3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

AN ARTICULATORY SYNTHESIZER FOR THE SIMULATION OF CONSONANTS

Paul Boersma

Institute of Phonetic Sciences, University of Amsterdam,
Herengracht 338, 1016CG Amsterdam, The Netherlands

ABSTRACT

We present a model of the lungs, glottis, and vocal tract, that computes numerically the acoustic output that results from the positions, motions, and tensions of the muscles involved. The model is designed to be able to produce almost any possible speech utterance.

Keywords: *Articulatory synthesis, Vocal-tract model*

1. INTRODUCTION

In our investigations into the articulatory and perceptual features of speech sounds, we need an articulatory-acoustic model of the production of speech utterances. The input to this model should bear a close relationship to the activities of the main muscles involved, and the output should be the resulting sound. For the model to be able to produce almost any speech utterance, it must be capable of having:
- yielding and vibrating walls everywhere (not only in glottis)
- noise generation in turbulent conditions anywhere
- time-varying lengths of tract regions

In contrast with some existing models [1, 2], computations of myo-elastic and aerodynamic quantities are carried out on a representation of the *entire* speech apparatus (from diaphragm to lips and nostrils) as a sequence of straight tubes with variable lengths and widths, each of which has two walls of equal mass, tension, and damping (figure 1). The acoustic output is computed from the time-derivative of the airflow at the lips and nostrils, and from the movements of the walls.

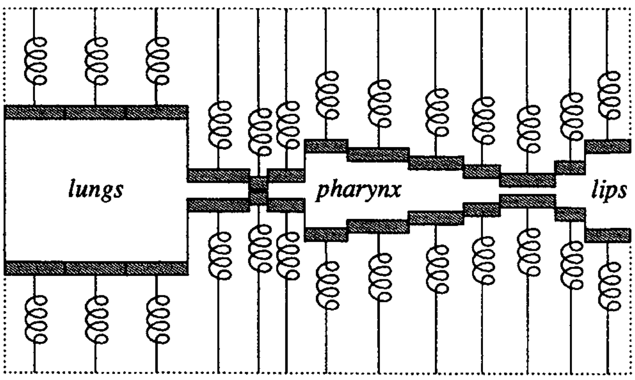

Fig. 1. The lungs, the glottis and the vocal tract are all treated in the same way.

2. MESHING

The lungs, the bronchi, the trachea, the glottis and the vocal tract are divided into straight tubes with lengths Δx and equilibrium widths w_{eq}. For the subglottal system, the lengths are constant, and varying the equilibrium widths is responsible for building up pressures behind constrictions in the glottis and in the vocal tract. The glottis consists of two tubes with constant lengths (i.e., vocal-cord thickness) and adjustable equilibrium widths and tensions. For the vocal tract, the lengths and equilibrium widths of the tubes are determined by two contours that represent the positions of the main muscles involved (figure 2). The outer contour is formed by the rear pharyngeal wall, the velum, the palate, the alveoli, the upper teeth, and the upper lips. The inner contour is formed by the hyoid bone, the tongue root, the tongue body, the tongue tip, the lower teeth, and the lower lips. Wherever the inner contour crosses the outer contour, the equilibrium width of the tract becomes negative (the walls are pressed together). Meshing of the vocal tract is performed in a way based on Mermelstein [3], see figure 2: the origin of the co-ordinate system is the centre of curvature of the velum and palate. Because the lengths of the tubes must be continuous over time, the mesh points are fixed on the outer contour, and the line pieces that represent the central cross sections of the tubes have a constant direction (horizontal lines below the origin, vertical lines to the right of the origin, and radial lines in the upper left quadrant). Because the equilibrium widths of the tubes must also be continuous over time, the absolute value of the width is taken as the distance from the mesh point to the *nearest* point on the inner contour; the sign of the width is negative if the mesh point is inside the closed inner contour. The length of a tube is the distance between two points that are each half-way between the midpoints of two adjacent mesh lines.

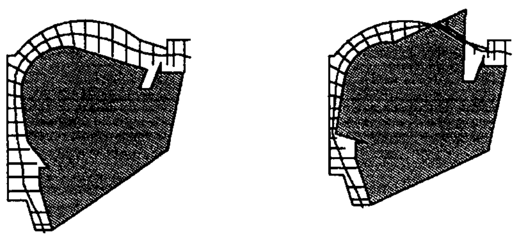

Fig. 2. Meshing of the vocal tract in a neutral position (left) and during the closure of the ejective stop [t'] (right). Two images from a film of the simulated utterance [ət'ə].

3. SPRINGS

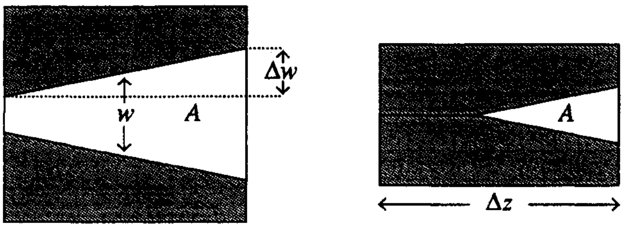

Fig. 3. The walls close upon each other like a zipper.

From the cross-sectional view of a tube in figure 3, we see that the walls of a tube are not exactly parallel, and that the average width w_{av} (always positive) is a smooth function of the relative wall position w, which can be negative:

$$w_{av} = \begin{cases} w + w_{min} & (w \ge \Delta w) \\ \dfrac{(\Delta w + w)^2}{4\Delta w} + w_{min} & (-\Delta w \le w \le \Delta w) \\ w_{min} & (w \le -\Delta w) \end{cases} \quad (1)$$

where $w_{min} = 0.01$ mm allows for a small leakage. The cross-sectional area A thus equals $w_{av}\,\Delta z$, where Δz is the third dimension of the tube (e.g., in the glottis it is the length of the vocal cords).

The equation of motion of a tube wall is

$$\tfrac{1}{2} m \frac{d^2 w}{dt^2} = F_{op} + F_{co} + F_{cl} - \tfrac{1}{2}\left(B_{op} + B_{cl}\right)\frac{dw}{dt} + P\,\Delta x\,\Delta z \quad (2)$$

where m is the mass of the wall and P is the mean excess pressure in the tube. The spring force F_{op} is given by

$$F_{op} = \tfrac{1}{2} k^{(1)}\left(w_{eq} - w\right) + \tfrac{1}{8} k^{(3)}\left(w_{eq} - w\right)^3 \quad (3)$$

where $k^{(1)}$ and $k^{(3)}$ are the linear and cubic spring constants of the wall. A coupling force F_{co} connects the walls of the mth tube with those of the $(m-1)$th and $(m+1)$th tube by a spring:

$$F_{co} = k^{(1)}_{m-1,m}\left(w_{m-1} - w_m\right) + k^{(3)}_{m-1,m}\left(w_{m-1} - w_m\right)^3 + \dots \quad (4)$$

The collision force F_{cl} due to the contact of both walls is

$$F_{cl} = \begin{cases} \dfrac{s^{(1)}(\Delta w - w)^2}{8\Delta w} + \dfrac{s^{(3)}(\Delta w - w)^4}{64\Delta w} & (-\Delta w \le w \le \Delta w) \\ -\tfrac{1}{2} s^{(1)} w - \tfrac{1}{8} s^{(3)} w\left(w^2 + \Delta w^2\right) & (w \le -\Delta w) \end{cases} \quad (5)$$

where $s^{(1)}$ and $s^{(3)}$ are the linear and cubic tissue stiffnesses. The damping of the spring B_{op} and the damping due to the contact of both walls B_{cl} are expressed relative to the critical dampings as

$$B_{op} = B_{op,rel}\, 2\sqrt{k_{eff}\, m} \quad ; \quad B_{cl} = B_{cl,rel}\, 2\sqrt{s_{eff}\, m_{eff}} \quad (6)$$

where the dynamic spring "constant" is

$$k_{eff} = -\frac{\partial F_{op}}{\partial (w/2)} = k^{(1)} + \tfrac{3}{4} k^{(3)}\left(w_{eq} - w\right)^2 \quad (7)$$

and the effective mass and stiffness are

$$m_{eff} = \begin{cases} m \dfrac{\Delta w - w}{2\Delta w} & (-\Delta w \le w \le \Delta w) \\ m & (w \le -\Delta w) \end{cases}$$

$$s_{eff} = \begin{cases} \dfrac{\Delta w - w}{2\Delta w}\left(s^{(1)} + \tfrac{1}{4} s^{(3)} (\Delta w - w)^2\right) & (-\Delta w \le w \le \Delta w) \\ s^{(1)} + \tfrac{1}{4} s^{(3)}\left(3w^2 + \Delta w^2\right) & (w \le -\Delta w) \end{cases} \quad (8)$$

We take $B_{op,rel} = 0.2$ in the glottis, 0.8 everywhere else, and $B_{cl,rel} = 1$ (critical damping) everywhere. These values and those for m, $k^{(1)}$, $k^{(3)}$, $s^{(1)}$, and $s^{(3)}$ are thought to represent an average adult female speaker.

In the digital simulation, time is spliced into pieces of a fixed duration Δt, which must be smaller than the time needed for sound to travel the length of the shortest tube. The difference equations that derive the wall positions w of the mth tube at a time $(n+1)\Delta t$ from the positions at $n\Delta t$, are

$$\dot{w}_m^{n+\frac{1}{2}} = \frac{\dot{w}_m^{n-\frac{1}{2}} + \dfrac{2\left(F^n_{op,m} + F^n_{cl,m} + P^n_m \Delta x^n_m \Delta z_m\right)}{m^n_m}\Delta t}{1 + \dfrac{B^n_{op,m} + B^n_{cl,m}}{m^n_m}\Delta t}$$

$$w_m^{n+1} = w_m^n + \dot{w}_m^{n+\frac{1}{2}}\Delta t \quad (9)$$

4. AERODYNAMIC EQUATIONS

The conservation of the mass in a tract is expressed as the pseudo-1-dimensional integral equation

$$\frac{\partial}{\partial t}\int_{x_1}^{x_2} \rho A\, dx = (\rho v A)_{x=x_1} - (\rho v A)_{x=x_2} \quad (10)$$

where x is the place co-ordinate along the tract, t is the time, $\rho(x,t)$ is the mass density of the fluid, $A(x,t)$ is the cross-sectional area of the tube, $v(x,t)$ is the mean particle velocity, and $x_1(t)$ and $x_2(t)$ are two positions in the tract.

The equation of motion for a particle at position x is

$$\rho\frac{dv}{dt} = \rho\frac{\partial v}{\partial t} + \rho v\frac{\partial v}{\partial x} = -\frac{\partial P}{\partial x} + Rv \quad (11)$$

where $R(x,t)$ is the resistance (Ns/m^4) due to viscous drag at the walls, which is the sum of the resistance due to the viscous boundary layer, which is dominant for large tube widths [4, 1], and the resistance of Hagen-Poiseuille (fully developed laminar) flow, which is a good approximation for small tube widths [5, 2]:

$$R = \frac{0.3\,\mathrm{Ns/m^3}}{w_{av}} + \frac{12 \cdot 1.86 \cdot 10^{-5}\,\mathrm{Ns/m^2}}{\left(w^2\right)_{av}} \tag{12}$$

In the lungs, we divide the widths in (12) by a factor 1000, in order to account for the multiple branching.

For small particle velocities, the fluid is nearly incompressible, so the equation of motion becomes

$$\frac{\partial(\rho v)}{\partial t} = -\frac{\partial\left(P + \frac{1}{2}\rho_0 v^2\right)}{\partial x} + Rv \tag{13}$$

The third aerodynamic equation, which couples the continuity and motion equations, is the adiabatic equation of state

$$P = (\rho - \rho_0)c^2 \tag{14}$$

where $\rho_0 = 1.14$ kg/m^3 is the mean atmospheric density, and $c = 353$ m/s is the velocity of sound in vivo.

We can define the mass flow J (kg/s), the kinetic (Bernoulli) pressure K, and the continuous pressure Q (N/m^2), by

$$J \equiv \rho v A \quad ; \quad K \approx \frac{(\rho v)^2}{2\rho_0} \quad ; \quad Q \equiv \rho c^2 + K \tag{15}$$

For laminar flow (no turbulence), J and Q are continuous at the boundaries between tubes. More precisely, their values at the boundary between the mth and (m+1)th tube are

$$J_{m+1-} = J_{m+} \quad ; \quad Q_{m+1-} = Q_{m+} + P_{turb,m,m+1} \tag{16}$$

where subscripts $m+$ and $m-$ refer to the limit values at the right and left boundaries of the mth tube, and where $P_{turb,m,m+1}$ is the pressure gain due to turbulence in going from the mth to the (m+1)th tube. A turbulence pressure drop exists at any tube boundary where the flow is directed from the narrower into the wider tube, if the particle velocity in the narrower tube is higher than a critical velocity $v_{crit} = 10$ m/s:

$$P_{turb} = -\frac{1}{2}\rho_0 v_{narrow}\left(|v_{narrow}| - v_{crit}\right)\left(1 - \frac{A_{narrow}}{A_{wide}}\right)^2 \cdot \left(1 + 0.1\,\mathrm{N}_m^n\right) \tag{17}$$

where N_m^n is unit-power Gaussian white noise. The last factor in (17) represents noise generated from turbulence.

The momentum density p (kg/m^{2}s) and the mass line density e (kg/m) (both discontinuous at tube boundaries) are

$$p \equiv \frac{J}{A} \quad ; \quad e \equiv \frac{QA}{c^2} - \frac{J^2}{2\rho_0 c^2 A} \tag{18}$$

With these, we can eliminate the equation of state from the hydrodynamic equations and write them in integral form as

$$\frac{\partial}{\partial t}\int e\,dx = -\Delta J \quad ; \quad \int \frac{\partial p}{\partial t}dx = -\Delta Q - \int \frac{R}{\rho_0} p\,dx \tag{19}$$

These equations can be integrated numerically after rewriting them as difference equations that derive the quantities J and Q at a time $(n+1)\Delta t$ from the quantities at $n\Delta t$, for every tube boundary. We use a scheme analogous to Lax-Wendroff integration [6]. First, the Lax step averages the values for e and p inside the tubes from J and Q at the boundaries:

$$e_{m\pm}^n \equiv \frac{Q_{m\pm}^n A_m^n}{c^2} - \frac{\left(J_{m\pm}^n\right)^2}{2\rho_0 c^2 A_m^n} \quad ; \quad p_{m\pm}^n = \frac{J_{m\pm}^n}{A_m^n} \tag{20}$$

$$e_m^n = \tfrac{1}{2}\left(e_{m-}^n + e_{m+}^n\right) \quad ; \quad p_m^n = \tfrac{1}{2}\left(p_{m-}^n + p_{m+}^n\right) \tag{21}$$

From these values we can calculate the mean pressure and particle velocity in the mth tube as

$$\rho_m^n = \frac{e_m^n}{A_m^n} \quad ; \quad P_m^n = \left(\rho_m^n - \rho_0\right)c^2 \quad ; \quad v_m^n = \frac{p_m^n}{\rho_m^n} \tag{22}$$

from which we compute the new tube widths w_m^{n+1} and areas A_m^{n+1} with equations (1) through (9), the resistances R_m^n with (12), the resistance factors r_m^n with

$$r_m^n \equiv 1 + \frac{R_m^n}{\rho_0}\Delta t \tag{23}$$

and the turbulence $P_{turb,m,m+1}^n$ with (17). The next step is integrating to find the half-way values of e, p, J, and Q inside the tubes, using a first-order explicit method:

$$e_m^{n+\frac{1}{2}} = \frac{e_m^n \Delta x_m^n + \frac{1}{2}\Delta t\left(J_{m-}^n - J_{m+}^n\right)}{\Delta x_m^{n+\frac{1}{2}}}$$

$$p_m^{n+\frac{1}{2}} = \frac{p_m^n + \frac{1}{2}\Delta t\,\dfrac{Q_{m-}^n - Q_{m+}^n}{\Delta x_m^n}}{1 + \frac{1}{2}\Delta t\,\dfrac{R_m^n}{\rho_0}} \tag{24}$$

$$J_m^{n+\frac{1}{2}} = p_m^{n+\frac{1}{2}} A_m^{n+\frac{1}{2}} \quad ; \quad Q_m^{n+\frac{1}{2}} = \frac{e_m^{n+\frac{1}{2}} c^2}{A_m^{n+\frac{1}{2}}} + \frac{\left(p_m^{n+\frac{1}{2}}\right)^2}{2\rho_0}$$

The next step is integrating to find implicit equations for the new limit values of e and p at tube boundaries:

$$\frac{p_{m+}^{n+1} - p_{m+}^n}{\Delta t}\frac{1}{2}\Delta x_m^{n+\frac{1}{2}} + \frac{p_{m+1-}^{n+1} - p_{m+1-}^n}{\Delta t}\frac{1}{2}\Delta x_{m+1}^{n+\frac{1}{2}} = Q_m^{n+\frac{1}{2}} - Q_{m+1}^{n+\frac{1}{2}} + P_{turb,m,m+1}^n - \frac{R_m^n}{\rho_0}p_{m+}^{n+1}\frac{1}{2}\Delta x_m^{n+\frac{1}{2}} - \frac{R_{m+1}^n}{\rho_0}p_{m+1-}^{n+1}\frac{1}{2}\Delta x_{m+1}^{n+\frac{1}{2}} \tag{25}$$

$$\frac{\left(e_{m+}^{n+1}\frac{1}{2}\Delta x_m^{n+1} + e_{m+1-}^{n+1}\frac{1}{2}\Delta x_{m+1}^{n+1}\right) - \left(\text{same}^n\right)}{\Delta t} = J_m^{n+\frac{1}{2}} - J_{m+1}^{n+\frac{1}{2}} \tag{26}$$

Substituting (20) at the boundaries yields the left-limit values

$$J_{m+}^{n+1} = \frac{\Delta x_m^{n+\frac{1}{2}} p_{m+}^n + \Delta x_{m+1}^{n+\frac{1}{2}} p_{m+1-}^n + 2\Delta t\left(Q_m^{n+\frac{1}{2}} - Q_{m+1}^{n+\frac{1}{2}} + P_{turb,m,m+1}^n \right)}{r_m^n \frac{\Delta x_m^{n+\frac{1}{2}}}{A_m^{n+1}} + r_{m+1}^n \frac{\Delta x_{m+1}^{n+\frac{1}{2}}}{A_{m+1}^{n+1}}} \tag{27}$$

$$Q_{m+}^{n+1} = \left(A_m^{n+1}\Delta x_m^{n+1} + A_{m+1}^{n+1}\Delta x_{m+1}^{n+1}\right)^{-1} \cdot \left(c^2\left(e_{m+}^n \Delta x_m^n + e_{m+1-}^n \Delta x_{m+1}^n + 2\Delta t\left(J_m^{n+\frac{1}{2}} - J_{m+1}^{n+\frac{1}{2}} \right)\right) + \frac{\left(J_{m+}^{n+1}\right)^2 \Delta x_m^{n+1}}{2\rho_0 A_m^{n+1}} + \frac{\left(J_{m+1-}^{n+1}\right)^2 \Delta x_{m+1}^{n+1}}{2\rho_0 A_{m+1}^{n+1}} - P_{turb,m,m+1} A_{m+1}^{n+1} \Delta x_{m+1}^{n+1} \right) \tag{28}$$

from which the right-limit values are computed with (16). These equations make the integration of the hyperbolic parts of (19) second-order accurate. The integration of the dissipative part is first-order implicit, which ensures stability.

At the diaphragm (the leftmost tube boundary in figure 1), there is no flow in the x-direction, so that

$$J_{1-}^{n+1} = 0 \quad ; \quad Q_{1-}^{n+1} = \frac{c^2\left(e_{1-}^n \Delta x_1^n - 2J_1^{n+\frac{1}{2}} \Delta t \right)}{A_1^{n+1} \Delta x_1^{n+1}} \tag{29}$$

The other boundary condition applies at the lips and at the nostrils, where we have approximately [4, 5]

$$\frac{\partial Q}{\partial t} - \frac{\partial\left(\frac{cJ}{A}\right)}{\partial t} + \frac{\left(Q - \rho_0 c^2\right)c}{a} = 0 \tag{30}$$

where a is the radius of the opening at the lips or nostrils. The analog of equation (25) at the boundary tube M is

$$\frac{p_{M+}^{n+1} - p_{M+}^n}{\Delta t} \frac{1}{2} \Delta x_M^{n+\frac{1}{2}} = Q_M^{n+\frac{1}{2}} - Q_{M+}^{n+\frac{1}{2}} - \frac{R_M^n}{\rho_0} p_{M+}^{n+1} \frac{1}{2} \Delta x_M^{n+\frac{1}{2}} \tag{31}$$

and the difference equation for (30) is

$$\frac{Q_{M+}^{n+1} - Q_{M+}^n}{\Delta t} - \frac{c p_{M+}^{n+1} - c p_{M+}^n}{\Delta t} + \frac{c}{a}\left(Q_{M+}^{n+\frac{1}{2}} - \rho_0 c^2 \right) = 0 \tag{32}$$

If we approximate $Q_{M+}^{n+\frac{1}{2}}$ in (31) and (32) as the average of Q_{M+}^{n+1} and Q_{M+}^n, we can solve for p_{M+}^{n+1} (and thus for J_{M+}^{n+1}) and for Q_{M+}^{n+1}.

At the three-way boundary of the velopharyngeal port we have a formula for Q that is analogous to (28). If the oral, nasal, and pharyngeal tubes adjacent to this boundary (denoted by subscripts or, na, and ph) all have equal lengths Δx, the formula for J becomes

$$\left(\frac{r_{or}^n}{A_{or}^{n+1}} + \frac{1}{\frac{A_{ph}^{n+1}}{r_{ph}^n} + \frac{A_{na}^{n+1}}{r_{na}^n}} \right) J_{or-}^{n+1} = \left(\frac{1}{A_{mo}^n} + \frac{1}{A_{ph}^n + A_{na}^n} \right) J_{or}^n + {} - \frac{2\Delta t}{\Delta x^{n+\frac{1}{2}}} \left(Q_{or}^{n+\frac{1}{2}} - \frac{A_{ph}^{n+\frac{1}{2}} Q_{ph}^{n+\frac{1}{2}} + A_{na}^{n+\frac{1}{2}} Q_{na}^{n+\frac{1}{2}}}{A_{ph}^{n+\frac{1}{2}} + A_{na}^{n+\frac{1}{2}}} \right) \tag{33}$$

$$J_{na-}^{n+1} = \text{(same, with } or \text{ and } na \text{ exchanged)}$$

$$J_{ph+}^{n+1} = J_{or-}^{n+1} + J_{na-}^{n+1}$$

5. RESULTS

Our model of the speech apparatus can simulate breathing, vowels (figure 4), plosives (figure 5), nasals, ejectives, implosives, clicks, trills (figure 4), and some fricatives.

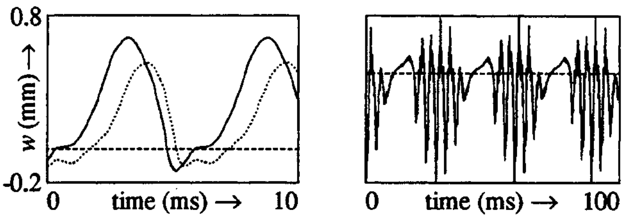

Fig. 4. To the left, the movements of the lower (drawn curve) and upper (stippled curve) parts of the vocal cords during the phonation of the vowel [a], with an equilibrium width of 0 mm. To the right, the sound from a passive vibration of the tongue tip, producing the voiced alveolar trill [r].

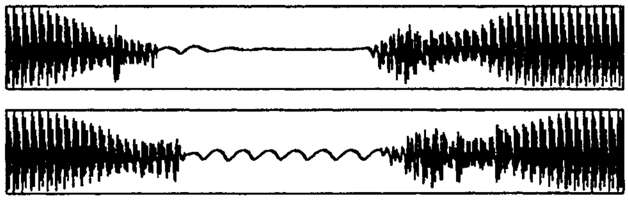

Fig. 5. 0.4 seconds from the acoustic outputs of two [apa]- and [aba]-like utterances whose articulations differ only in supraglottal wall tension. Glottal parameters are the same and constant. Even the timing of the lip movements is the same.

REFERENCES

[1]: *Ishizaka, K.; Flanagan, J.L.:* Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords. Bell System Technical Journal 51, pp. 1233-1268, 1972

[2]: *Maeda, S.:* A Digital Simulation Method of the Vocal-tract System. Speech Communication 1, pp. 199-229, 1982

[3]: *Mermelstein, P.:* Articulatory model for the study of speech production. J. Acoust. Soc. Am. 53, pp. 1070-1082, 1973

[4]: *Morse, P.M.; Ingard, K.U.:* Theoretical Acoustics. McGraw-Hill, New York 1968

[5]: *Flanagan, J.L.:* Speech Analysis Synthesis and Perception. Second, Expanded Edition. Springer, Berlin 1972

[6]: *LeVeque, R.J.:* Numerical Methods for Conservation Laws. Birkhäuser, Basle 1992