Simulation of flow and acoustics in the vocal tract

Peter Birkholz and Dietmar Jackèl
Institute for Computer Sciences, University of Rostock, Albert-Einstein-Str. 21, 18055 Rostock, Germany (Europe)
phone: +49 (381) 498-3428, fax: +49 (381) 498-3426
email:{piet,dj}@informatik.uni-rostock.de

Introduction

A widely used approach for the simulation of vocal tract acoustics is the transmission line model (TLM) with lumped elements [2, 3]. The TLM is based on one-dimensional acoustic propagation in a tube with a piecewise constant cross-sectional area. The governing equations for the TLM were originally derived under the following assumptions: The sound field causes only small perturbations of the thermodynamic equilibrium and the particle velocities are small compared to the speed of sound. However, the assumption of small fluid velocities in the vocal tract is often not justifiable. For instance at the constriction of fricatives or shortly after the release of plosives the particle velocities can exceed 0.1 Mach [1, p. 47-51]. The peak velocities of the airflow through the glottis during normal phonation are of the same order of magnitude, too. The consequences of these high velocities are the Bernoulli effect and, if the occasion arises, energy losses due to flow separation from the vocal tract walls at sudden expansions (shock losses) or due to flow contractions at sudden constrictions (e.g., at the inlet of the glottis).

These effects are not captured by the linear equations of sound propagation. Therefore, with regard to an acoustic simulation of the vocal tract system in the time domain, we present in this paper the spatial discretization of the equation of motion which *includes* the nonlinear "Bernoulli term". However, we approximate this term by a linear expression during the *temporal* discretization. The results of the implementation of this term in our vocal tract simulation will be discussed at the end of the paper.

Equation of motion

Conservation of momentum in a one-dimensional, frictionless fluid is fully described by Euler's equation,

$$-\frac{\partial p}{\partial x} = \rho\frac{\partial v}{\partial t} + \rho v\frac{\partial v}{\partial x}, \quad (1)$$

where p is the pressure, ρ is the density, v the particle velocity and x the spatial coordinate. In acoustics, density and pressure are each written as the sum of a static part and a small perturbation part: $p = p_0 + p'$ und $\rho = \rho_0 + \rho'$. When we furthermore substitute $v\frac{\partial v}{\partial x}$ by $\frac{1}{2}\frac{\partial v^2}{\partial x}$, Eq. (1) becomes

$$-\frac{\partial(p_0 + p')}{\partial x} = (\rho_0 + \rho')\frac{\partial v}{\partial t} + (\rho_0 + \rho')\frac{1}{2}\frac{\partial v^2}{\partial x}. \quad (2)$$

This equation is usually linearized in acoustics by neglecting products of small quantities (ρ', p' and v). This yields the well known equation of motion for the one-dimensional sound field:

$$-\frac{\partial p'}{\partial x} = \rho_0\frac{\partial v}{\partial t}.$$

However, in the vocal tract, velocities are not small in any case, so that the v^2-term in Eq. (2) should not be neglected. Thus, we use the following equation as starting point for the discretization:

$$-\frac{\partial p'}{\partial x} = \rho_0\frac{\partial v}{\partial t} + \frac{\rho_0}{2}\frac{\partial v^2}{\partial x}. \quad (3)$$

Spatial discretization

For the computer simulation, Eq. (3) must be transformed into a discrete representation. The grid for the spatial discretization is provided by the division of the vocal tract in short abutting tube sections. The pressure p is sampled in the middle of each tube section and the volume velocity u is sampled at the borders of adjacent sections[1]. A short piece of the vocal tract, consisting of two sections, is shown in Fig. 1. The lengths and areas of the two sections are denoted as l_{i-1}, l_i and A_{i-1}, A_i, respectively. We discretize Eq. (3) by the multiplication with an infinitesimal distance dx and the integration between the positions x_{i-1} and x_i (position coordinates in the middle of the individual tube sections). When we assume a time-independent area function $A(x)$ and consider that the volume velocity satisfies $u(x,t) = v(x,t)A(x)$, we get

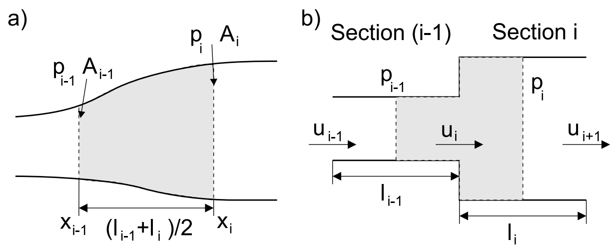

Figure 1: Spatial discretization of a piece of the vocal tract. a) continual case, b) discrete case.

$$-\int_{x_{i-1}}^{x_i} \frac{\partial p(x,t)}{\partial x}dx = \rho_0\int_{x_{i-1}}^{x_i} \frac{\partial}{\partial t}\left(\frac{u(x,t)}{A(x)}\right)dx + \frac{\rho_0}{2}\int_{x_{i-1}}^{x_i} \frac{\partial}{\partial x}\left(\frac{u^2(x,t)}{A^2(x)}\right)dx.$$

[1]We drop the prime to denote the sound pressure in the following.

Due to the spatial sampling, $u(x,t) \equiv u_i(t)$ for $x_{i-1} \leq x \leq x_i$. The partial derivatives on the right-hand side of the equation can now be transformed into total derivatives, yielding

$$p_{i-1}(t)-p_i(t) = \rho_0 \dot{u}_i(t) \int_{x_{i-1}}^{x_i} \frac{dx}{A(x)} + u_i^2(t)\frac{\rho_0}{2} \int_{x_{i-1}}^{x_i} \frac{\partial}{\partial x}\frac{dx}{A^2(x)}.$$

The integration over the piecewise constant area function $A(x)$ gives

$$p_{i-1} - p_i = \dot{u}_i(L_{i-1} + L_i) + u_i^2(W_i - W_{i-1}), \qquad (4)$$

where $L_i = \rho_0 l_i/(2A_i)$ and $W_i = \rho_0/(2A_i^2)$. Recall that this equation was derived from Euler's equation for the frictionless case. However, in reality the vocal tract has several loss mechanisms. Losses due to viscous friction can approximately be considered by means of a linear resistance R that causes a pressure drop proportional to u [3]. Other losses of energy can be found at sudden expansions or contractions of the vocal tract due to flow separation from the vocal tract walls [2]. These cases can be considered by means of a "correction coefficient" η_i in front of the second order term in Eq. (4). A sudden expansion in the vocal tract geometry in conjunction with sufficiently high fluid velocities usually produces turbulence noise that can be represented as a source of sound pressure at the appropriate place. This source of noise can be considered by an additional pressure term p_i^*. With these modifications, Eq. (4) becomes

$$p_{i-1}-p_i = \dot{u}_i(L_{i-1}+L_i)+u_i(R_{i-1}+R_i)+u_i^2\eta_i(W_i-W_{i-1})-p_i^*. \qquad (5)$$

This equation describes the relationship between pressure and volume velocity depicted by the circuit in Fig. 2. Compared to the conventional circuits, we have represented the nonlinear resistances (pressure drop proportional to u^2) by the diamond-shaped symbols for better discrimination.

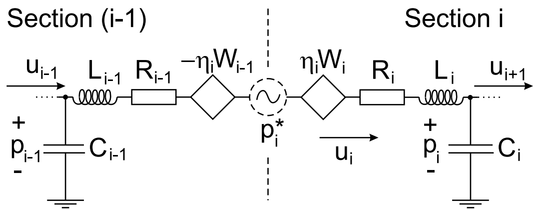

Figure 2: Lumped transmission-line representation of the vocal tract from the center of section $i-1$ to the center of section i including the nonlinear resistances.

Each of the capacitors represents the compressibility of the air in one of the two tube sections. They result from the spatial discretization of the continuity equation. The relation between pressure and volume velocity for the capacities reads

$$\dot{p}_i = \frac{1}{C_i}(u_i - u_{i+1}), \qquad (6)$$

where $C_i = l_i A_i/(\rho_0 c^2)$ and c is the speed of sound.

Temporal discretization

The equations (5) and (6) are coupled by the quantities $p_i(t)$ and $u_i(t)$ and are the foundation for the acoustic simulation in our synthesizer. The temporal discretization is implemented by means of the trapezoid rule

$$f[n] = f[n-1] + \frac{\Delta t}{2}(\dot{f}[n-1] + \dot{f}[n]) \qquad (7)$$

for $f \in \{p, u\}$. In order to linearize the second-order term in Eq. (5) in $u[n]$, we make the following approximation:

$$u^2[n] \approx -u^2[n-1] + 2u[n]u[n-1]. \qquad (8)$$

By combining the equations (5), (6), (7) and (8) we obtain a linear system of equations, whose solution is the wanted vector of the volume velocities. At the glottal end and at the mouth of the vocal tract, the corresponding boundary conditions have to be implemented (lung/glottis and radiation impedance) [3, 2].

Results and Discussion

In some implementations of transmission line models, nonlinear resistances have been used by other researchers before, but usually only with regard to vocal fold vibration or in order to model pressure losses after narrow constrictions in the vocal tract [2, 4]. However, at the other places in the vocal tract, where no energy losses are expected (either due to turbulence or flow separation), the Bernoulli effect is not considered.

In our simulation of the TLM we can choose to switch the nonlinear resistances either on or off along the entire length of the vocal tract in order to examine their influence on the spatial and temporal pressure distribution during speech production. For instance, when we simulate "breathing out" with a wide open glottis we can clearly observe the change of pressure depending on the cross-sectional area along the tube axis, just as we would expect it due to the Bernoulli effect for a stationary flow. However, during the simulation of speech with a *time varying* vocal tract geometry we could in some cases detect numeric instabilities, especially in the time intervals immediately preceding or following alveolar or labial closure. These instabilities do not occur when we switch the nonlinear resistances off. Thus, we attribute the mistakes to the approximation made in Eq. (8) and are currently seeking to improve the method.

References

[1] J. L. Flanagan, *Speech Analysis Synthesis and Perception.* Academic Press Inc., 1965.

[2] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *The Bell System Technical Journal*, **51**, no. 6, pp. 1233–1268, 1972.

[3] S. Maeda, "A digital simulation method of the vocal-tract system," *Speech Communication*, **1**, pp. 199–229, 1982.

[4] K. N. Stevens, *Acoustic Phonetics.* The MIT Press, 1998.