5 to 3 cm water. Finally, figure 6 shows the particle velocity in the glottis. During the labial closure, it is small, because the glottis is open. Comparison with figure 7a shows that the velocity oscillation lags somewhat behind the glottal-width oscillation; this phase difference keeps the cords vibrating.

Figure 7 shows several attempts to make [apa]- and [aba]-like utterances from the same contraction of the lungs and lip movements that were used in figure 6. In figure 7a, there are no articulatory gestures, apart from the lip movements, and voicing stops one period after the closing of the lips (the rest width of the glottis stays 1 mm, the relative tension of the vocal cords stays 1.5). We see that the glottal width during oral

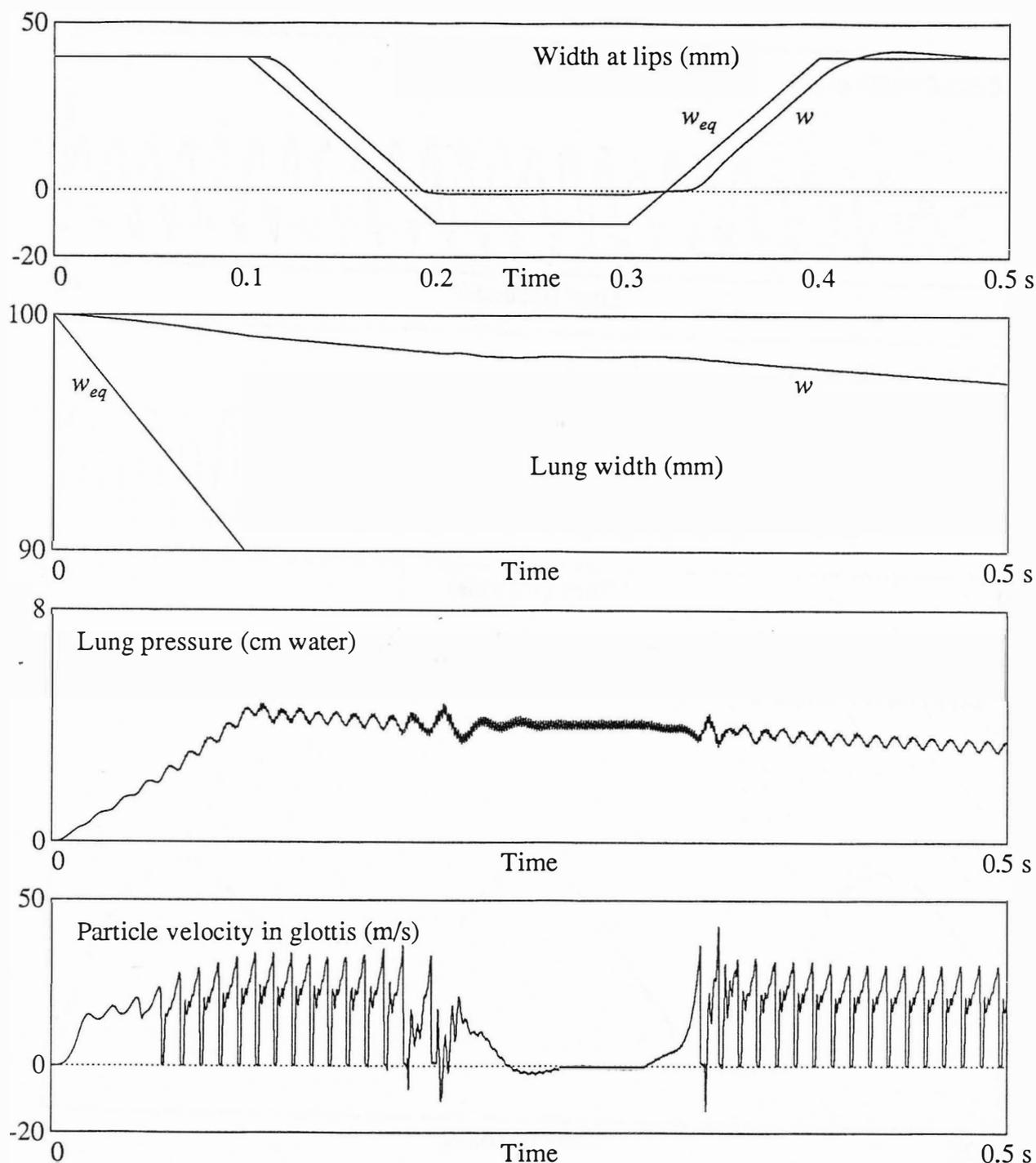


Fig. 6. Some articulatory and realized properties of an [apa]-like articulation. The top figure shows the target position (equilibrium width) of the lips (straight lines) and the position realized. The second figure shows the same for the lungs. The third figure shows the pressure at the bottom of the lungs. The bottom figure shows the mean particle velocity in the glottis.

closure is greater than the maximum width during phonation, even without any activity of the larynx.

The consonant can be made more voiceless by spreading the glottis to 4 mm (in a movement synchronous with the movement of the lips), as is seen in figure 7b. This gesture is used in aspiration. Another voiceless consonant is made if the glottis is constricted to -1 mm (figure 7d); this is used in ejectives, in which the release burst, deprived from pulmonic excitation, is obtained by pulling up the larynx. Yet another voiceless consonant is made if the vocal-tract walls are stiff (figure 7f).

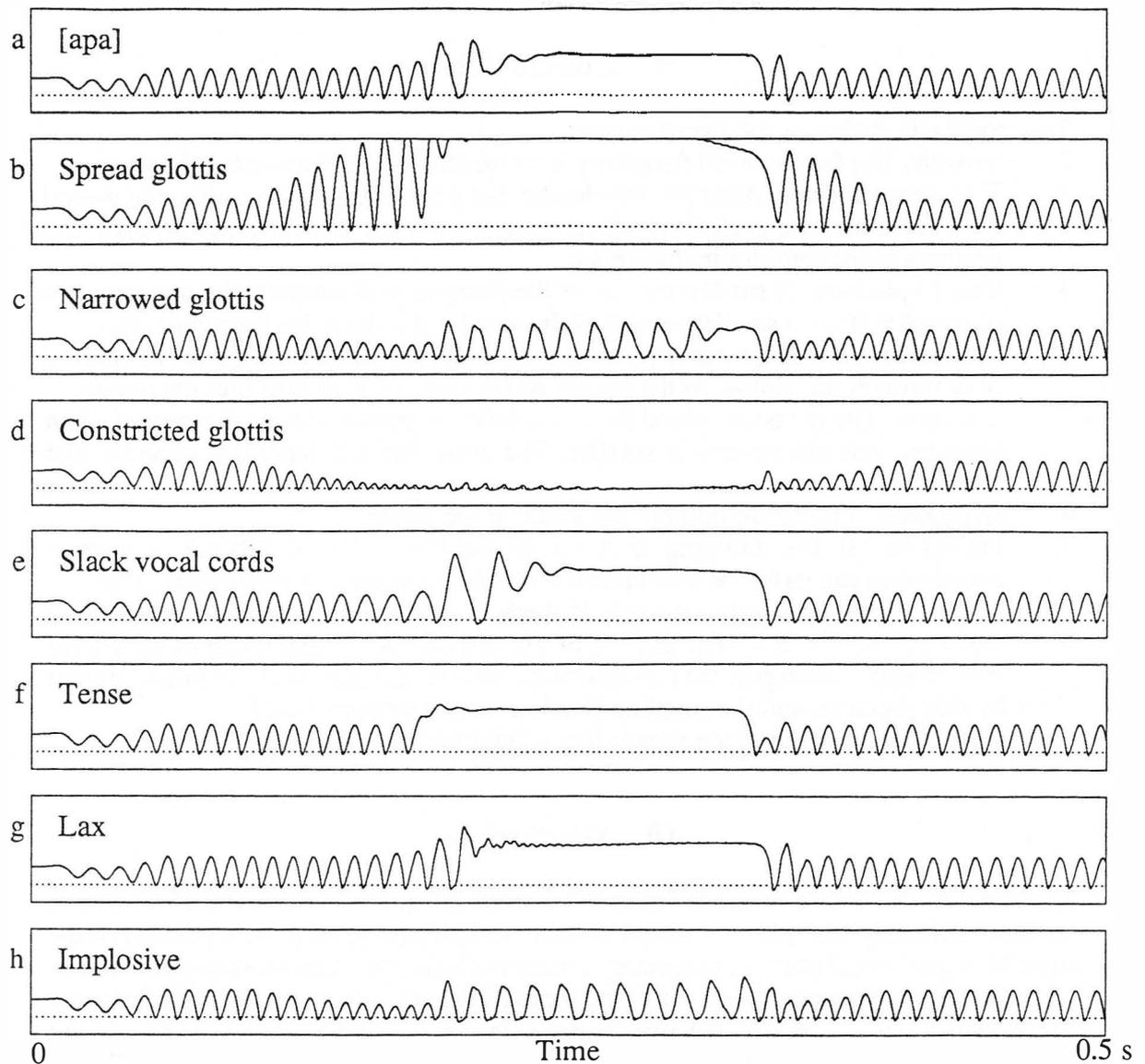


Fig. 7. The glottal widths of eight [aba]- and [apa]-like utterances as functions of time. The plots have been scaled between -1 and $+5$ millimetres.

Voicing can be maintained longer if the vocal cords are brought together (to 0 mm), as we see in figure 7c. If this narrowing is done during phonation, the acoustic power diminishes, as is also seen in figure 7c. Note that the fundamental frequency falls even if there is no adjustment of the tension of the vocal cords. An additional lowering of the larynx between 0.2 and 0.4 seconds so that at 0.4 seconds the length of the pharynx is 1.3 times longer than it was at 0.2 seconds, helps voicing considerably (figure 7h); this is the implosive articulation.

Making the vocal cords slack during closure causes phonation to continue only a little longer (figure 7e). The same goes for reducing the tension of the pharynx walls (this is the “lax” or “lenis” articulation, see figure 7g); if the cubic spring constant of the speaker is reduced as well, the consonant does become very voiced (not in the figure), which points to the possibility that the perfect-string approximation is not realistic here. Thus, voicing can be maintained during labial closure, either by relaxing the wall muscles in the vocal tract, or by simultaneously narrowing the glottis and lowering the larynx (thus making an implosive consonant).

9 Conclusion

The model can make the following sounds:

1. Vowels. The fundamental frequency is influenced by the vocal-tract shape.
2. Voiceless plosives. After an oral closure, the intra-oral pressure rises, the glottal flow falls, and the cords stop vibrating and part. All this does not involve any gestures of the muscles in the larynx.
3. Voiced plosives. A smaller tension of the pharynx wall supports the maintenance of a glottal flow; a smaller tension of the vocal cords helps the Bernoulli force.
4. Implosives. The larynx is lowered during oral closure; this helps to maintain a flow through the glottis. At the release of the stop, air is sucked into the mouth.
5. Ejectives. The pressure behind the constriction is greater than in the case of plain plosives, and the volume is smaller. The noise burst is therefore stronger and shorter.
6. Aspiration. Enlarge the equilibrium width of the glottis.
7. Fricatives. If the damping is too high for the walls to vibrate, a narrow constriction can easily be maintained; this is accompanied by turbulence noise.
8. Clicks. These can only be made if there is a coupling between the different regions in the mouth. After making two constrictions, the rest width of the region in-between is enlarged, the pressure falls, the two constrictions are pulled tighter by this pressure, and the coupling finally causes a strong release.
9. Trills. The damping of the apex is low enough for it to be able to vibrate.

10 Discussion

The focus of this article is on the "manner" features of speech production. If we like to model realistically the "place" features as well, we have to use other subdivisions of the supralaryngeal vocal tract. In particular, a more realistic transition from the position of the articulators to the area functions of the supraglottal vocal tract has been modelled by Mermelstein (1973). Ideally, the articulatory parameters should be the activities of the twenty-odd most important muscles, instead of our seven equilibrium positions, two relative tensions and two relative lengths.

Breathy voicing can be included by making Δw of the vocal cords dynamically dependent on w_{eq} , thus maintaining a triangular rest shape of the glottis. The vocal cords can then vibrate without closing. However, realistic values of Δw (about 1 mm) do not lead to vibration, due to the rigidity of the walls in our model. Breathy voicing is heard in our model output if Δw is about 0.1 mm.

The model makes no nasal sounds. A nasal tract, however, can be added in a straightforward manner.

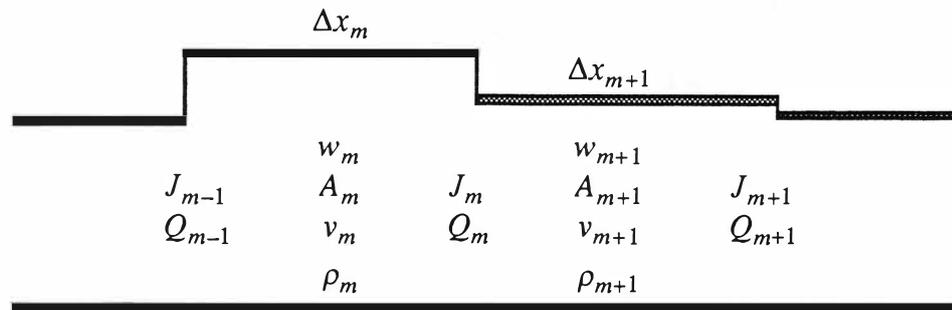


Fig. A1. Some inner tubes, showing some aerodynamic quantities with place labels. The only quantities defined on boundaries, are the mass flow J and the continuous pressure Q .

Appendix: difference equations

Difference equations relate the continuous quantities at a time $t + (n+1)\Delta t$ to those at a time $t + n\Delta t$, where Δt is the *sampling period* or *integration-time step*, which will be constant in our evaluations. They will do so for every tube m from the lungs ($m = 1$) to the lips ($m = M$) and for every tube boundary m from the bottom of the lungs ($m = 0$) through the inner boundaries ($m = 1 \dots M-1$, where M is the total number of tubes) to the place where sound radiates from the lips ($m = M$).

In this section we present the complete algorithm for the integration of the myo-elastic and aerodynamic differential equations. In deriving the states of the springs at the time $(n+1)\Delta t$ from those at the time $n\Delta t$, a simple explicit integration scheme suffices, because the spring motion is relatively slow and sufficiently damped. For the hyperbolic part of the aerodynamic equations we will use a second-order accurate integration scheme that is based on the *Lax-Wendroff* method, which consists of the following two steps (see Mitchell 1969, Press et al. 1989):

1. Compute the half-way values (at a time $(n+\frac{1}{2})\Delta t$) of the "conserved fluxes" (in our case, the mass flow J and the continuous pressure Q), from the values at time $n\Delta t$, using first-order accurate explicit integration (forward Euler). Second-order accuracy could be achieved by taking into account the values at $(n-\frac{1}{2})\Delta t$, which makes this the *staggered-leapfrog* method instead of Lax-Wendroff; in our case, however, this would lead to instabilities.
2. Use these half-way values for computing the values at the new time $(n+1)\Delta t$ to second-order accuracy.

For our case, the Lax-Wendroff method needs two modifications:

1. Because of the discontinuities at the tube boundaries, we have to average in a way the left- and right-limit values of the mass flow densities and of the masses. Their proper weighting is suggested by the results of an integration along characteristics for constant and equal tube lengths.
2. Because the lengths of the tubes are not equal, we also have to average the distances that appear in the gradient parts of the equations of motion. The method of averaging that we use, produces the correct resonance frequencies in the tract.

The initial state of the system is defined as (upper indices count time steps, lower indices count tubes or tube boundaries)

$$\begin{aligned}
 J_m^0 &= 0 & , & & Q_m^0 &= 0 & & \text{for } m = 0 \dots M \\
 w_m^0 &= w_{eq,m}^0 \text{ or } w_{eq,m} & , & & \dot{w}_m^0 &= 0 & & \text{for } m = 1 \dots M \\
 \Delta x_m^0 &= \Delta x_{rel,m}^0 \Delta x_{eq,m} \text{ or just } \Delta x_{eq,m} & & & & & & \text{for } m = 1 \dots M
 \end{aligned} \tag{A1}$$

For an integration of the time evolution of the system, we write the aerodynamic differential equations in the following way:

$$\begin{aligned}
 \frac{\partial(\rho A \Delta x)}{\partial t} &= J_{left} - J_{right} \\
 \frac{\partial(\rho v)}{\partial t} &= - \frac{\partial Q}{\partial x} - R v
 \end{aligned} \tag{A2}$$

The quantities that appear in the left-hand side of (A2) are called the *mass* $\rho A \Delta x$ and the *momentum density* ρv . We can write the continuous quantities J and Q in terms of these as

$$\begin{aligned}
 J &= (\rho v) A \\
 Q &= \left(\frac{\rho A \Delta x}{A \Delta x} - \rho_{atm} \right) c^2 + \frac{(\rho v)^2}{2 \rho_{atm}}
 \end{aligned} \tag{A3}$$

The inverse relations are given by

$$\begin{aligned}
 \rho v &= \frac{J}{A} \\
 \rho A \Delta x &= \left(\rho_{atm} + \frac{Q}{c^2} \right) A \Delta x - \frac{J^2 \Delta x}{2 \rho_{atm} c^2 A}
 \end{aligned} \tag{A4}$$

For every sampling period, we proceed by the following steps:

Step 1: "Compute the momentum densities, the masses of air, and the state variables in the tubes, from the mass flows and continuous pressures at the tube boundaries."

The mean values of the continuous quantities inside tube m are

$$\bar{Q}_m^n = \frac{1}{2} (Q_{m-1}^n + Q_m^n) \quad , \quad \bar{J}_m^n = \frac{1}{2} (J_{m-1}^n + J_m^n) \tag{A5}$$

The mean momentum density is

$$(\rho v)_m^n = \frac{\bar{J}_m^n}{A_m^n} \quad (\text{A6})$$

The mean value of the state variable P (to be used in the mass-spring equations) is

$$P_m^n = \bar{Q}_m^n - \frac{\left((\rho v)_m^n \right)^2}{2 \rho_{atm}} \quad (\text{A7})$$

The mean values of the state variables ρ and v are

$$\rho_m^n = \rho_{atm} + \frac{P_m^n}{c^2}$$

$$v_m^n = \frac{(\rho v)_m^n}{\rho_m^n} \quad (\text{A8})$$

This velocity v can now be used to compute the resistances. Finally, the total mass of air inside the tube is computed as

$$(\rho A \Delta x)_m^n = \rho_m^n A_m^n \Delta x_m^n \quad (\text{A9})$$

Step 2: "Compute the new cross sections of the tubes from the old cross sections, the articulation data and the old mean pressures, using first-order explicit integration. Interpolate to find the half-way values of the cross sections."

The second-order differential equation (5) is divided up into two parts as

$$\dot{w}_m^{n+1} = \frac{w_m^n + \frac{\Delta t}{m_m^n} \left(tension_m^n + 2 P_m^n \Delta z_m \Delta x_m^n \right)}{1 + \frac{B_m^n \Delta t}{m_m^n}} \quad (\text{A10})$$

$$w_m^{n+1} = w_m^n + \dot{w}_m^{n+1} \Delta t$$

where the tension is computed from (2), (6), (7) and (8), and the damping from (30) through (35). This integration is first-order explicit for the harmonic part, and first-order implicit for the dissipative part. The method of the integration of the harmonic part, which uses the new value of \dot{w} to compute the new w , conserves energy.

The new value of the cross section A_m^{n+1} is derived from w_m^{n+1} with the help of equation (9), and the half-way value of the cross section is interpolated as

$$A_m^{n+1/2} = \frac{1}{2} (A_m^{n+1} + A_m^n) \quad (\text{A11})$$

Step 3: "Compute the new lengths of the tubes, from the articulation data. Interpolate to find the half-way values of these lengths."

$$\Delta x_m^{n+1} = \Delta x_{rel,m}^n \Delta x_{eq,m} \quad (\text{A12})$$

$$\Delta x_m^{n+1/2} = \frac{1}{2} (\Delta x_m^{n+1} + \Delta x_m^n)$$

Step 4: "Compute the half-way values of the mean mass flow densities inside the tubes, from the old mass flow densities inside the tubes and the old continuous pressures at the tube boundaries, using first-order explicit integration. From these, compute the half-way values of the mean mass flows inside the tubes."

For the sake of stability, we use implicit (backward) integration for the resistance part:

$$(\rho v)_m^{n+1/2} = (\rho v)_m^n + \frac{1}{2} \Delta t \left(\frac{Q_{m-1}^n - Q_m^n}{\Delta x_m^n} - R_m^n \frac{(\rho v)_m^{n+1/2}}{\rho_{atm}} \right)$$

or

$$\left(1 + \frac{R_m^n \Delta t}{2 \rho_{atm}} \right) (\rho v)_m^{n+1/2} = (\rho v)_m^n + \frac{1}{2} \Delta t \frac{Q_{m-1}^n - Q_m^n}{\Delta x_m^n} \quad (\text{A13})$$

$$\bar{J}_m^{n+1/2} = (\rho v)_m^{n+1/2} A_m^{n+1/2}$$

Step 5: "Compute the half-way values of the masses inside the tubes, from the old masses inside the tubes and the old mass flows at the tube boundaries, using first-order explicit integration. From these and the half-way values of the mass flow densities inside the tubes, compute the half-way values of the mean continuous pressures inside the tubes."

$$\begin{aligned}
(\rho A \Delta x)_m^{n+1/2} &= (\rho A \Delta x)_m^n + \frac{1}{2} \Delta t (J_{m-1}^n - J_m^n) \\
\bar{Q}_m^{n+1/2} &= \left(\frac{(\rho A \Delta x)_m^{n+1/2}}{A_m^{n+1/2} \Delta x_m^{n+1/2}} - \rho_{atm} \right) c^2 + \frac{((\rho v)_m^{n+1/2})^2}{2 \rho_{atm}}
\end{aligned} \tag{A14}$$

Step 6: "Compute the new mass flows at the tube boundaries from their old values and from the half-way values of the mean continuous pressures inside the tubes, using an appropriate weighting of the left- and right-limit values of the new mass flow densities at the tube boundaries."

The momentum density is not continuous at tube boundaries. The equation of motion is therefore integrated with the use of equation A4:

$$\left(\frac{r_m^n}{A_m^{n+1}} + \frac{r_{m+1}^n}{A_{m+1}^{n+1}} \right) J_m^{n+1} = \left(\frac{1}{A_m^n} + \frac{1}{A_{m+1}^n} \right) J_m^n + \frac{4\Delta t}{\Delta x_m^{n+1/2} + \Delta x_{m+1}^{n+1/2}} (\bar{Q}_m^{n+1/2} - \bar{Q}_{m+1}^{n+1/2}) \tag{A15}$$

where

$$r_m^n \equiv 1 + \frac{R_m^n \Delta t}{\rho_{atm}} \quad \text{for } m = 1 \dots M$$

Equation (A15) is second-order accurate for the hyperbolic part, and only first-order accurate for the dissipative part (for dissipation, only an implicit, first-order accurate integration guarantees results that are stable in the sense of reaching equilibrium faster when there are stronger resistances; whereas second-order accuracy features stability only in the Von Neumann sense).

Step 7: "Compute the new continuous pressures at the tube boundaries from their old values, from the half-way values of the mean mass flows inside the tubes, and from the new mass flows at the tube boundaries, using an appropriate weighting of the left- and right limit values of the new "masses" at the tube boundaries."

The masses are not continuous at boundaries. Again, we must use equation (A4):

$$\begin{aligned}
& \left(\rho_{atm} c^2 + Q_m^{n+1} \right) \left(A_m^{n+1} \Delta x_m^{n+1} + A_{m+1}^{n+1} \Delta x_{m+1}^{n+1} \right) - \frac{(J_m^{n+1})^2}{2\rho_{atm}} \left(\frac{\Delta x_m^{n+1}}{A_m^{n+1}} + \frac{\Delta x_{m+1}^{n+1}}{A_{m+1}^{n+1}} \right) \\
&= \left(\rho_{atm} c^2 + Q_m^n \right) \left(A_m^n \Delta x_m^n + A_{m+1}^n \Delta x_{m+1}^n \right) - \frac{(J_m^n)^2}{2\rho_{atm}} \left(\frac{\Delta x_m^n}{A_m^n} + \frac{\Delta x_{m+1}^n}{A_{m+1}^n} \right) \\
& \quad + 2 \Delta t \left(\bar{J}_m^{n+1/2} - \bar{J}_{m+1}^{n+1/2} \right) c^2
\end{aligned} \tag{A16}$$

This integration is second-order accurate, too, as it is equally balanced in time. The boundary conditions at the lungs are

$$J_0^{n+1} = 0 \quad (\text{A17})$$

$$(\rho_{atm}c^2 + Q_0^{n+1}) A_1^{n+1} \Delta x_1^{n+1} = (\rho_{atm}c^2 + Q_0^n) A_1^n \Delta x_1^n - 2 \Delta t \bar{J}_1^{n+1/2} c^2$$

and the conditions at the lips are derived from integrating equation (20) to second-order precision as

$$\begin{aligned} 0 &= Q_M^{n+1} - Q_M^n - \frac{cJ_M^{n+1}}{A_M^{n+1}} + \frac{cJ_M^n}{A_M^n} + \frac{c\Delta t}{a_{lip}} \frac{Q_M^{n+1} + Q_M^n}{2} \\ &= \frac{Q_M^{n+1}}{r_{rad}} - \frac{Q_M^n}{g_{rad}} - \frac{cJ_M^{n+1}}{A_M^{n+1}} + \frac{cJ_M^n}{A_M^n} \end{aligned} \quad (\text{A18})$$

where

$$r_{rad} \equiv \frac{1}{1 + \frac{c\Delta t}{2a_{lip}}}, \quad g_{rad} \equiv \frac{1}{1 - \frac{c\Delta t}{2a_{lip}}}$$

so that the new flow and pressure at the lips are computed from

$$\frac{r_M^n + r_{rad}}{A_M^{n+1}} J_M^{n+1} = \frac{1 + r_{rad}}{A_M^n} J_M^n + \frac{2\Delta t}{\Delta x_M^{n+1/2}} (\bar{Q}_M^{n+1/2} - Q_M^n) + \left(Q_M^n - \frac{r_{rad} Q_M^n}{g_{rad}} \right) c^{-1} \quad (\text{A19})$$

$$\frac{Q_M^{n+1}}{r_{rad}} = \frac{Q_M^n}{g_{rad}} + c \left(\frac{J_M^{n+1}}{A_M^{n+1}} - \frac{J_M^n}{A_M^n} \right)$$

The space-averaging that occurs in equations (A15), (A16) and (A19), was suggested by the results of an integration along the characteristics of the hyperbolic parts of the aerodynamic equations. These characteristics are the lines $x=x_0 \pm ct$, and an integration along them can only be done if the lengths of the tubes are equal and constant. This integration runs as follows. An alternative way of writing the aerodynamic difference equations (18) is

$$0 = \frac{\partial J}{\partial x} + A \frac{\partial \rho}{\partial t} + \rho \frac{\partial A}{\partial t} = \frac{\partial J}{\partial x} + \frac{A}{c^2} \left(\frac{\partial Q}{\partial t} - \rho v \frac{\partial v}{\partial t} \right) + \rho \frac{\partial A}{\partial t}$$

$$0 = \rho \frac{\partial \left(\frac{J}{\rho A} \right)}{\partial t} + \frac{\partial Q}{\partial x} + R v = \frac{1}{A} \frac{\partial J}{\partial t} - \frac{v}{A} \frac{\partial(\rho A)}{\partial t} + \frac{\partial Q}{\partial x} + R v$$

$$= \frac{1}{A} \frac{\partial J}{\partial t} + \frac{\partial Q}{\partial x} + \frac{v}{A} \frac{\partial J}{\partial x} + R v$$

These equations can be made to look more similar:

$$0 = c \frac{\partial J}{\partial x} + A \frac{\partial Q}{\partial ct} - J \frac{\partial v}{\partial ct} + \rho c^2 \frac{\partial A}{\partial ct}$$

$$0 = c \frac{\partial J}{\partial ct} + A \frac{\partial Q}{\partial x} + v \frac{\partial J}{\partial x} + R v A$$

These equations are integrated along the x -coordinate over the entire m th tube, which has a constant length Δx :

$$0 = c (J_m - J_{m-1}) + \frac{1}{2} A_m \frac{\partial(Q_{m-1} + Q_m)}{\partial ct} \Delta x + \left(-\bar{J}_m \frac{\partial v_m}{\partial ct} + \rho_m c^2 \frac{\partial A_m}{\partial ct} \right) \Delta x$$

$$0 = \frac{1}{2} c \frac{\partial(J_{m-1} + J_m)}{\partial ct} \Delta x_m + A_m (Q_m - Q_{m-1}) + v_m (J_m - J_{m-1}) + R_m v_m A_m \Delta x$$

Integrating over time yields:

$$0 = \frac{1}{2} c \left(J_m^n + J_m^{n+1} - J_{m-1}^n - J_{m-1}^{n+1} \right) \Delta ct$$

$$+ \frac{1}{2} A_m^{n+1/2} \left(Q_m^{n+1} + Q_{m-1}^{n+1} - Q_m^n - Q_{m-1}^n \right) \Delta x$$

$$+ \left(-\bar{J}_m^{n+1/2} (v_m^{n+1} - v_m^n) + \rho_m^{n+1/2} c^2 (A_m^{n+1} - A_m^n) \right) \Delta x$$
(A20)

$$0 = \frac{1}{2} c \left(J_{m-1}^{n+1} + J_m^{n+1} - J_{m-1}^n - J_m^n \right) \Delta x$$

$$+ \frac{1}{2} A_m^{n+1/2} \left(Q_m^n + Q_m^{n+1} - Q_{m-1}^n - Q_{m-1}^{n+1} \right) \Delta ct$$

$$+ \left(v_m^{n+1/2} (J_m^{n+1/2} - J_{m-1}^{n+1/2}) + R_m^{n+1/2} J_m^{n+1} \Delta x / \rho_m^{n+1/2} \right) \Delta ct$$

We can define (approximate) the two correction pressures (they represent the deviation from simple acoustic waves):

$$P_{WB,m}^{n+1/2} \equiv \frac{\rho_m^n c^2 (A_m^{n+1} - A_m^n) - \bar{J}_m^n (v_m^n - v_m^{n-1})}{A_m^{n+1/2}}$$

$$P_{C,m}^{n+1/2} \equiv \frac{v_m^n (J_m^n - J_{m-1}^n)}{A_m^{n+1/2}}$$

If $\Delta x = \Delta ct$, we can lose some terms (this is called integration along characteristics; a vocal-tract integration that uses this is found in Sondhi & Resnick (1983)). Equations (A20) are added to one another to give

$$\frac{r_m^n c J_m^{n+1} - c J_{m-1}^n}{A_m^{n+1/2}} + Q_m^{n+1} - Q_{m-1}^n + P_{WB,m}^{n+1/2} + P_{C,m}^{n+1/2} = 0 \quad (\text{A21})$$

Subtraction of equations (A20) gives

$$\frac{c J_m^n - r_m^n c J_{m-1}^{n+1}}{A_m^{n+1/2}} - Q_m^n + Q_{m-1}^{n+1} + P_{WB,m}^{n+1/2} - P_{C,m}^{n+1/2} = 0 \quad (\text{A22})$$

A boundary between two tube sections is a left and a right boundary at the same time. Thus, equation (A22) can also be written as

$$\frac{c J_{m+1}^n - r_{m+1}^n c J_m^{n+1}}{A_{m+1}^{n+1/2}} - Q_{m+1}^n + Q_m^{n+1} + P_{WB,m+1}^{n+1/2} - P_{C,m+1}^{n+1/2} = 0 \quad (\text{A23})$$

We can now solve J_m^{n+1} from equations (A21) and (A23):

$$\left(\frac{r_m^n}{A_m^{n+1/2}} + \frac{r_{m+1}^n}{A_{m+1}^{n+1/2}} \right) c J_m^{n+1} = \quad (\text{A24})$$

$$\frac{c J_{m-1}^n}{A_m^{n+1/2}} + \frac{c J_{m+1}^n}{A_{m+1}^{n+1/2}} + Q_{m-1}^n - Q_{m+1}^n - P_{WB,m}^{n+1/2} - P_{C,m}^{n+1/2} + P_{WB,m+1}^{n+1/2} - P_{C,m+1}^{n+1/2}$$

The new pressures Q_m^{n+1} are computed from J_m^{n+1} by equations (A21) or (A23) or directly by

$$\left(\frac{A_m^{n+1/2}}{r_m^n} + \frac{A_{m+1}^{n+1}}{r_{m+1}^n} \right) Q_m^{n+1} = \frac{A_m^{n+1/2}}{r_m^n} \left(Q_{m-1}^n - P_{WB,m}^n - P_{C,m}^n + cJ_{m-1}^n \right) + \frac{A_{m+1}^{n+1/2}}{r_{m+1}^n} \left(Q_{m+1}^n - P_{WB,m+1}^n + P_{C,m+1}^n - cJ_{m+1}^n \right) \quad (\text{A25})$$

Nothing flows into or out of the lungs other than via the windpipe:

$$J_0^{n+1} = \bar{0} \quad (\text{A26})$$

This can be combined with equation (A23) to give

$$Q_0^{n+1} = Q_1^n - \frac{cJ_1^n}{A_1^{n+1/2}} - P_{WB,1}^n + P_{C,1}^n \quad (\text{A27})$$

At the lips, equation (A18) can be combined with the equation for the rightmost tube (A21), which yields

$$\left(\frac{r_M^n}{A_M^{n+1/2}} + \frac{r_{rad}}{A_M^{n+1}} \right) cJ_M^{n+1} = \frac{cJ_{M-1}^n}{A_M^{n+1/2}} + \frac{r_{rad} cJ_M^n}{A_M^n} + Q_{M-1}^n - \frac{r_{rad} Q_M^n}{g_{rad}} - P_{WB,M}^n - P_{C,M}^n \quad (\text{A28})$$

after which Q_M^{n+1} is computed from (A18) or from (A21). Note that if v is small, and $\partial A/\partial t=0$, and $\Delta x=c\Delta t$, then (A19) and (A28) are the same formula.

REFERENCES

- Flanagan, J.L. & L.L. Landgraf (1968): "Self-oscillating source for vocal-tract synthesizers", *IEEE Transactions on Audio and Electroacoustics* AU-16: 57-64. Reprinted in J.L. Flanagan & L.R. Rabiner (eds.) (1973): *Speech Synthesis*, Dowden, Hutchinson & Ross, Stroudsburg.
- Flanagan, J.L. & K. Ishizaka (1977): "Acoustic characterization and computer simulation of the air volume displaced by the vibrating vocal cords: lateral and longitudinal motion", in: R. Carre, R. Descout & M. Wajskop (eds.): *Articulatory Modeling and Phonetics*, Proceedings of a Symposium held at Grenoble, G.A.L.F. Groupe de la Communication Parlée.
- Ishizaka, K. & J.L. Flanagan (1972): "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell System Technical Journal* 51: 1233-1268. Reprinted in J.L. Flanagan & L.R. Rabiner (eds.) (1973): *Speech Synthesis*, Dowden, Hutchinson & Ross, Stroudsburg.
- Landau, L.D. & E.M. Lifshitz (1953): *Gidrodinamika*. French translation: *Mécanique des fluides*, Editions Mir, Moscow 1971.
- Mermelstein, P. (1973): "Articulatory model for the study of speech production", *Journal of the Acoustical Society of America* 53: 1070-1082.
- Mitchell, A.R. (1969): *Computational Methods in Partial Differential Equations*, John Wiley, London.
- Press, W.H., B.P. Flannery, S.A. Teukolsky & W.T. Vetterling (1989): *Numerical Recipes in Pascal*, Cambridge University Press.

Sondhi, M.M. & J.R. Resnick (1983): "The inverse problem for the vocal tract: Numerical methods, acoustical experiments, and speech synthesis", *Journal of the Acoustical Society of America* **73** (3): 985-1002.