Chapter 5

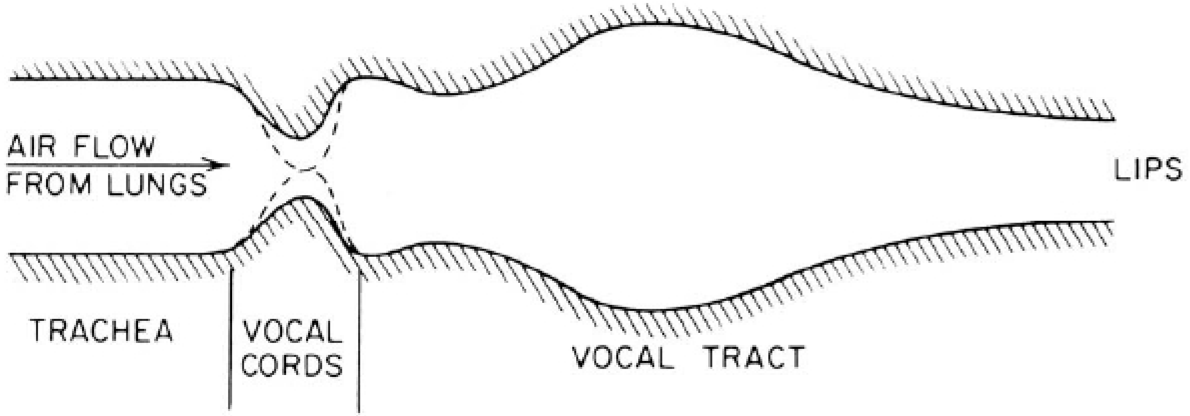
Sound Propagation in the Human Vocal Tract

声道中的声音传播

Basics

- can use basic physics to formulate air flow equations for vocal tract
- need to make simplifying assumptions about vocal tract shape and energy losses to solve air flow equations

Sound in the Vocal Tract



- Issues in creating a detailed physical model
  - time variation of the vocal tract shape (we will look mainly at fixed shapes)
  - Losses due to heat conduction and friction in the walls (we will first assume no loss, then a simple model of loss)
  - softness of vocal tract walls (leads to sound absorption issues)
  - radiation of sound at lips (need to model how radiation occurs)
  - nasal coupling (complicates the tube models as it leads to multi-tube solutions)
  - excitation of sound in the vocal tract (need to worry about vocal source coupling to vocal tract as well as source-system interactions)

Vocal Tract Transfer Function

Schematic Vocal Tract

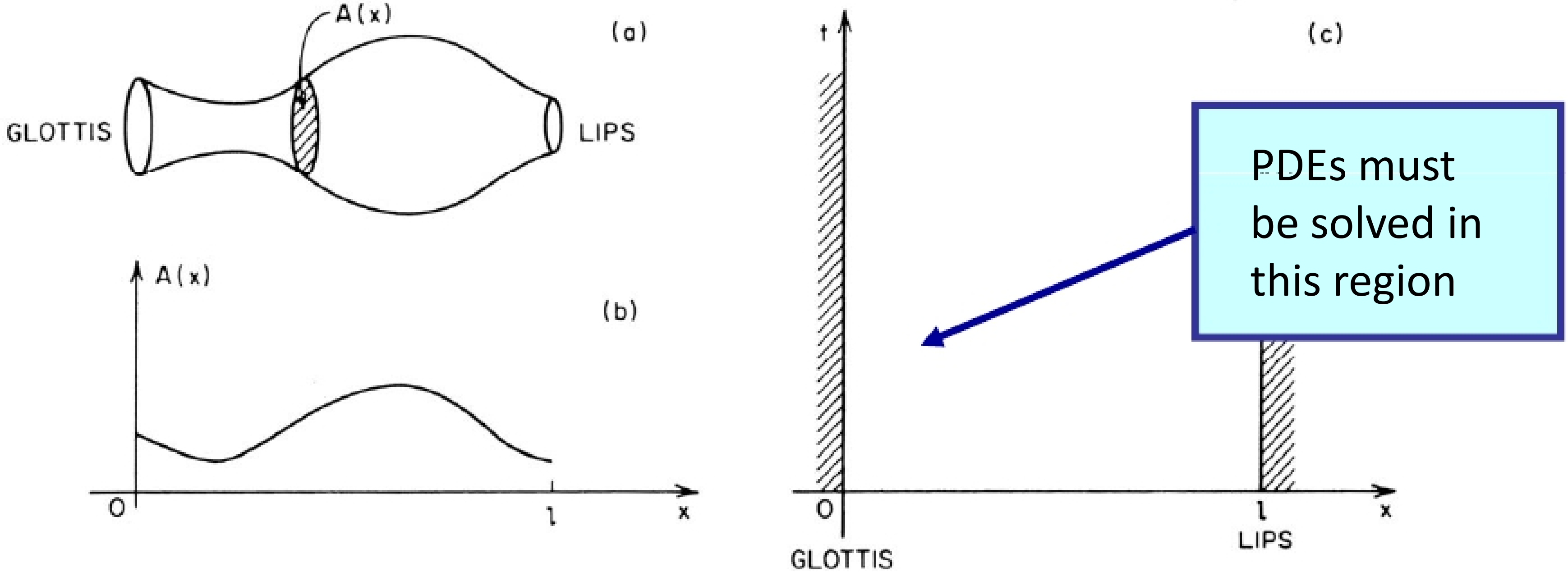


Fig. 3.13 (a) Schematic vocal tract; (b) corresponding area function; (c) $x-t$ plane for solution of wave equation.

- simplified vocal tract => non-uniform tube with time varying cross section area
- plane wave propagation along the axis of the tube (this assumption valid for frequencies below about 4000 Hz)
- no losses at walls

Sound Wave Propagation

- using the laws of conservation of mass, momentum and energy (质量、动量、能量守恒定律), it can be shown that sound wave propagation in a tube satisfies the equations:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial (u/A)}{\partial t}$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2}\frac{\partial (pA)}{\partial t} + \frac{\partial A}{\partial t}$$

- Where
  - $p = p(x,t)$ = variation in sound pressure in the tube at position x and time t
  - $u = u(x,t)$ = variation in volume velocity flow at position x and time t
  - $\rho =$ the density of air in the tube
  - $c =$ the velocity of sound
  - $A = A(x,t) =$ the 'area function' of the tube,

    i.e., the cross-sectional area normal to the axis of the tube(与声管轴方向正交的截面积), as a function of the distance along the tube and as a function of time

Solutions to Wave Equation

- no closed form solutions exist for the propagation equations
  - need boundary conditions, namely $u(0,t)$ (the volume velocity flow at the glottis), and $p(l,t)$, (the sound pressure at the lips) to solve the equations numerically (by a process of iteration)
  - need complete specification of $A(x,t)$, the vocal tract area function; for simplification purposes we will assume that there is no time variability in $A(x,t)$ => the term related to the partial time derivative of A becomes 0
  - even with these simplifying assumptions, numerical solutions are very hard to compute

Consider simple cases and extrapolate results to more complicated cases

Uniform Lossless Tube

- Assume uniform lossless tube => $A(x,t)=A$ (shape consistent with /UH/ vowel)

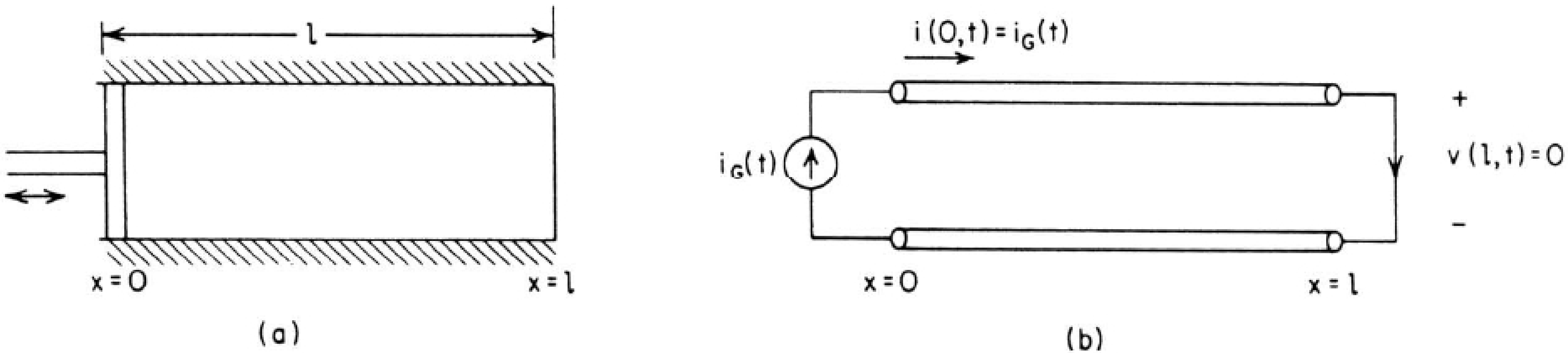


Fig. 3.14 (a) Uniform lossless tube with ideal terminations; (b) corresponding electrical transmission line analogy.

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A}\frac{\partial u}{\partial t} \qquad -\frac{\partial v}{\partial x} = L\frac{\partial i}{\partial t}$$

$$-\frac{\partial u}{\partial x} = \frac{A}{\rho c^2}\frac{\partial p}{\partial t} \qquad -\frac{\partial i}{\partial x} = C\frac{\partial v}{\partial t}$$

Acoustic-Electrical Analogs

Acoustic	Electrical
p = pressure	*v* = voltage
u = volume velocity	*i* = current
ρ/A= acoustic inductance	*L* = inductance 电感
$A/(\rho c^2)$= acoustic capacitance	*C* = capacitance 电容
uniform lossless acoustic tube $u(0,t) = U_G(\Omega)e^{j\Omega t}$ $p(\ell,t) = 0$	lossless transmission line terminated in a short circuit, $v(l,t) = 0$ at one end, excited by a current source $i(0,t) = i_G(t)$ at the other end

Traveling Wave Solution

- solve for $u(x,t)$ and $p(x,t)$

$$u(x,t) = U_G(\Omega)e^{j\Omega t}\left[\frac{e^{j\Omega(2\ell - x)/c} + e^{j\Omega x/c}}{1 + e^{2j\Omega\ell/c}}\right]$$

$$p(x,t) = \frac{\rho c}{A}U_G(\Omega)e^{j\Omega t}\left[\frac{e^{j\Omega(2\ell - x)/c} - e^{j\Omega x/c}}{1 + e^{2j\Omega\ell/c}}\right]$$

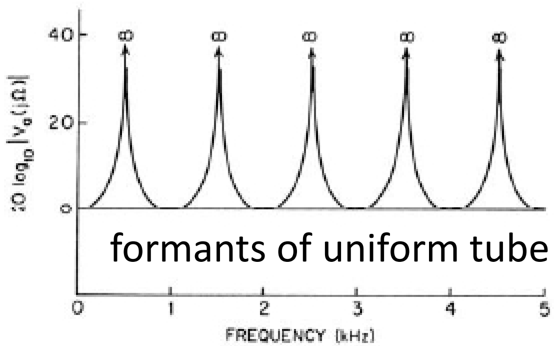
- look at solution for $u(\ell,t)$

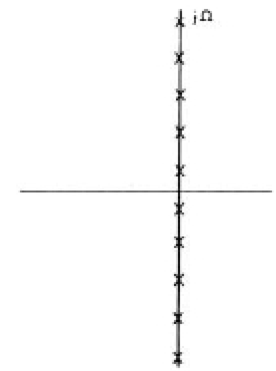
$$u(\ell,t) = U_G(\Omega)e^{j\Omega t}\left[\frac{2e^{j\Omega\ell/c}}{1 + e^{2j\Omega\ell/c}}\right] = U(\ell,\Omega)e^{j\Omega t}$$

$$= \frac{1}{\cos(\Omega\ell/c)}U_G(\Omega)e^{j\Omega t}$$

Overall Transfer Function

- consider the volume velocity at the lips (*x=l) as a* function of the source (at the glottis)





$$\frac{U(\ell,\Omega)}{U_G(\Omega)} = V_a(\Omega) = \frac{1}{\cos(\Omega\ell / c)}$$

↑

Frequency response of uniform tube in terms of volume velocities

$\Omega = 2\pi f$; $c = 35{,}000$ cm/sec; ℓ=17.5 cm

$$\frac{\Omega\ell}{c} = \frac{2\pi f \ell}{c}$$

$$\cos\left(\frac{2\pi f \ell}{c}\right) = 0 \text{ when } \frac{2\pi f \ell}{c} = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$$

i.e., when $f_n = \frac{c}{4\ell} \cdot (2n+1),\ n = 0,1,2,\ldots$

$f_0 = 500$Hz; $f_1 = 1500$ Hz; $f_2 = 2500$ Hz,...

Effects of Losses in VT

- several types of losses to be considered
  - viscous friction 粘性摩擦 at the walls of the tube
  - heat conduction 热传导 through the walls of the tube
  - vibration 振动 of the tube walls
- loss will change the frequency response of the tube
- consider first wall vibrations
  - assume walls are elastic => cross-sectional area of the tube will change with pressure in the tube
  - assume walls are 'locally' reacting => *A(x,t) ~ p(x,t)*
  - assume pressure variations are very small

$$A(x,t) = A_0(x,t) + \delta A(x,t)$$

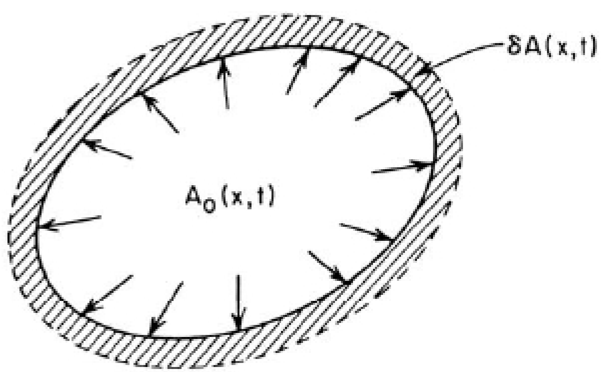


Fig. 3.16 Illustration of the effects of wall vibration.

Effects of Wall Vibration

- there is a differential equation relationship between area perturbation $\delta A(x,t)$ and the pressure variation, $p(x,t)$ of the form

$$m_W \frac{d^2(\delta A)}{dt^2} + b_W \frac{d(\delta A)}{dt} + k_W(\delta A) = p(x,t) \quad \text{where}$$

$$m_W(x) = \text{mass/unit length of the vocal tract wall}$$

$$b_W(x) = \text{damping/unit length of the vocal tract wall}$$

$$k_W(x) = \text{stiffness/unit length of the vocal tract wall}$$

- neglecting second order terms in u/A and pA, the basic wave equations become

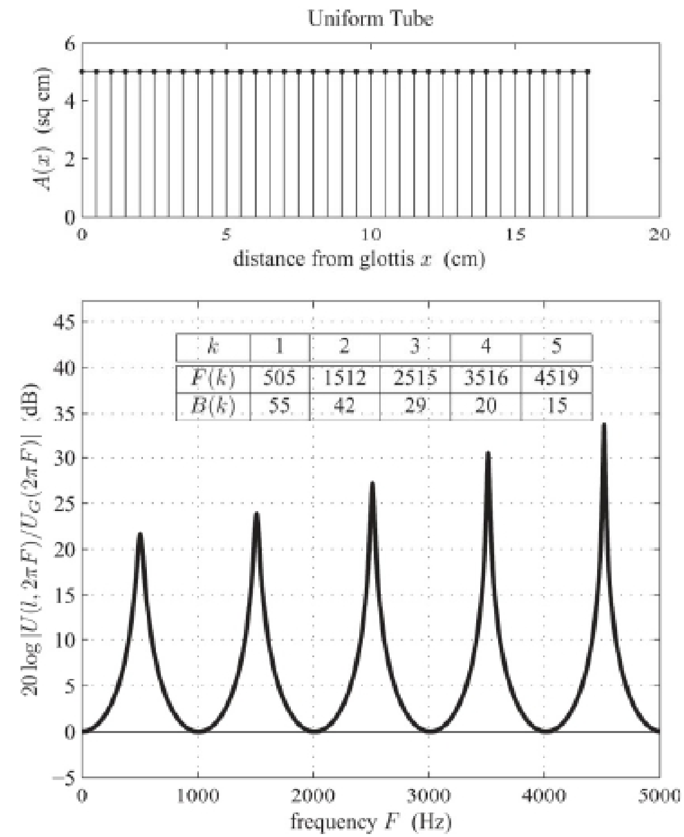
$$-\frac{\partial p}{\partial x} = \rho \frac{\partial (u/A_0)}{\partial t}$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2}\frac{\partial (pA_0)}{\partial t} + \frac{\partial A_0}{\partial t} + \frac{\partial (\delta A)}{\partial t}$$

Effects of Wall Vibration on FR

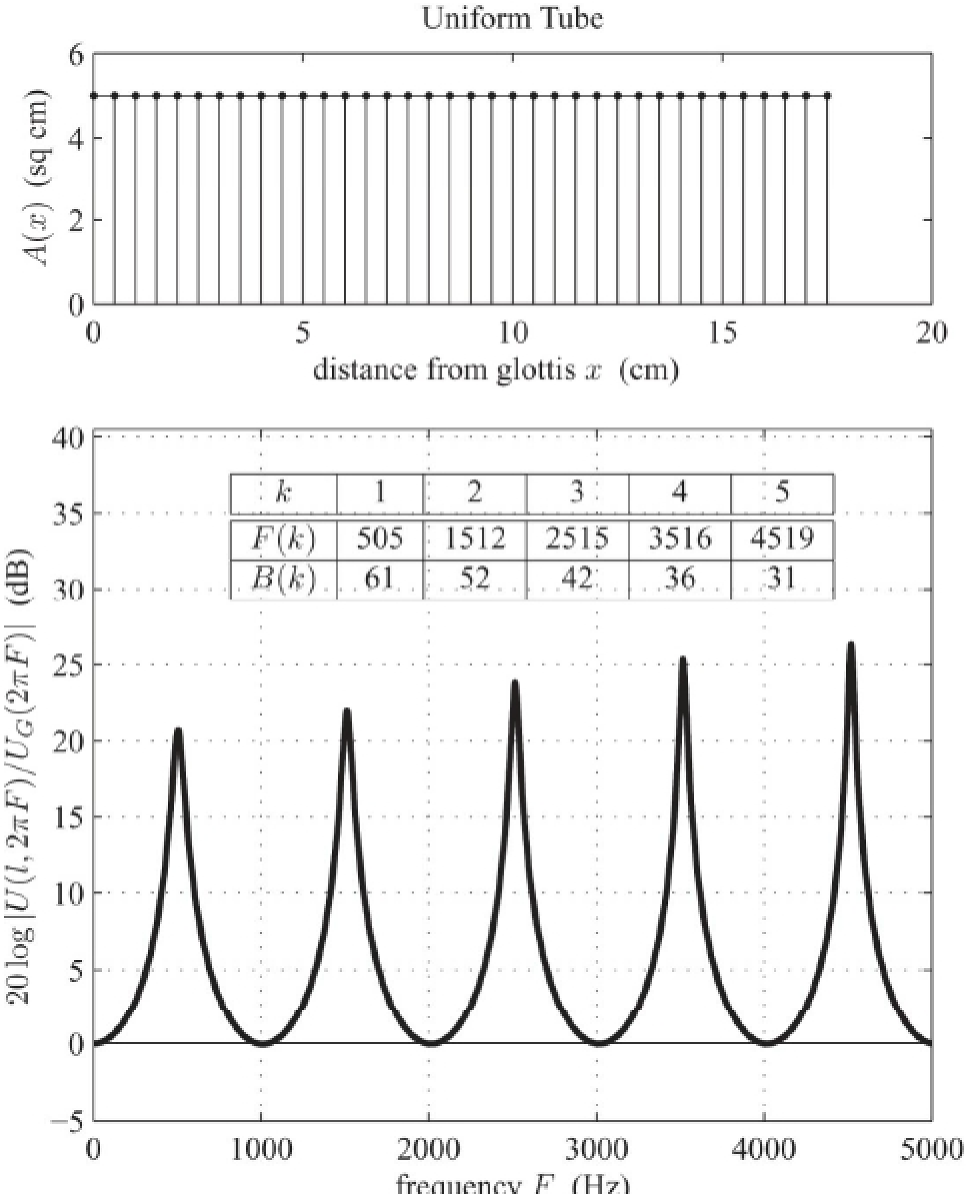
- using estimates for m_W, b_W, and k_W from measurements on body tissue, and with boundary condition at lips of *p(l,t)=0,* we get:

$$V_a(j\Omega) = \frac{U(\ell,\Omega)}{U_G(\Omega)}$$



- complex poles with non-zero bandwidths
- slightly higher frequencies for resonances
- most effect at lower frequencies

Friction and Thermal Conduction Losses



k	1	2	3	4	5
F(k)	505	1512	2515	3516	4519
B(k)	61	52	42	36	31

- Main effect of friction and thermal conduction losses is that the formant bandwidths increase
  - since friction and thermal losses increase with $\Omega^{1/2}$, the higher frequency resonances experience a greater broadening than the lower resonances
  - the effects of friction and thermal loss are small compared to the effects of wall vibration for frequencies below 3-4 kHz

Effects of at Radiation Lips

- we have assumed *p*(*l*,*t*)=*0* at the lips (the acoustical analog of a short circuit) => no pressure changes at the lips no matter how much the volume velocity changes at the lips
- in reality, vocal tract tube terminates with open lips, and sometimes open nostrils (for nasal consonants)
- this leads to two models for sound radiation at the lips

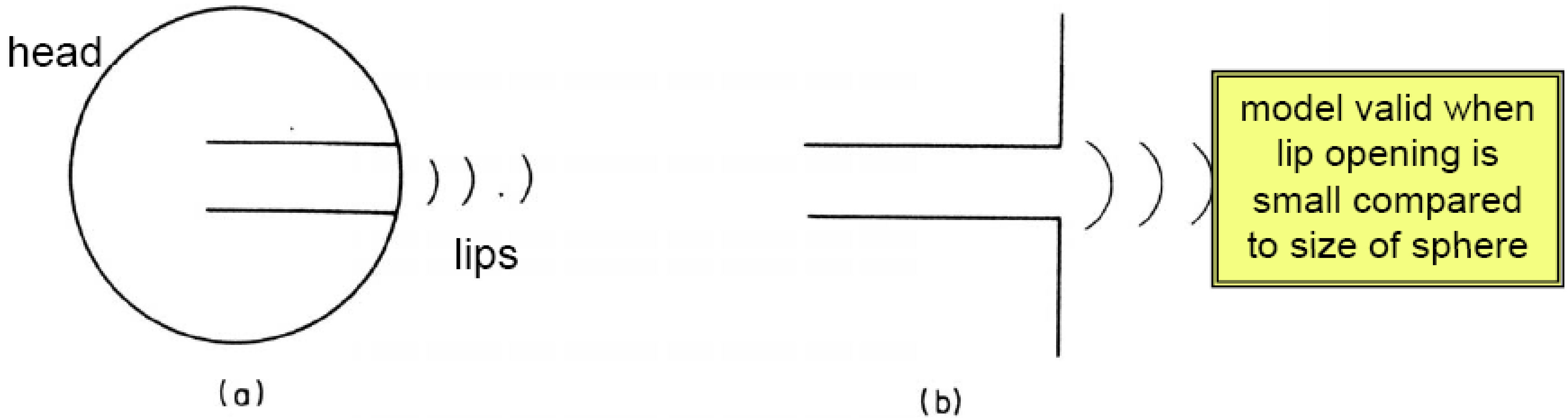


Fig. 3.19 (a) Radiation from a spherical baffle; (b) radiation from an infinite plane baffle.

Radiation at Lips

- using the infinite plane baffle model for radiation at the lips, can replace the boundary condition for a complex sinusoid input with the following:

$P(\ell,\Omega) = Z_L(\Omega)U(\ell,\Omega)$ where

$$Z_L(\Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r}$$ -- 'radiation impedance' or 'radiation load' at lips

- this 'radiation load' is the equivalent of a parallel connection of a radiation resistance, R_r , and a radiation inductance, L_r. Suitable values for these components are:

$R_r = \frac{128}{9\pi^2}$, $L_r = \frac{8a}{3\pi c}$, where a is the radius of the opening and c is the velocity of sound

Behavior of Radiation Load

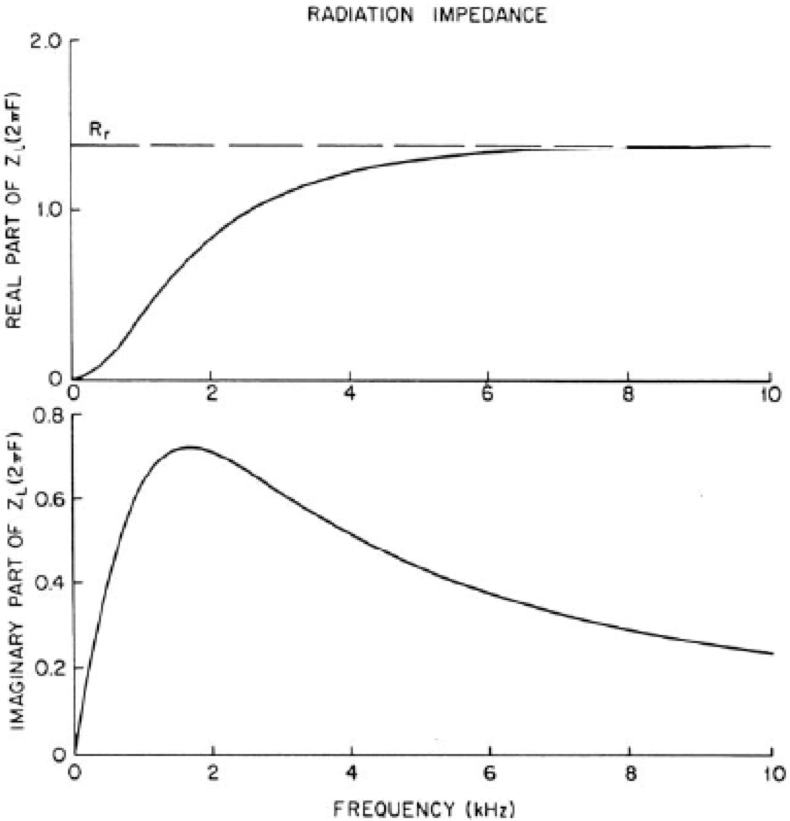


Fig. 3.20 Real and imaginary parts of the radiation impedance.

radiation losses most significant at higher frequencies

$$Z_L(\Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r}$$

- at low frequencies, $Z_L(\Omega) \approx 0$ (short circuit termination) $\Rightarrow$ old solution
- at mid-range frequencies, $Z_L(\Omega) \approx j\Omega L_r$ (inductive load) $\Rightarrow R_r \gg \Omega L_r$
- at higher frequencies, $Z_L(\Omega) \approx R_r$ (resistive load) $\Rightarrow \Omega L_r \gg R_r$

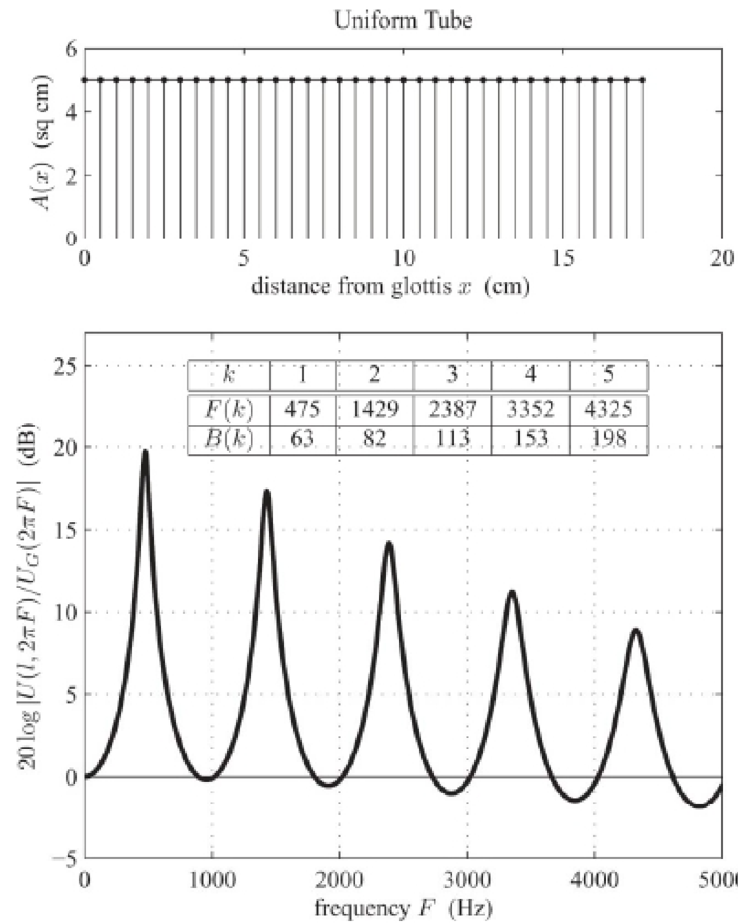
Overall Transfer Function

- for the case of a uniform, time invariant tube with yielding walls, friction and thermal losses, and radiation loss of an infinite plane baffle, can solve the wave equations for the transfer function:

$$V_a(j\Omega) = \frac{U(\ell,\Omega)}{U_G(\Omega)}$$

- assuming input at glottis of form:

$$U(0,t) = U_G(\Omega)e^{j\Omega t}$$

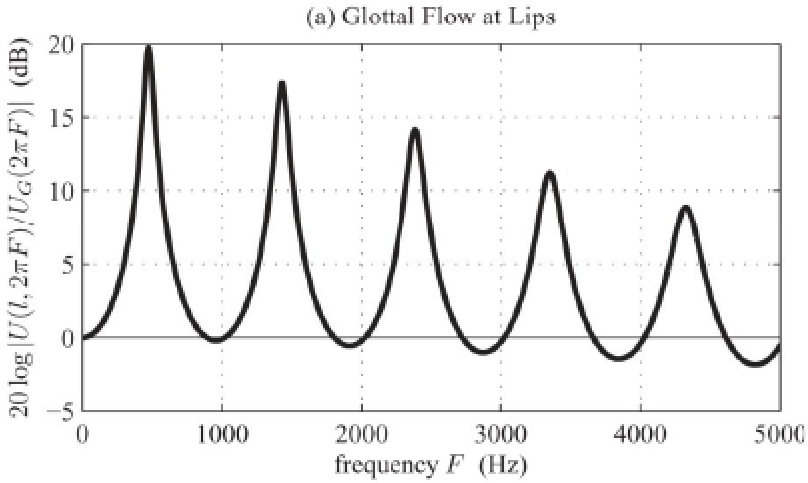


Higher bandwidths, lower resonance frequencies

- first resonance is primarily determined by wall loss
- higher resonance bandwidths are primarily determined by radiation losses

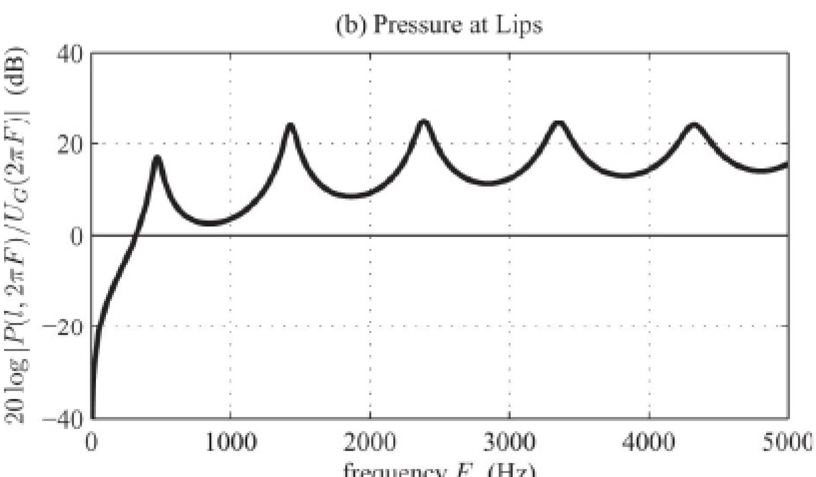
Vocal Tract Transfer Function

- look at transfer function of pressure at the lips and volume velocity at the glottis, which is of the form:



$$H_a(\Omega) = \frac{P(\ell,\Omega)}{U_G(\Omega)} = \frac{P(\ell,\Omega)}{U(\ell,\Omega)} \cdot \frac{U(\ell,\Omega)}{U_G(\Omega)}$$

$$= Z_L(\Omega) \cdot V_a(\Omega)$$



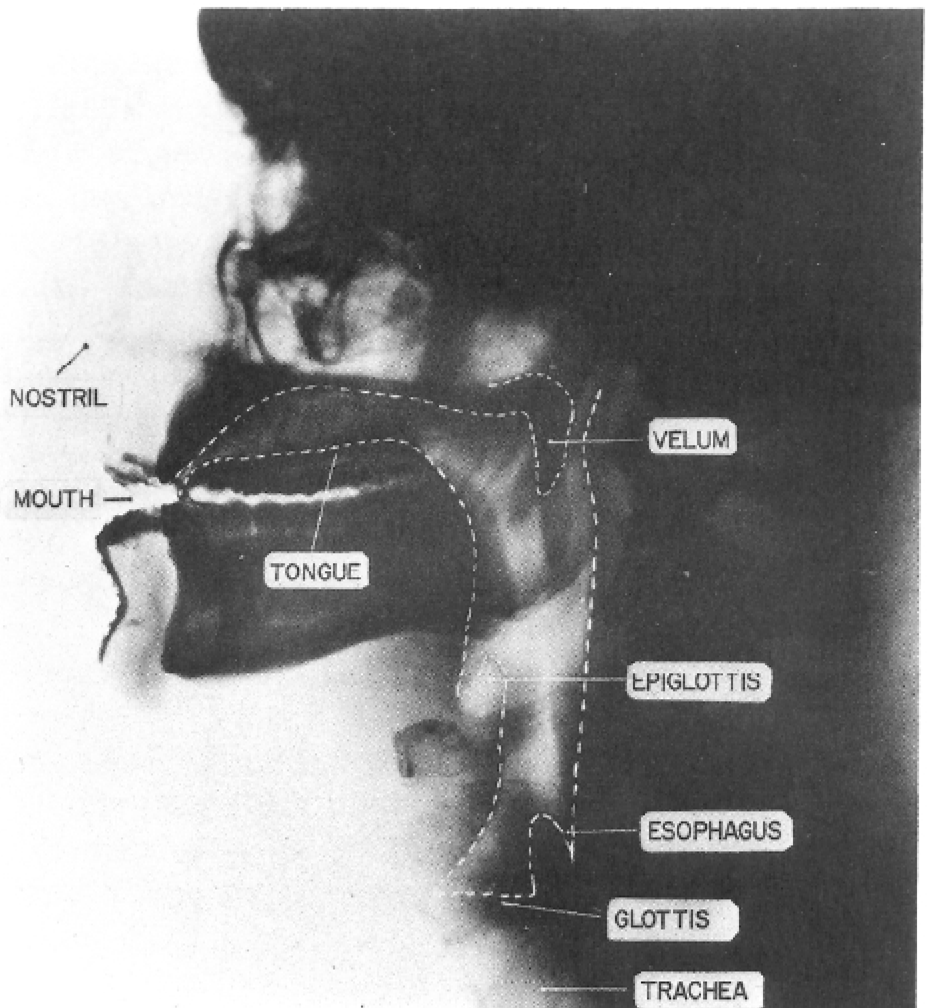
Notice:

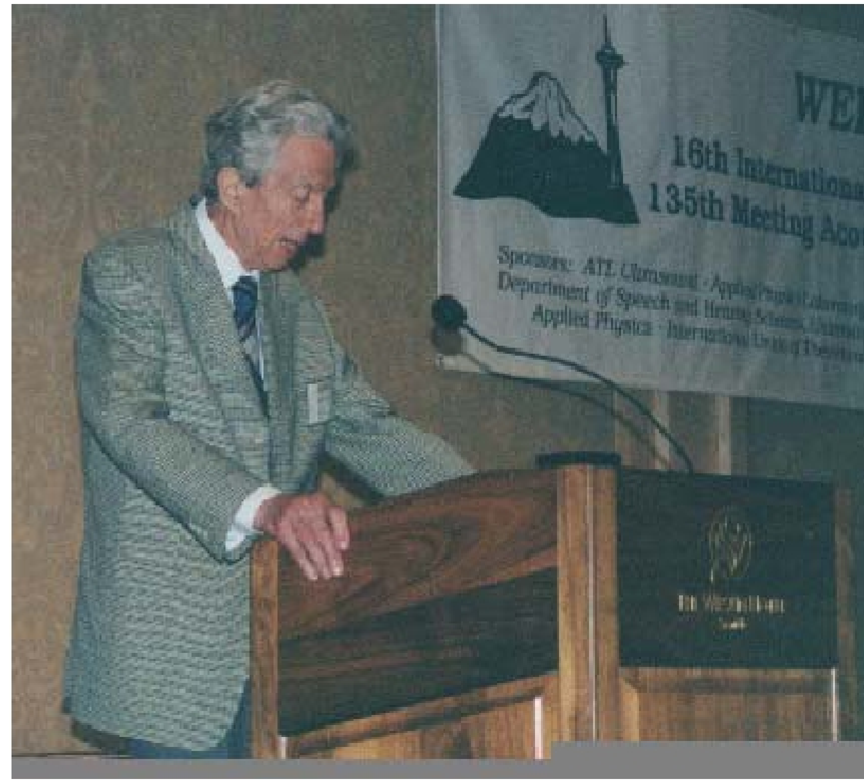
- zero at $\Omega=0$
- high frequency emphasis (compare with previous chart)

Vocal Tract Transfer Functions for Vowels

- using the frequency domain equations, can compute the frequency response functions for a set of area functions of the vocal tract for various vowel sounds, using all the loss mechanisms, assuming:
  - $A(x), 0 \leq x \leq l$ (glottis-to-lips) measured and known
  - steady state sounds ($dA/dt=0$)
  - measure $U(l,\Omega)/U_G(\Omega)$ for the vowels /AA/ /EH/ /IY/ /UW/

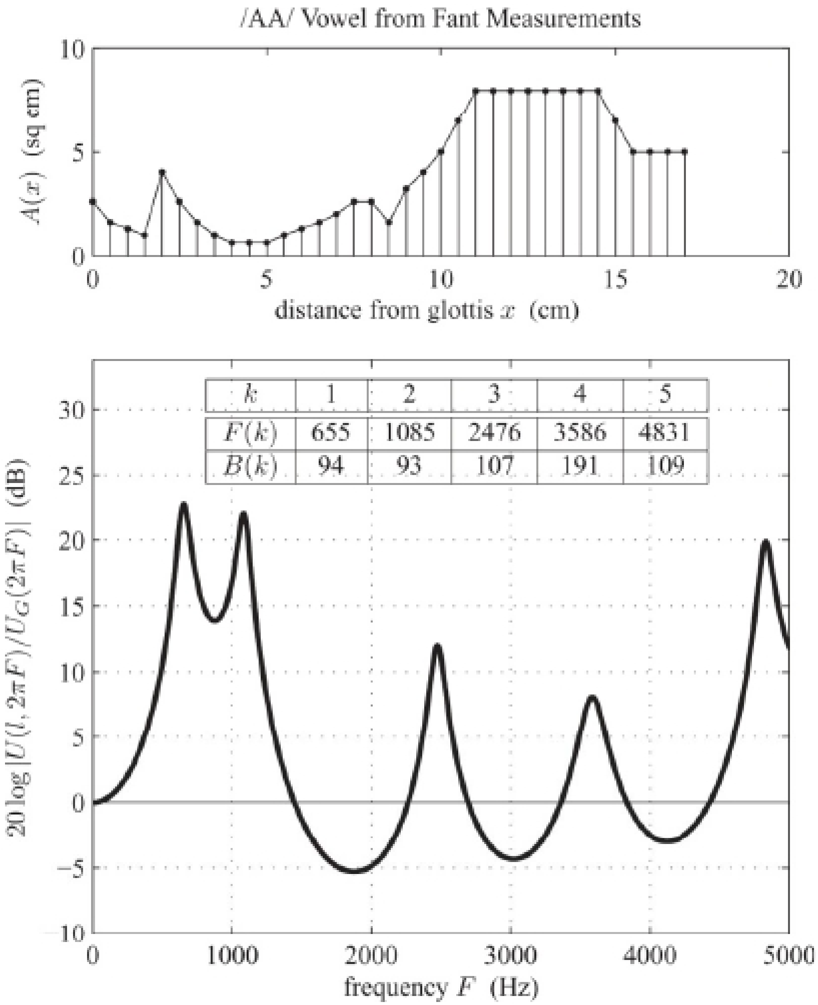
Area Function from X-Ray Photographs

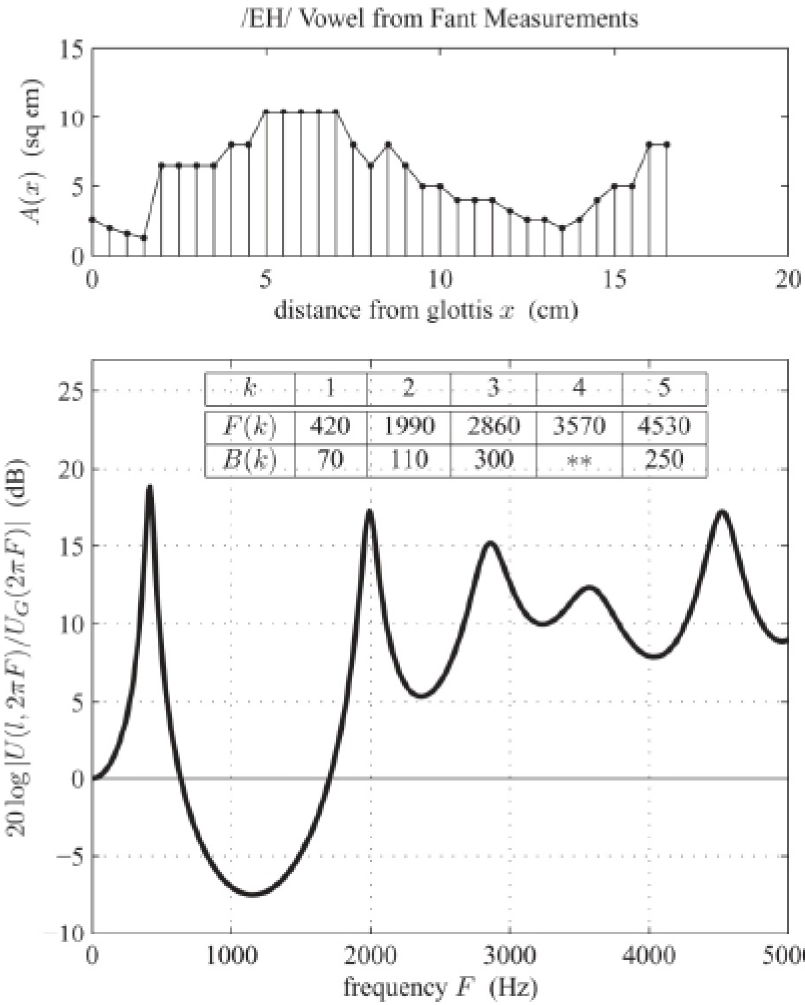




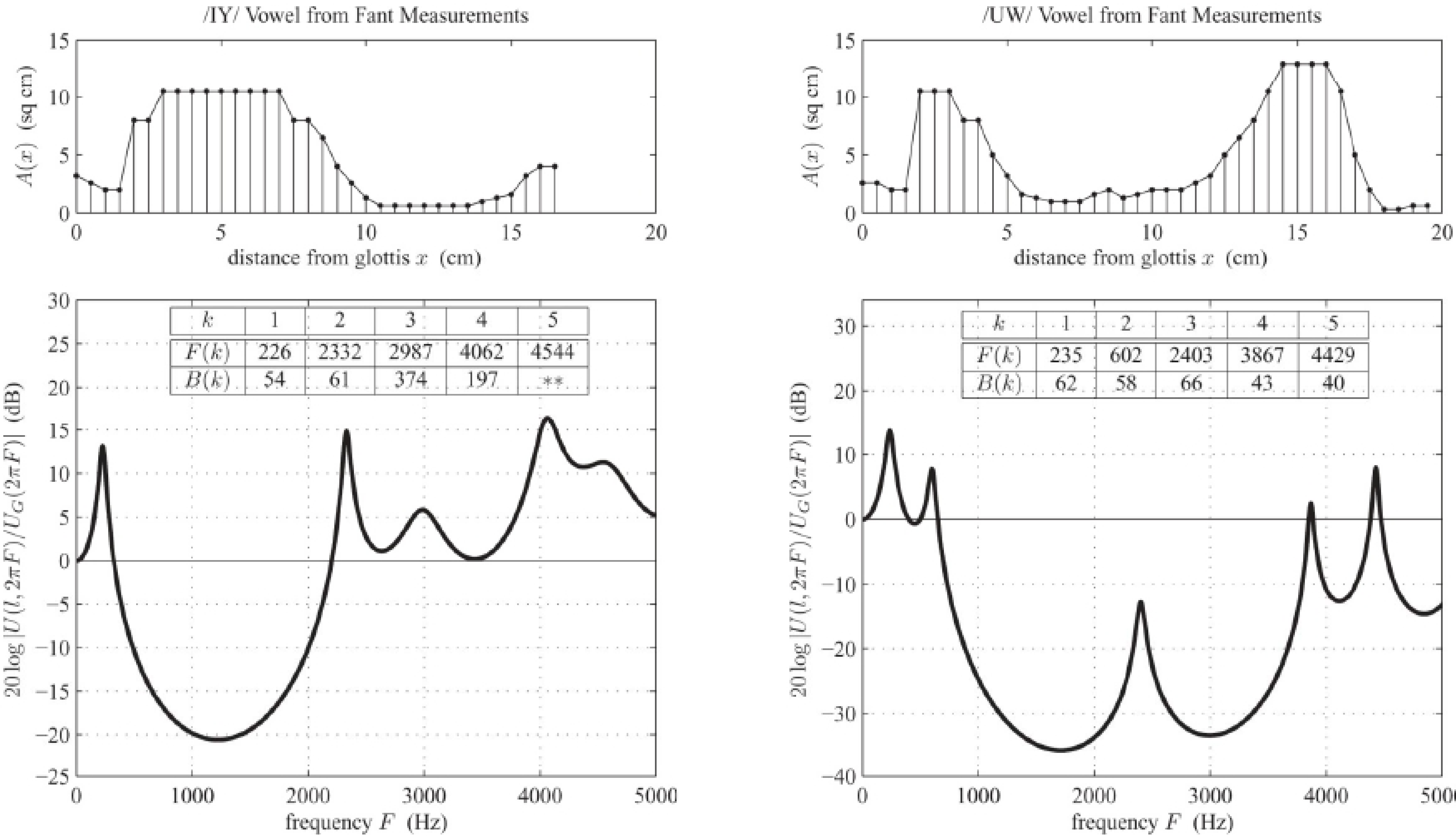
Gunnar Fant,
Acoustic Theory of Speech Production,
Mouton, 1970

Area Functions and FR for Vowels /AA/ and /EH/





Area Functions and FR for Vowels /IY/ and /UW/



VT Transfer Functions

- the vocal tract tube can be characterized by a set of resonances (formants) that depend on the vocal tract area function-with shifts due to losses and radiation
- the bandwidths of the two lowest resonances (F1 and F2) depend primarily on the vocal tract wall losses
- the bandwidths of the highest resonances (F3, F4, ...) depend primarily on viscous friction losses friction, thermal losses, and radiation losses

Nasal Coupling Effects

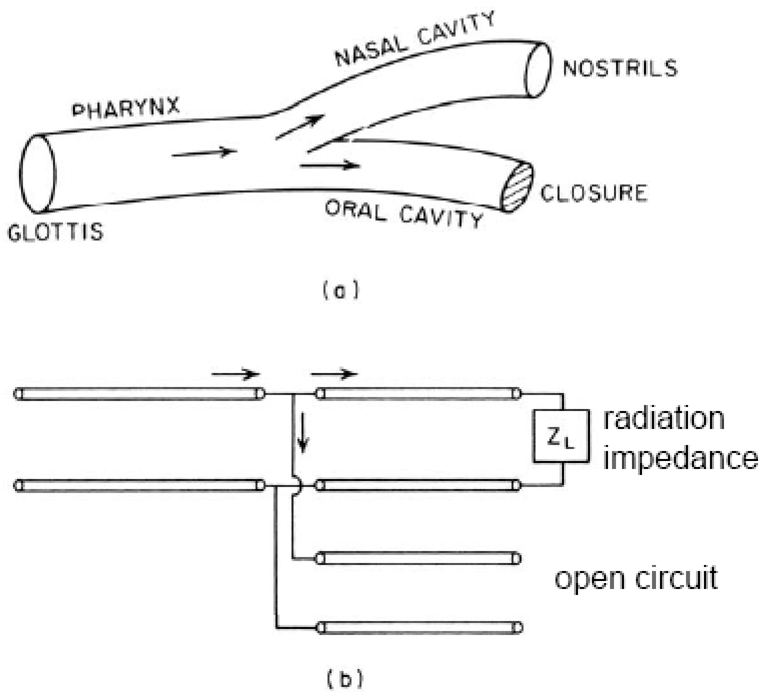


Fig. 3.27 (a) Tube model for production of nasals; (b) corresponding electrical analog.

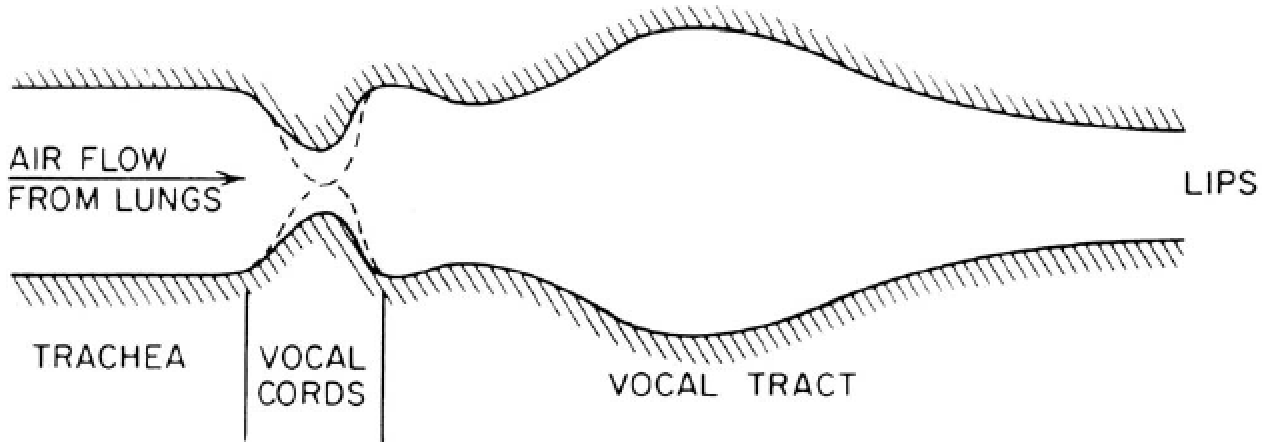
- at the branching point
  - sound pressure the same as at input of each tube
  - volume velocity is the sum of the volume velocities at inputs to nasal and oral cavities
- can solve flow equations numerically
  - results show resonances dependent on shape and length of the 3 tubes
- closed oral cavity can trap energy at certain frequencies, preventing those from appearing in the nasal output => anti-resonances or zeros of the transfer function
- nasal resonances have broader bandwidths than non-nasal voiced sounds => due to greater viscous friction and thermal loss due to large surface area of the nasal cavity

Excitation Sources

Sound Excitation in VT

1. air flow from lungs is modulated by vocal cord vibration, resulting in a quasi-periodic pulse-like source
2. air flow from lungs becomes turbulent as air passes through a constriction in the vocal tract, resulting in a noise-like source
3. air flow builds up pressure behind a point of total closure in the vocal tract => the rapid release of this pressure, by removing the constriction, causes a transient excitation (pop-like sound)

Voiced Excitation in VT



- lung pressure is increased, causing air to flow out of the lungs and through the opening between the vocal cords (the glottis)
- according to Bernoulli's law, if the tension in the vocal cords is properly adjusted, the reduced pressure in the constriction allows the cords to come together, thereby constricting air flow (see dotted lines above)
- because of closure of the vocal cords, pressure increases behind the vocal cords and eventually reaches a level sufficient to force the vocal cords to open and allows air to flow through the glottis again

- sustained Bernoulli oscillations => rate of opening and closing is controlled by air pressure in the lungs, tension 张力 and stiffness 刚性 of the vocal cords, and area of the glottal opening; the vocal tract area at the glottis also effects the rate

Glottal Excitation Model

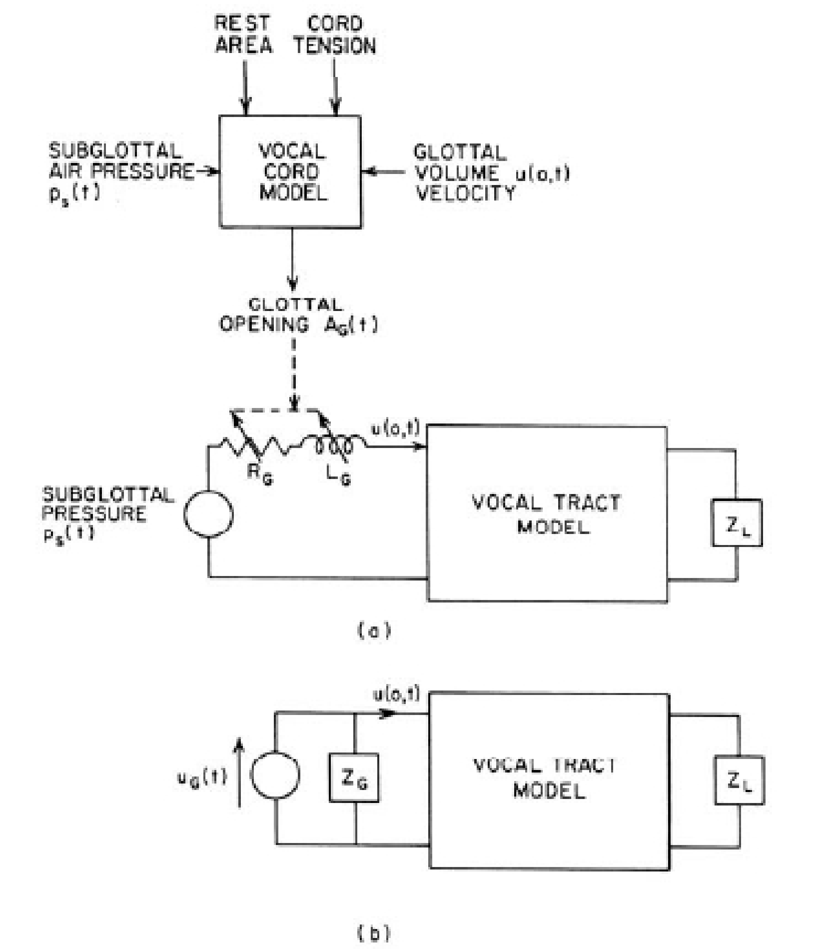
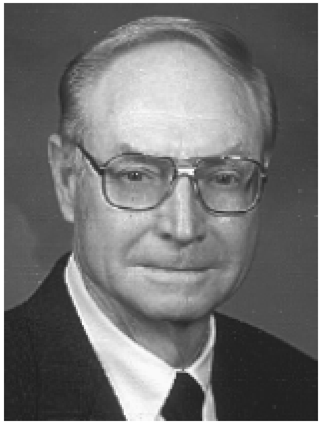


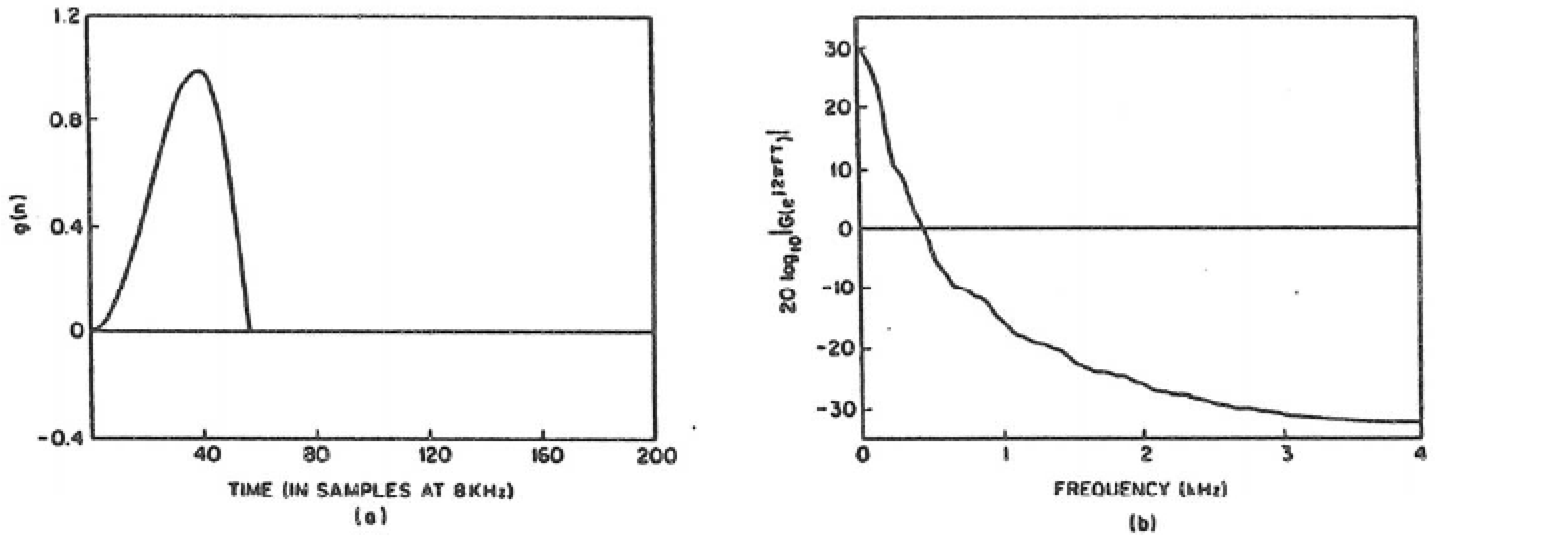
Fig. 3.29 (a) Diagram of vocal cord model; (b) approximate model for vocal cords.

J. L. Flanagan and K. Ishizaka, did the first detailed simulations of vocal cord oscillators. Subsequent researchers have refined the model for singing voice.



- vocal tract acts as a load on the vocal cord oscillator
- time varying glottal resistance and inductance-both functions of $1/A_G(t)$ => when $A_G(t)=0$ (total closure), impedance is infinite and volume velocity is zero

Rosenberg Glottal Pulse and Spectrum



$$
\begin{aligned}
g[n] &= 0.5[1-\cos(\pi n / N_1)] && 0 \le n \le N_1 \\
&= \cos[\pi(n-N_1)/(2N_2)] && N_1 \le n \le N_1 + N_2 \\
&= 0 && \text{otherwise}
\end{aligned}
$$

Note the high frequency fall off due to the lowpass pulse shape

Other Excitation Sources

- voiceless excitation occurs at a constriction of the vocal tract when volume velocity exceeds a critical value (called the Reynolds number) => this can be modeled using a randomly time varying source at the point of constriction
- a combination of voiced and voiceless excitation is used for voiced fricatives
- a total closure of the tract is used for stop consonants

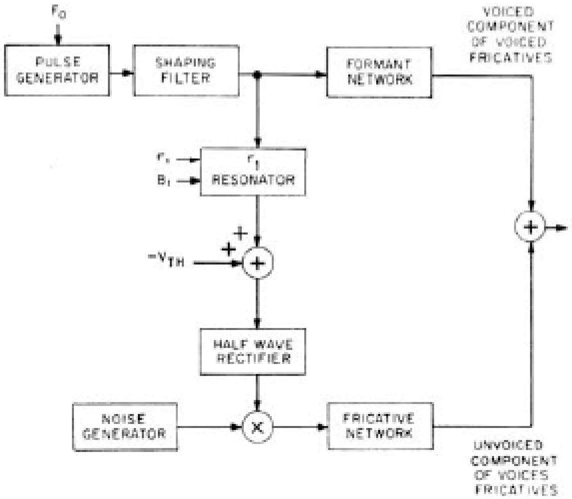


FIG. 2. Excitation network for voiced fricatives.

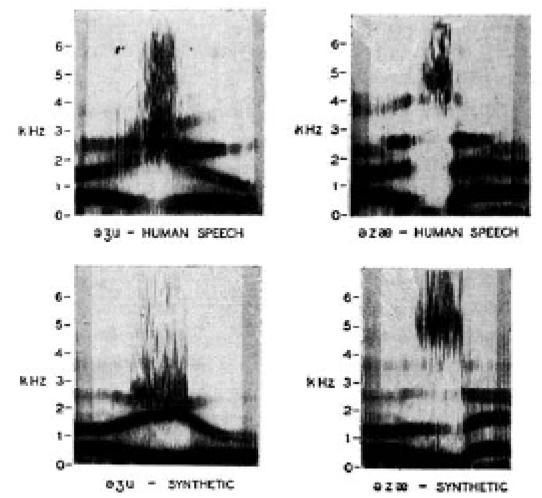
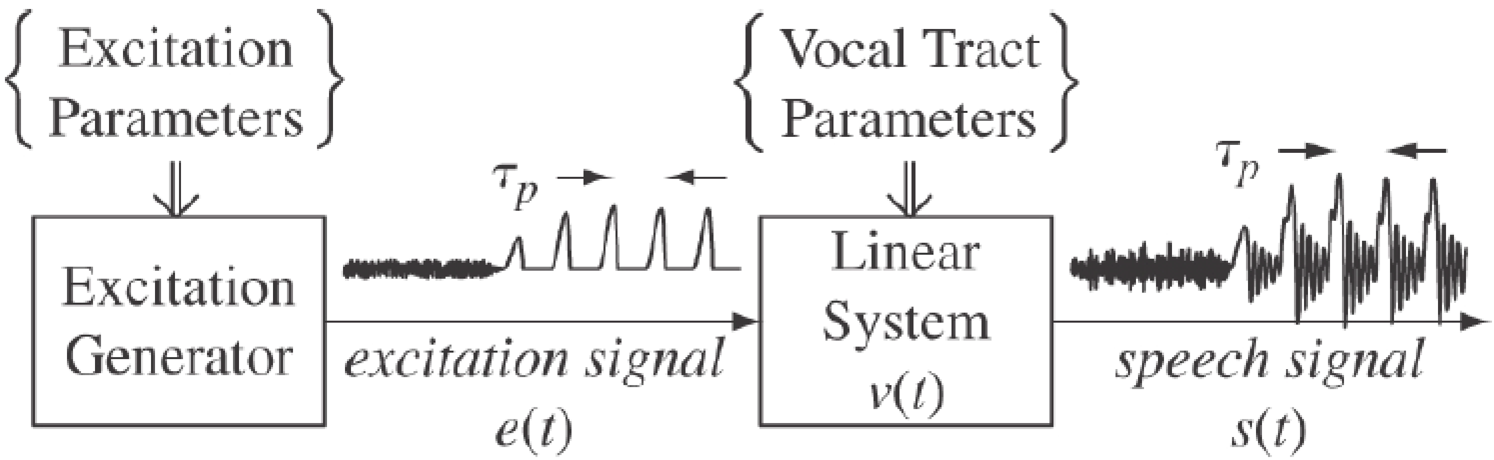


FIG. 3. Spectrographic examples of voiced fricatives.

Source-System Model



Pitch,
Voiced/Unvoiced,
Amplitude

Formants, Vocal Tract Area
Functions, Articulatory
Paramameters

Summary of Losses, Radiation and Boundary Condition Effects

- considered losses due to friction at walls, heat conduction through walls, vibration of walls
- losses introduce new terms into sound propagation equations
- effects of losses are increased bandwidth of complex poles (from 0 to a finite quantity) and changes in the regular spacing of the resonance (formant) frequencies of the tract
- radiation at lips adds a parallel resistance and inductance component and is most significant at higher frequencies
- nasal coupling adds components to solution which include anti-resonances (frequency response zeros)
- sound excitation models lead to simplified model with a distinct glottal pulse (for voiced speech) with strong high frequency drop-off in level
- the overall vocal tract is well modeled as a variable excitation generator exciting a linear time-varying system

Concatenated Lossless Tubes

Lossless Tube Models

- approximate $A(x)$ by a series of lossless, constant cross sectional area, acoustic tubes of the form shown at the right
- as the number of tubes becomes larger (smaller approximation error for the vocal tract area function), the approximation error for modeling the vocal tract goes to zero

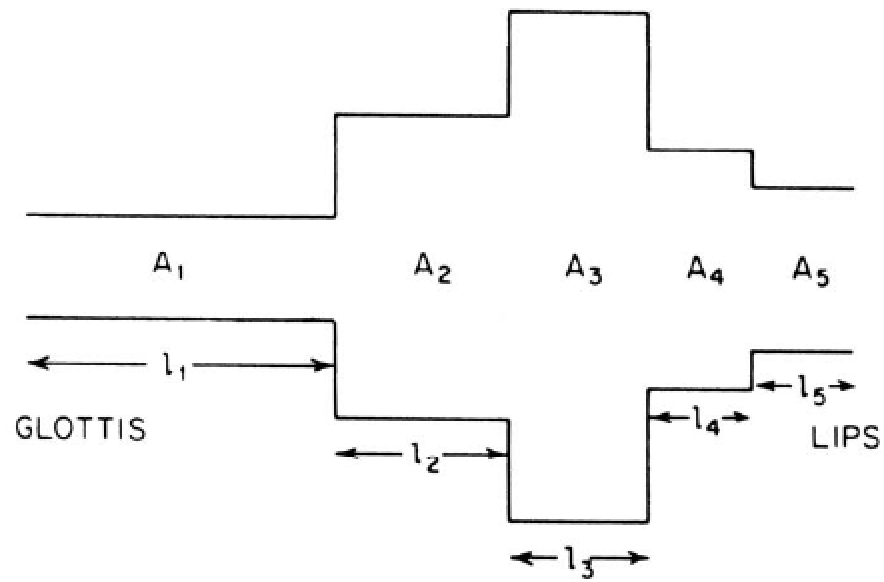
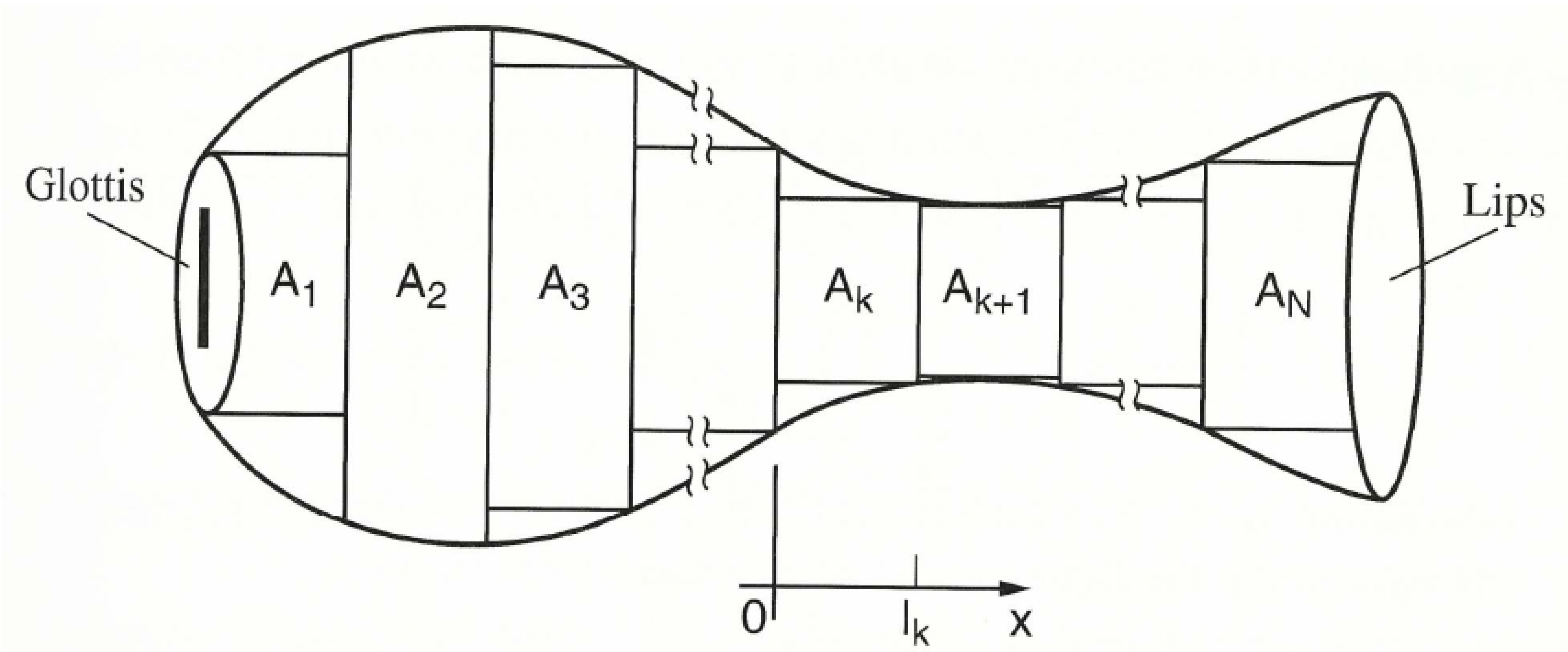


Fig. 3.32 Concatenation of 5 lossless acoustic tubes.

Concatenated (拼接) Tube Models



Lossless Tube Models

1. The vocal tract area function, *A,* is now a function of *x, A(x)*

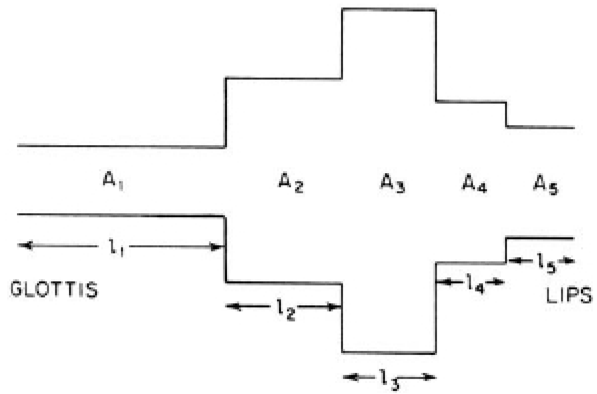


Fig. 3.32 Concatenation of 5 lossless acoustic tubes.

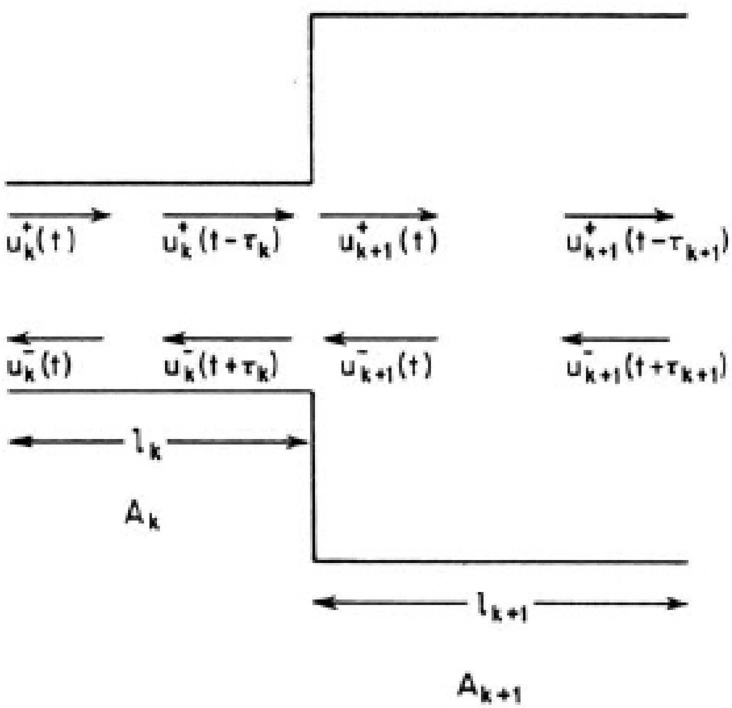
2. Solve the wave equation for the *kth* tube

$$p_k(x,t) = \frac{\rho c}{A}\left[u_k^+(t - x/c) + u_k^-(t + x/c)\right], \qquad 0 \le x \le \ell_k$$

$$u_k(x,t) = \left[u_k^+(t - x/c) - u_k^-(t + x/c)\right], \qquad 0 \le x \le \ell_k$$

Lossless Tube Models

3. add boundary conditions at the edges of adjacent tubes: both pressure and volume velocity must be continuous in both time and space at boundaries



$$p_k(\ell_k, t) = p_{k+1}(0, t)$$
$$u_k(\ell_k, t) = u_{k+1}(0, t)$$

Fig. 3.33 Illustration of the junction between two lossless tubes.

Lossless Tube Models

4. at each junction:
   – part of the positive going wave is propagated to the right while part is reflected back to the left
   – part of the negative going wave is propagated to the left while part is reflected back to the right

5. combine 2 & 3

$$u_{k+1}^{+}(t) = (1 + r_k) u_k^{+}(t - \tau_k) + r_k u_{k+1}^{-}(t)$$

$$u_k^{-}(t + \tau_k) = -r_k u_k^{+}(t - \tau_k) + (1 - r_k) u_{k+1}^{-}(t)$$

$$\tau_k = \ell_k / c$$

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} =$$ reflection coefficient for the k^{th} junction

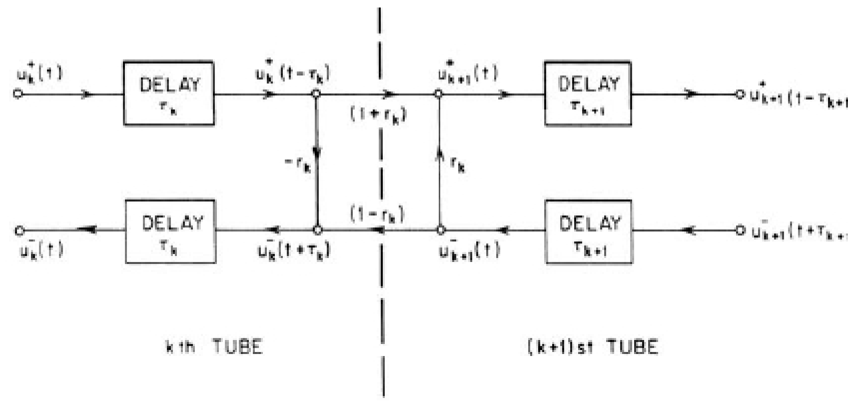


Fig. 3.34 Signal-flow representation of the junction between two lossless tubes.

Lossless Tube Models

6. for an N-tube model there are (N-1) junctions with reflection coefficients
   – there are boundary conditions at the lips and glottis
7. relating $p_N(l_N, t)$ and $u_N(l_N, t)$ to pressure and volume velocity at the lips via $P_N(\ell_N,\Omega) = z_L U_N(\ell_N,\Omega)$ where z_L is lip impedance, we get

$$u_N^-(t+\tau_N) = -r_L u_N^+(t-\tau_N) \qquad r_L = \frac{\rho c / A_N - z_L}{\rho c / A_N + z_L}$$

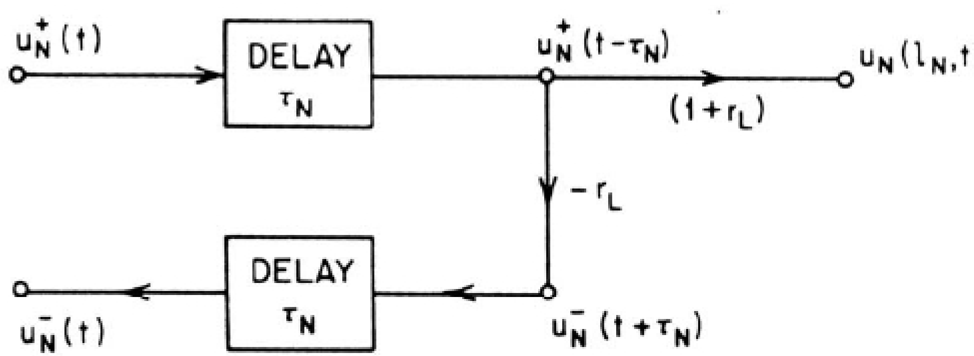
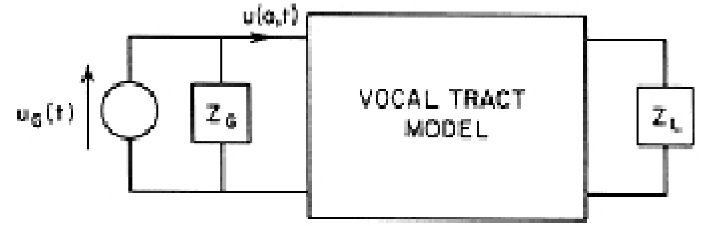


Fig. 3.35 Termination at lip end of a concatenation of lossless tubes.

Lossless Tube Models

8. using boundary condition at the glottis

$$U_1(0,\Omega) = U_G(\Omega) - P_1(0,\Omega)/Z_G$$



where z_G is glottal impedance, we get

$$u_1^+(t) = \frac{(1+r_G)}{2} u_G(t) + r_G u_1^-(t) \qquad r_G = \frac{z_G - \rho c / A_1}{z_G + \rho c / A_1}$$

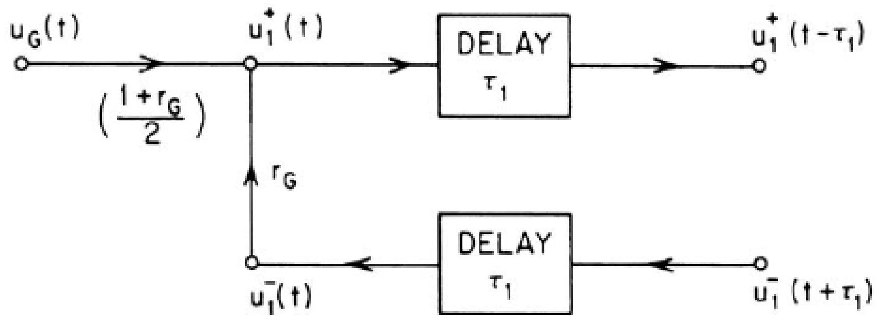


Fig. 3.36 Termination at glottal end of a concatenation of lossless tubes.

Lossless Two Tube Model

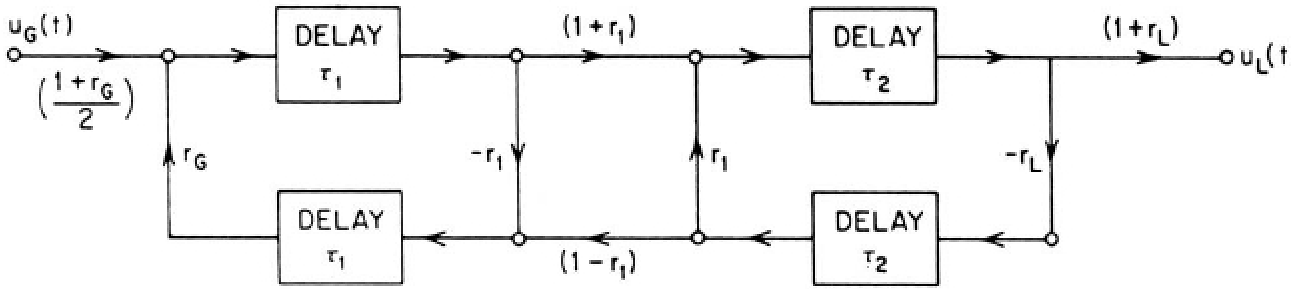


Fig. 3.37 Complete flow diagram of a two-tube model.

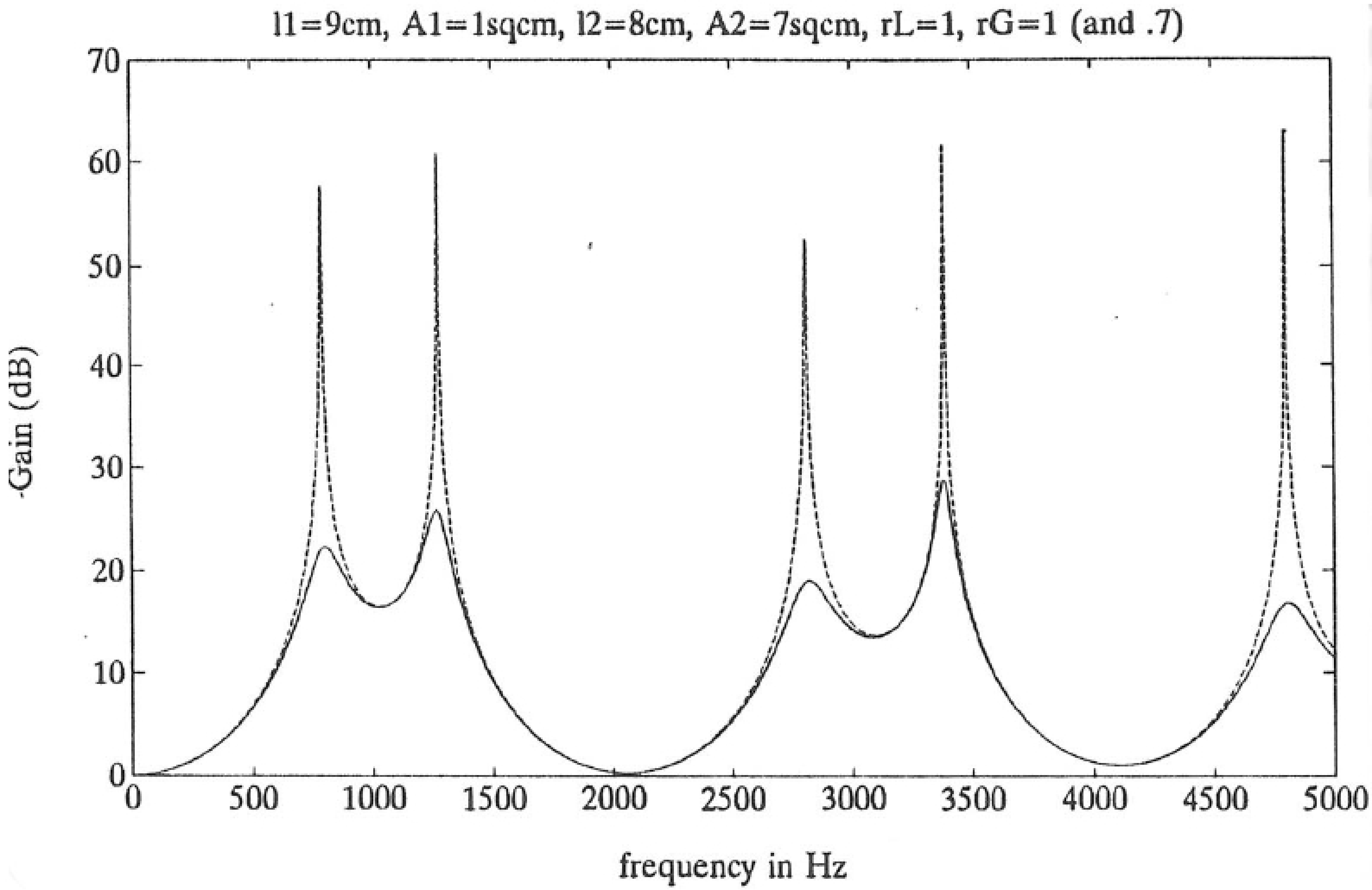
- volume velocity at lips is $u_L(t) = u_2(\ell_2, t)$
- transfer function from glottis to lips is

$$V_a(\Omega) = \frac{U_L(\Omega)}{U_G(\Omega)}$$

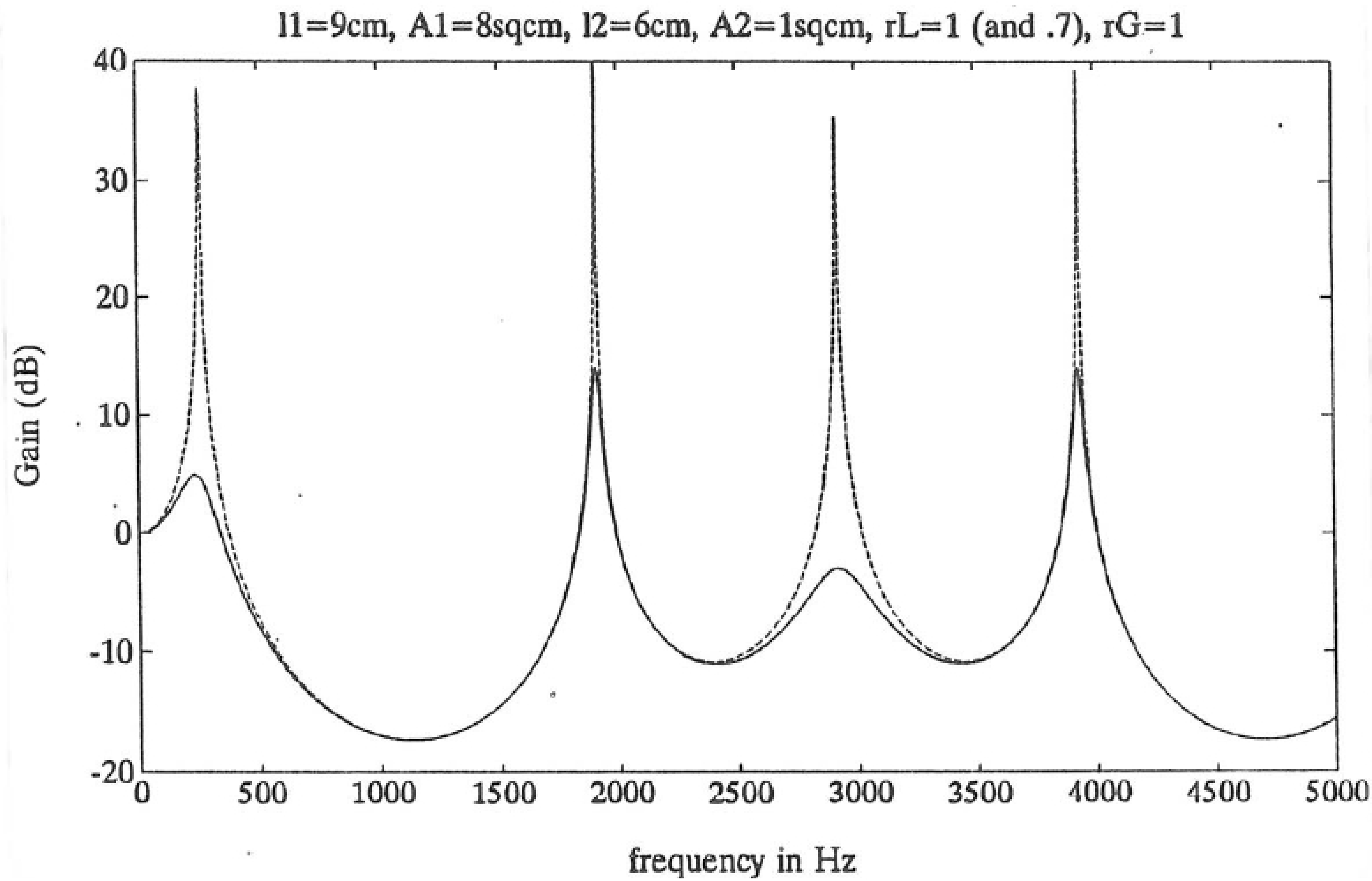
$$= \frac{0.5(1+r_G)(1+r_L)(1+r_1)e^{-j\Omega(\tau_1+\tau_2)}}{1 + r_1 r_G e^{-j\Omega 2\tau_1} + r_1 r_L e^{-j\Omega 2\tau_2} + r_L r_G e^{-j\Omega 2(\tau_1+\tau_2)}}$$

- note total delay of $(\tau_1 + \tau_2)$ is total propagation delay from glottis to lips

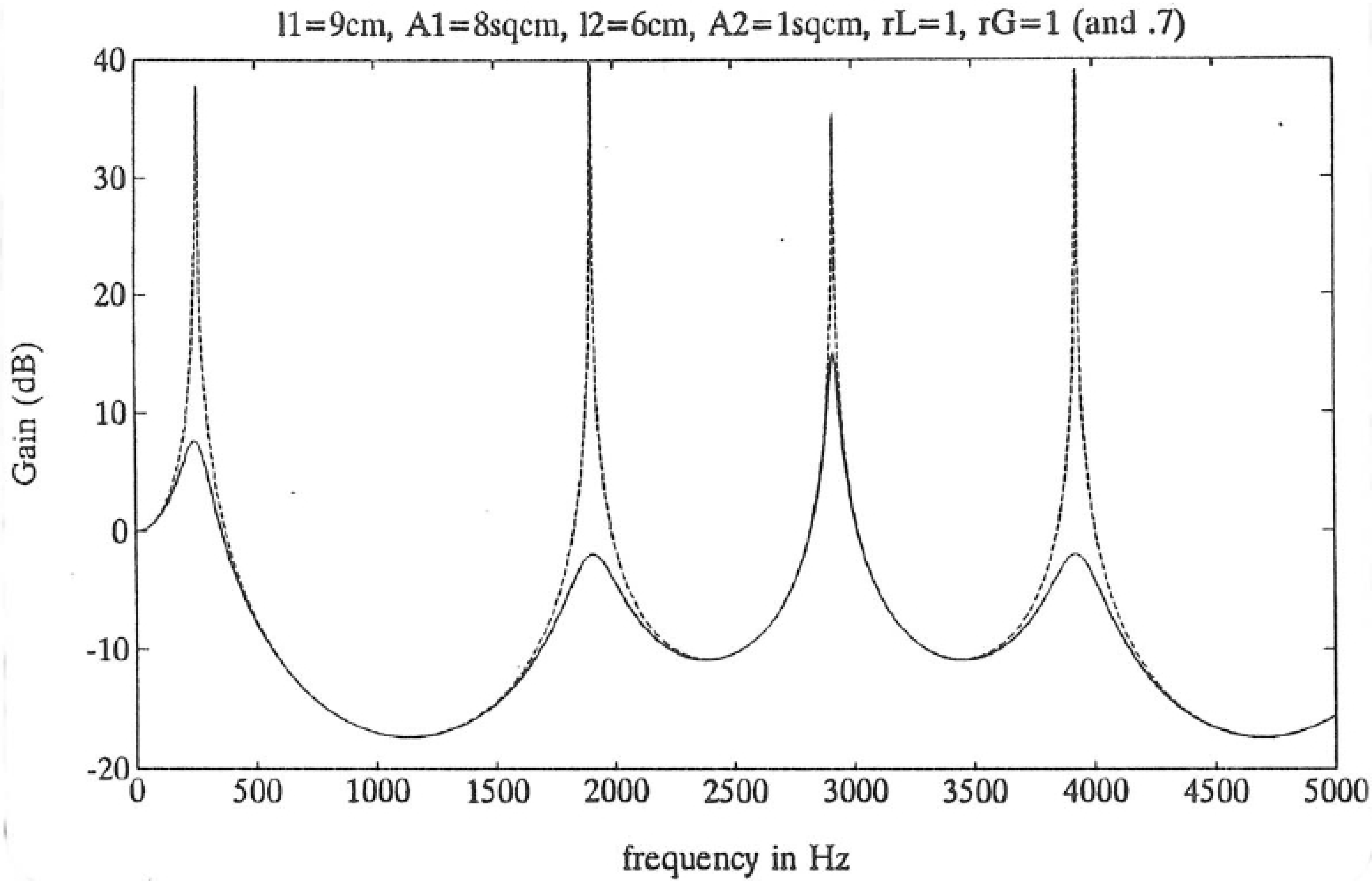
Two-Tube Model for Vowel /AA/



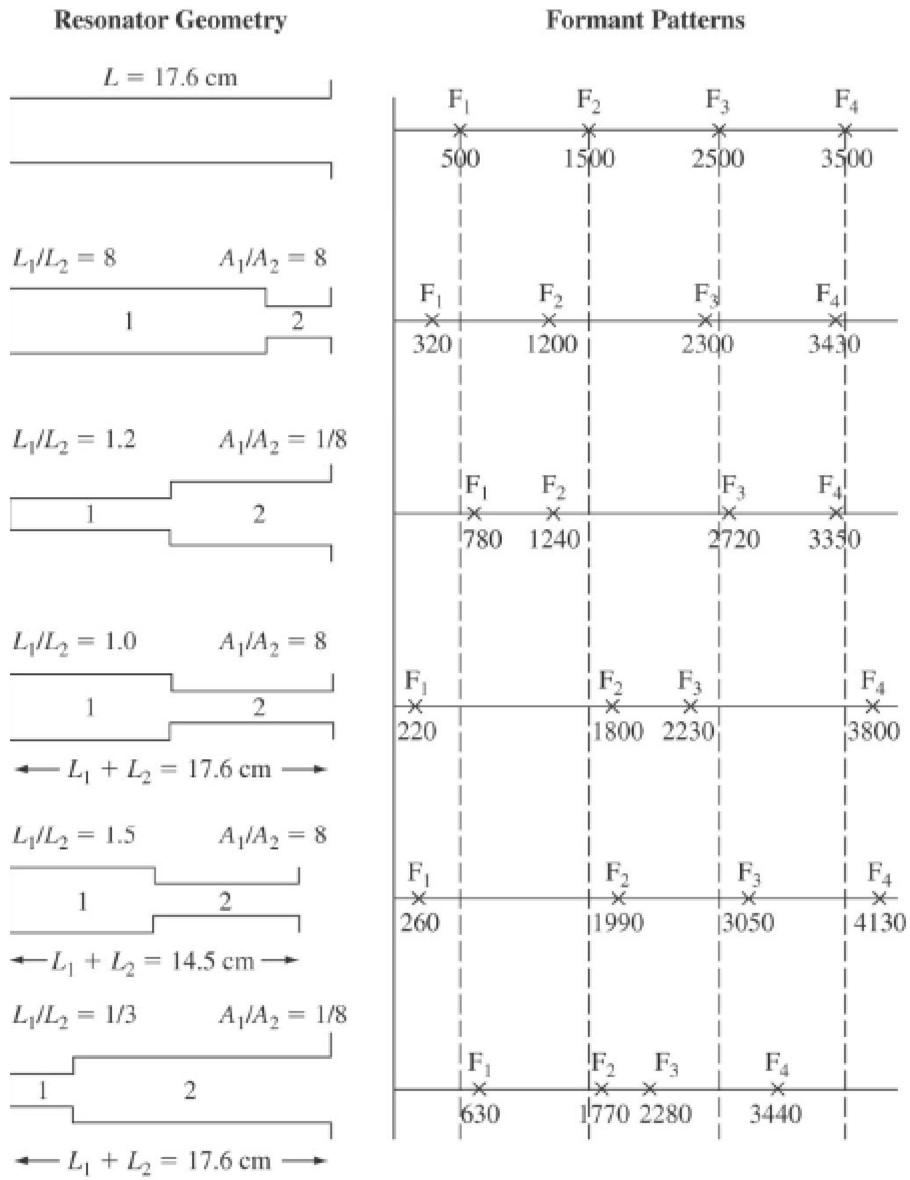
Two-Tube Model for Vowel /IY/ (Losses at Lips)



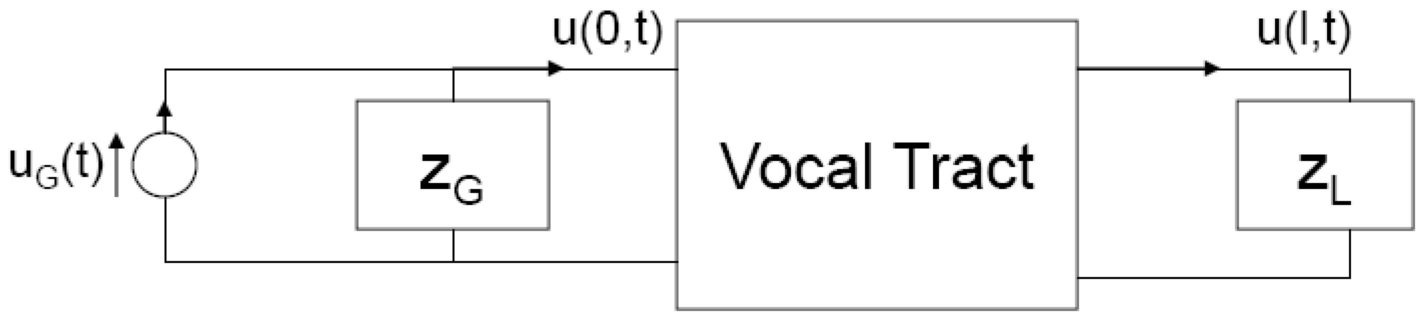
Two-Tube Model for Vowel /IY/ (Losses at Glottis)



Two Tube Model Resonances



Summary of Lossless Tube Models



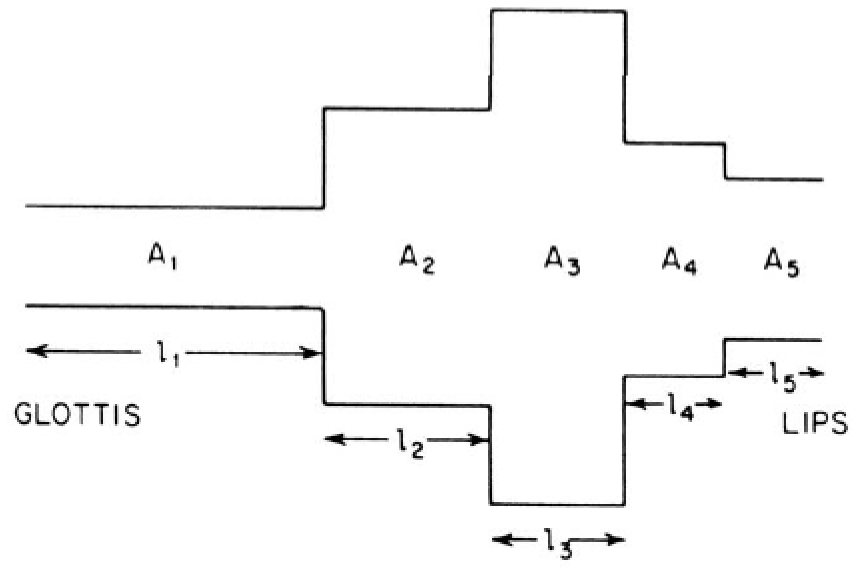
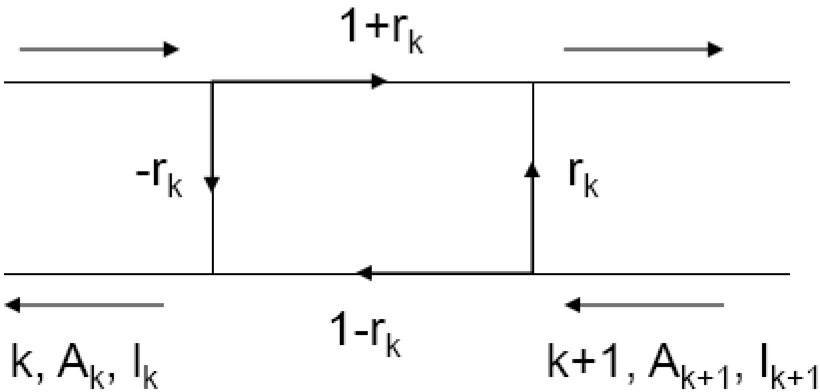


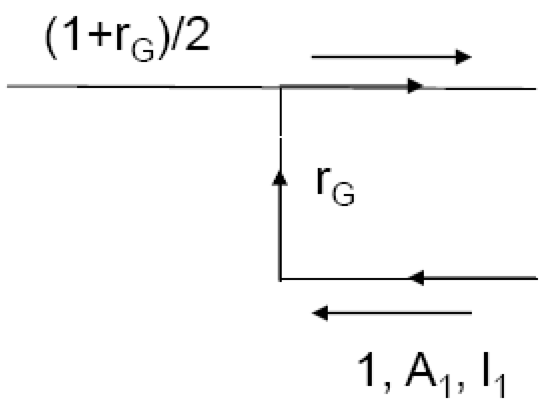
Fig. 3.32 Concatenation of 5 lossless acoustic tubes.

Summary of Lossless Tube Models

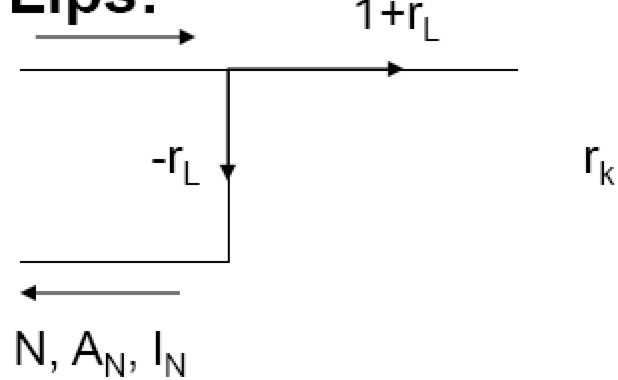


$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}$$; reflection coefficient

Glottis:



Lips:



$$r_L = \frac{\frac{\rho c}{A_N} - z_L}{\frac{\rho c}{A_N} + z_L}$$; reflection coefficient

$$r_G = \frac{-\frac{\rho c}{A_1} + z_G}{\frac{\rho c}{A_1} + z_G}$$; reflection coefficient

Relationship to Digital Filters

- observation that lossless tube model appears similar to digital filter implementations => consider a system of *N* lossless tubes, each of length $\Delta x = l/N$ where *l* is the overall length of the vocal tract

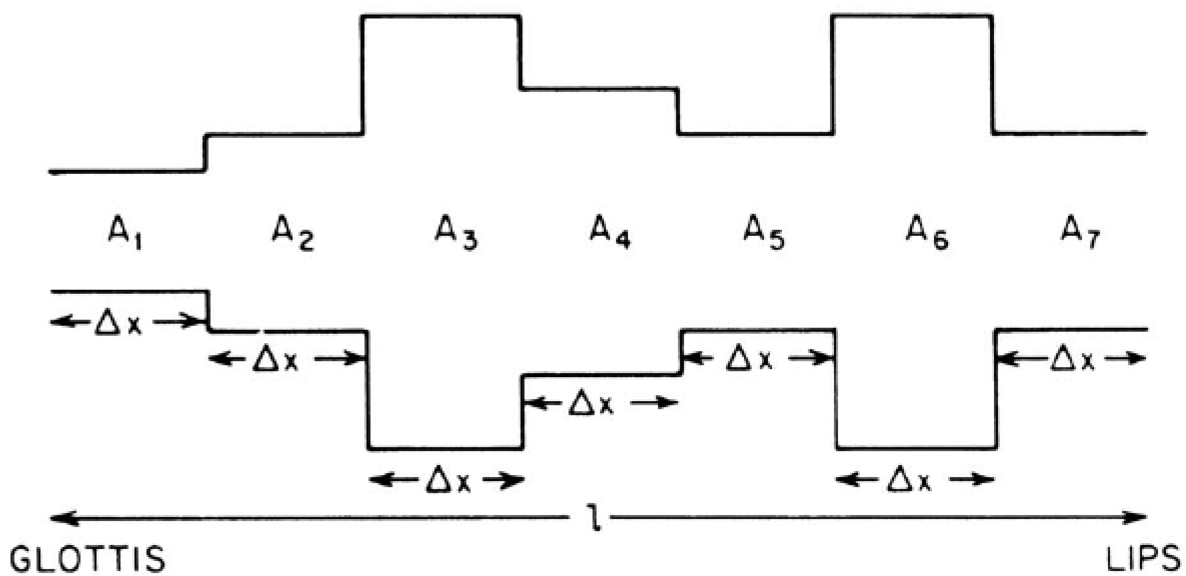
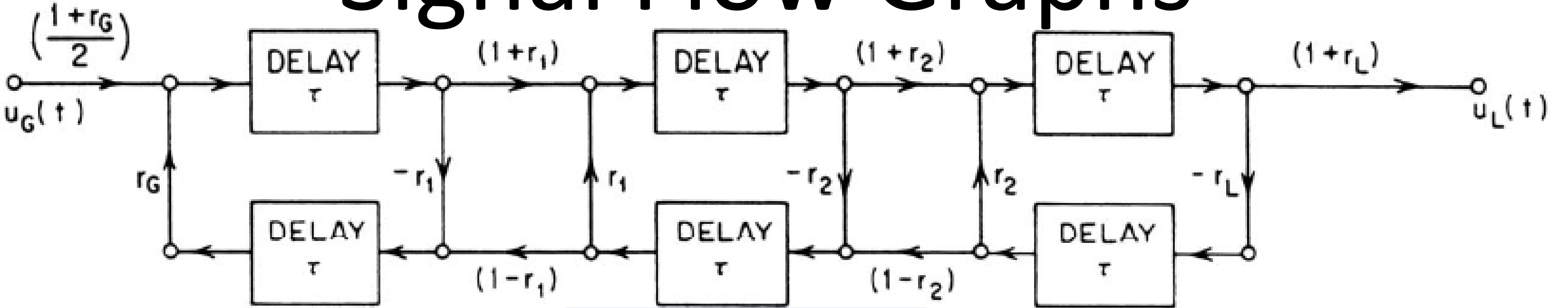


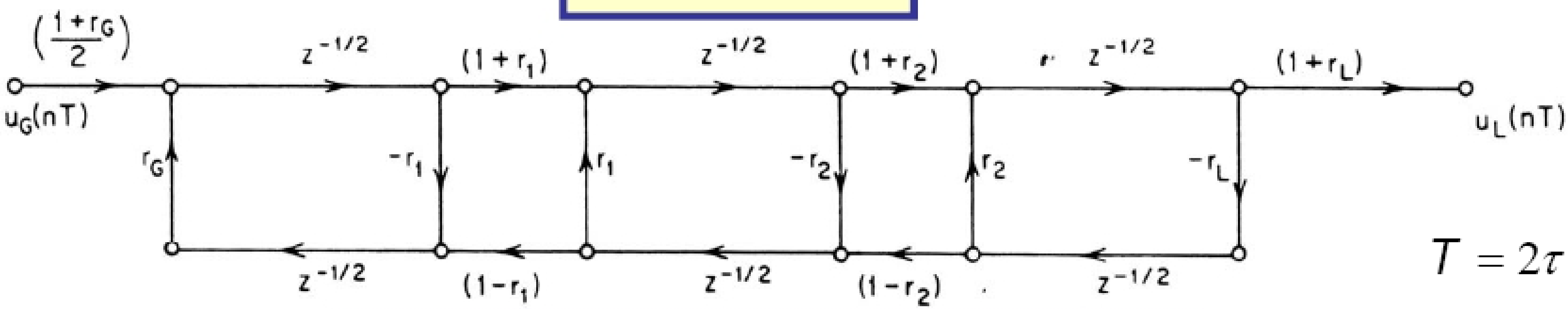
Fig. 3.38 Concatenation of $(N=7)$ lossless tubes of equal length.

- all delays equal to $\tau = \Delta x / c$, the time to traverse the length of one tube

Signal Flow Graphs

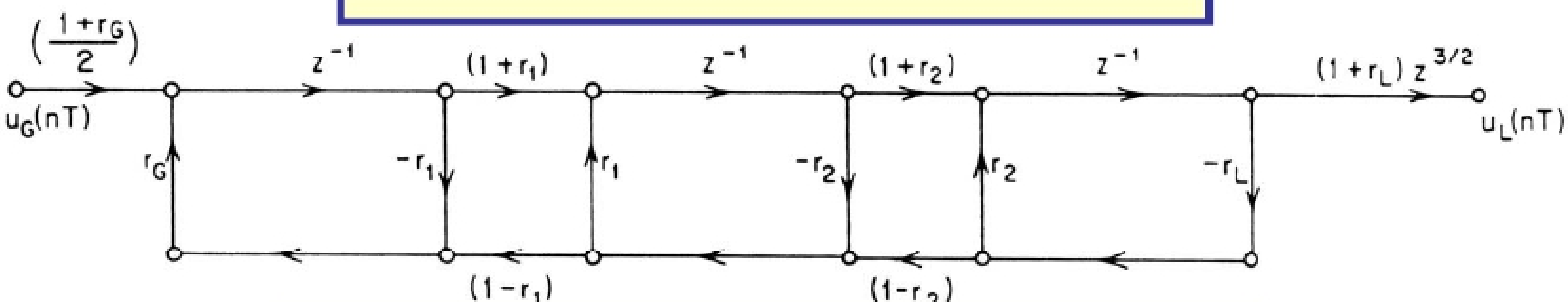


Analog model



$T = 2\tau$

D-T equivalent for bandlimited inputs



D-T equivalent without half-sample delays

Transfer Function of Lossless Tube Model

- want to determine

$$V(z)=\frac{U_L(z)}{U_G(z)}$$

- at junctions we have the relations

$$U_{k+1}^{+}(z)=U_k^{+}(z)z^{-1/2}\left(1+r_k\right)+U_{k+1}^{-}(z)r_k \;\Rightarrow U_k^{+}(z)=\frac{z^{1/2}}{1+r_k}U_{k+1}^{+}(z)-\frac{r_k z^{1/2}}{1+r_k}U_{k+1}^{-}(z)$$

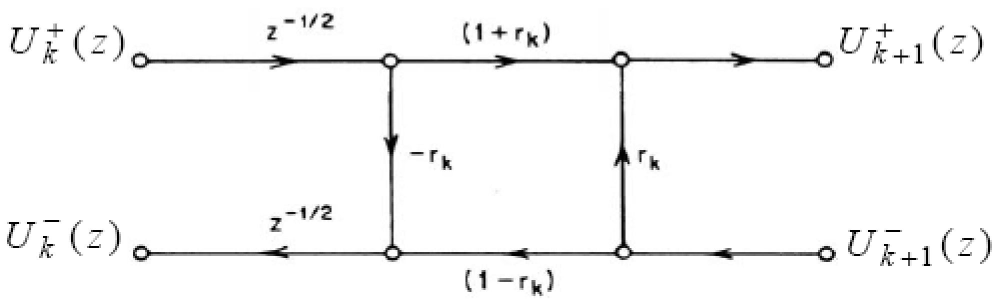
$$U_k^{-}(z)=U_k^{+}(z)z^{-1}\left(-r_k\right)+U_{k+1}^{-}(z)(1-r_k)z^{-1/2} \;\Rightarrow\; U_k^{-}(z)=\frac{-r_k z^{-1/2}}{1+r_k}U_{k+1}^{+}(z)+\frac{z^{-1/2}}{1+r_k}U_{k+1}^{-}(z)$$

- at lips use same formulation with fictitious (*N*+1)st tube that is infinitely long (no negative going wave)=> (*N*+1)st tube terminated in its characteristic impedance

$$U_{N+1}^{+}(z)=U_L(z)$$

$$U_{N+1}^{-}(z)=0$$

$$A_{N+1}=\frac{\rho c}{Z_L}, \qquad r_N=r_L$$



Transfer Function of Lossless Tube Model

- at the glottis

$$U_G(z) = \frac{2}{1+r_G} U_1^+(z) - \frac{2r_G}{1+r_G} U_1^-(z)$$

- putting it all together gives

$$\frac{U_G(z)}{U_L(z)} = \frac{1}{V(z)} = \left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \prod_{k=1}^{N} Q_k \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$Q_k = z^{1/2} \begin{bmatrix} \frac{1}{1+r_k} & \frac{-r_k}{1+r_k} \\ \frac{-r_k z^{-1}}{1+r_k} & \frac{z^{-1}}{1+r_k} \end{bmatrix} = z^{1/2} \hat{Q}_k$$

$$\frac{1}{V(z)} = z^{N/2} \left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \prod_{k=1}^{N} \hat{Q}_k \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Transfer Function of Lossless Tube Model

- consider a 2-section tube (*N=2)*

$$\frac{1}{V(z)} = \frac{2(1+r_1 r_2 z^{-1} + r_1 r_G z^{-1} + r_2 r_G z^{-2})z}{(1+r_G)(1+r_1)(1+r_2)}$$

$$V(z) = \frac{0.5(1+r_G)(1+r_1)(1+r_2)z^{-1}}{1+(r_1 r_2 + r_1 r_G)z^{-1} + r_2 r_G z^{-2}}$$

- in general

$$V(z) = \frac{0.5(1+r_G)\left[\prod_{k=1}^{N}(1+r_k)\right]z^{-N/2}}{D(z)}$$

$$D(z) = [1, -r_G]\begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_N \\ -r_N z^{-1} & z^{-1} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$D(z) = 1 - \sum_{k=1}^{N} \alpha_k z^{-k}$$

Transfer Function of Lossless Tube Model

- Choose N=10 as a reasonable number of tubes for model
- special case

$$r_G = 1 \; (Z_G = \infty)$$

$$r_N = 1 \Rightarrow A_{N+1} = \infty \quad \text{(infinite tube at lips)}$$

$$r_N = 0.714 \;\; \Rightarrow \;\; A_{N+1} = 28 \text{ cm}^2$$

Transfer Function of Lossless Tube Model

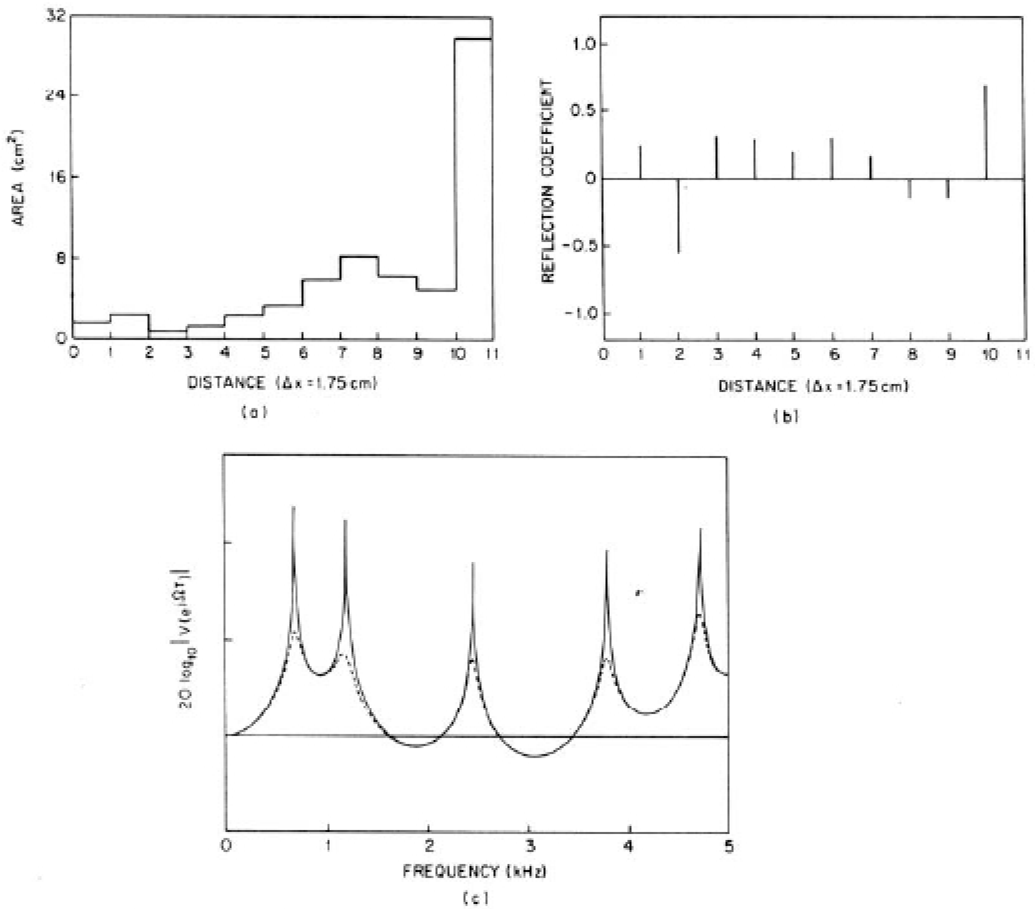


Fig. 3.43 (a) Area function for 10 section lossless tube terminated with reflectionless section of area 30 cm^2; (b) reflection coefficients for 10 section tube; (c) frequency response of 10 section tube; dotted curve corresponds to conditions of (b); solid curve corresponds to short-circuit termination. (Note area data of (a) estimated from data given by Fant [1] for the Russian vowel /a/.)

Other Synthesis Implementations

- direct form difference equation
- cascade of second order systems

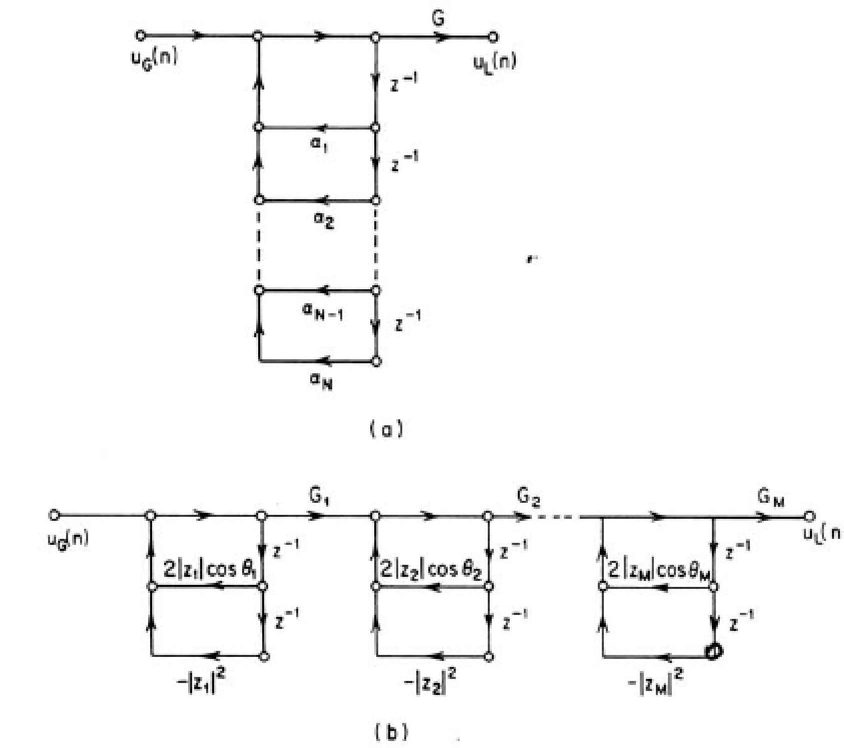


Fig. 3.46 (a) Direct form implementation of all-pole transfer function; (b) cascade implementation of all-pole transfer function ($G_k = 1 - 2|z_k|cos\theta_k + |z_k|^2$).

$$V(z) = \prod_{k=1}^{M} V_k(z)$$

$$V_k(z) = \frac{1-2|z_k|\cos(2\pi F_k T)+|z_k|^2}{1-2|z_k|\cos(2\pi F_k T)z^{-1}+|z_k|^2 z^{-2}}$$

Radiation at Lips

$P_L(z) = R(z)U_L(z)$ -- high pass filtering

$R(z) = R_0(1 - z^{-1})$ -- crude differentiator

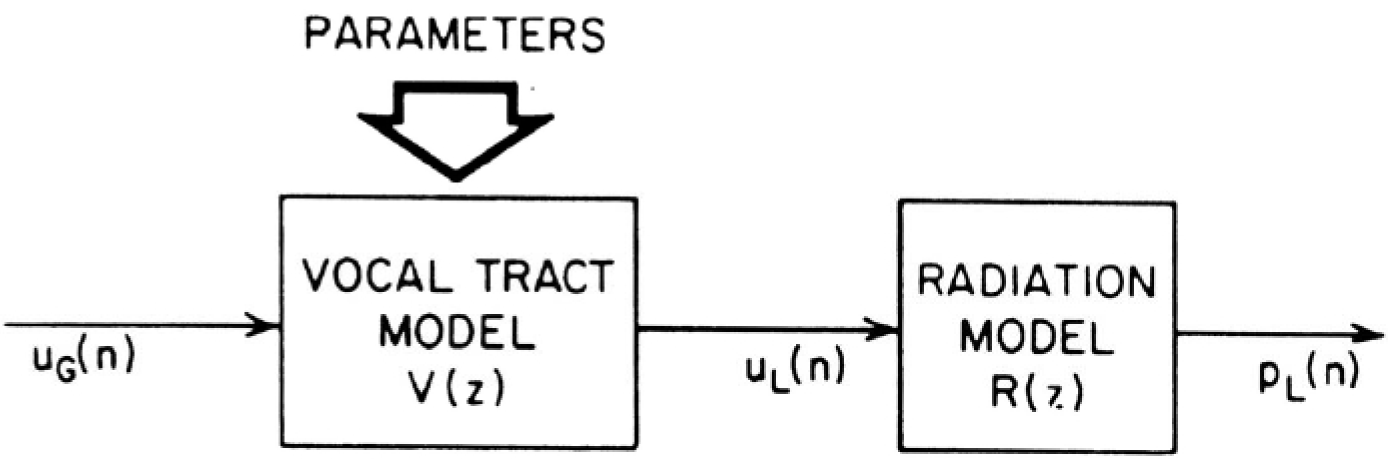


Fig. 3.47 Terminal analog model including radiation effects.

Excitation Model

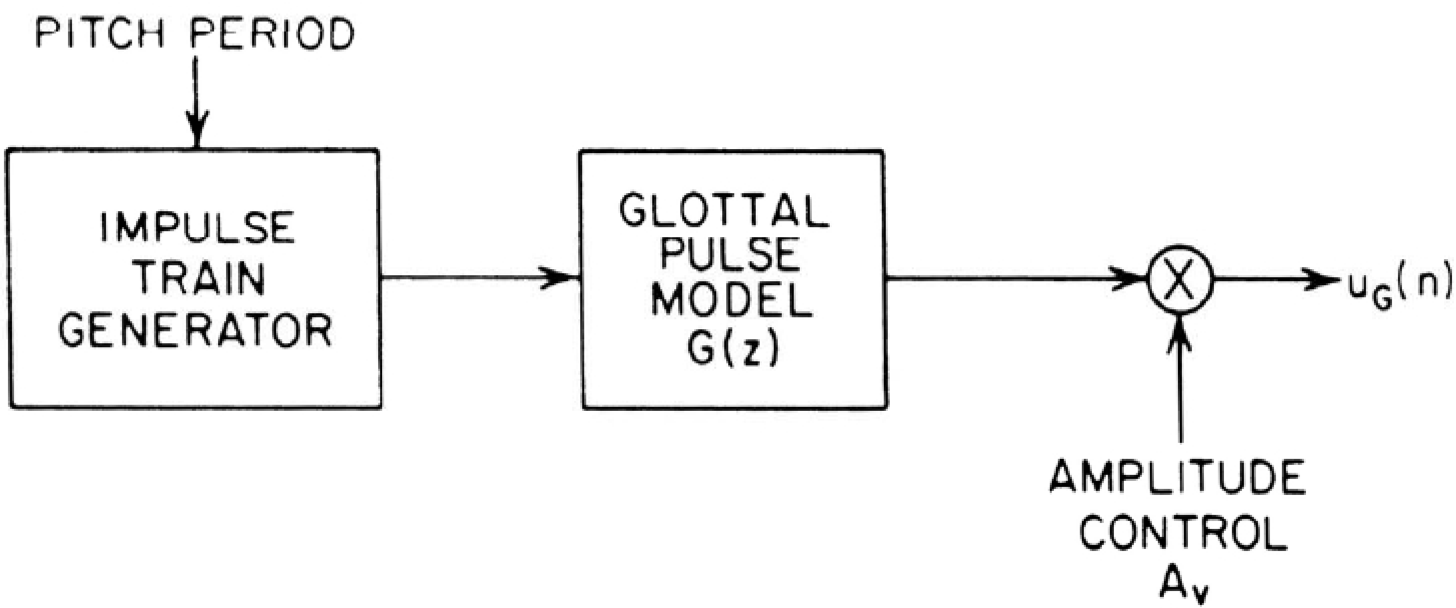
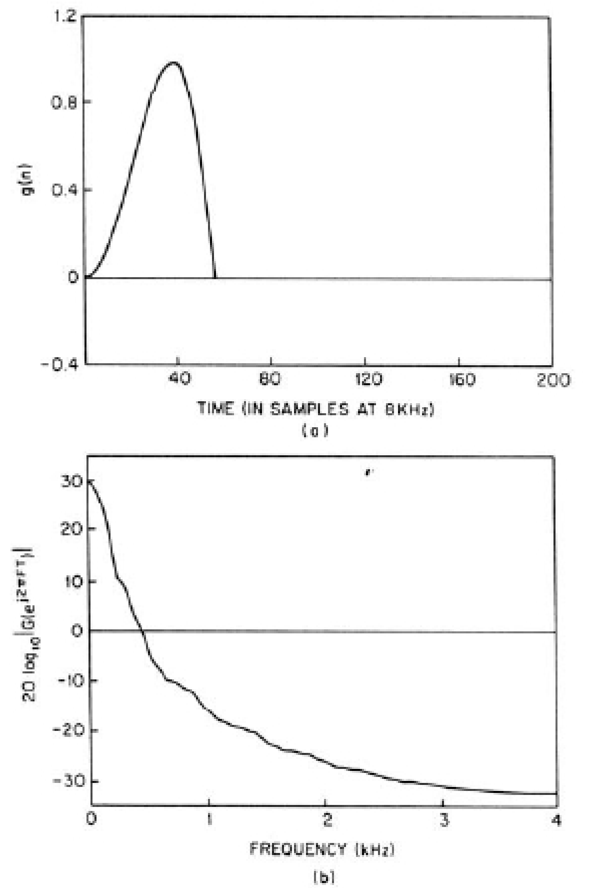


Fig. 3.48 Generation of the excitation signal for voiced speech.

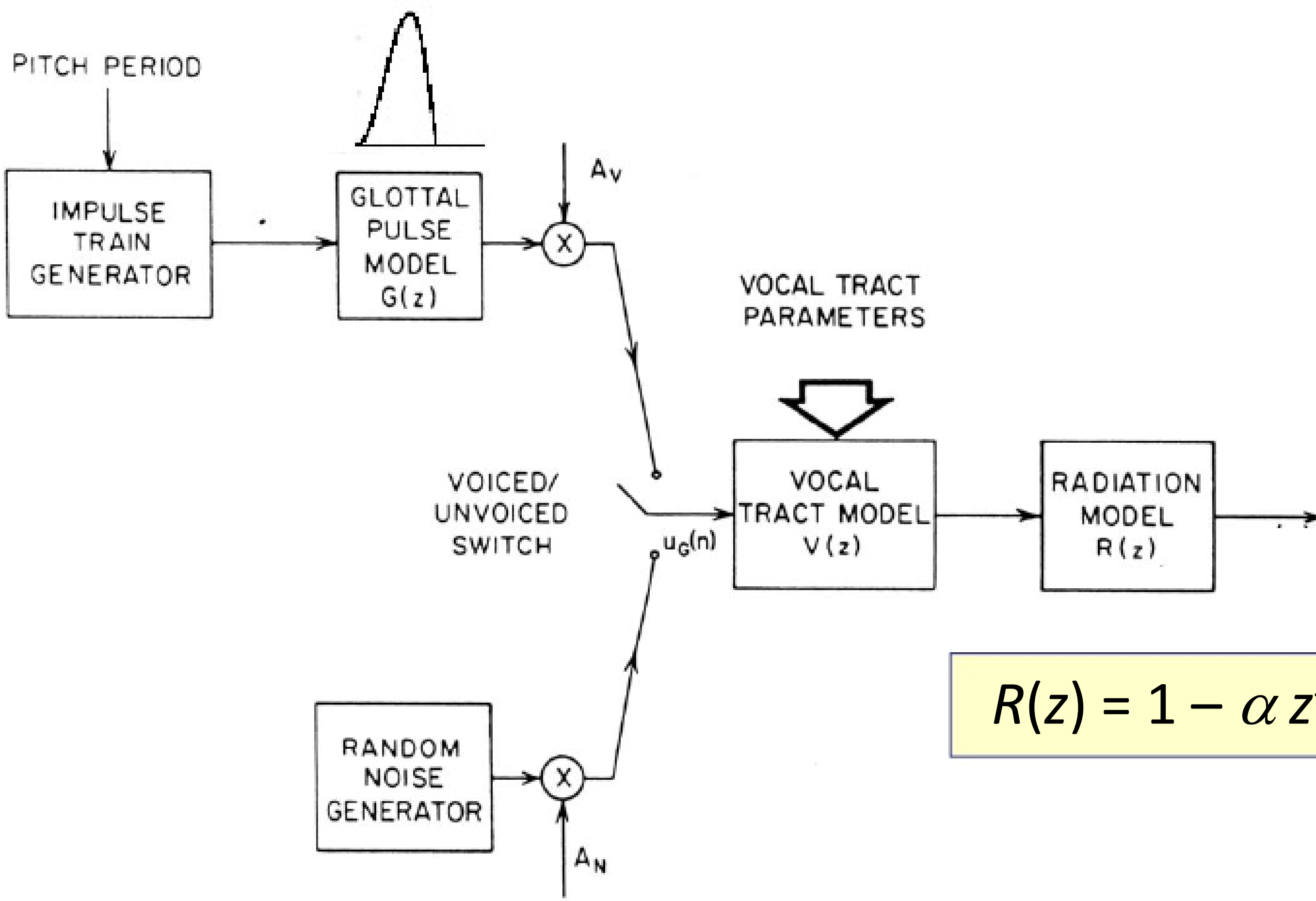
Glottal Pulse Model

$$
\begin{aligned}
g[n] &= 0.5\left[1-\cos(\pi n / N_1)\right], \quad & 0 \le n \le N_1 \\
&= \cos(\pi(n-N_1)/2N_2), \quad & N_1 \le n \le N_1+N_2 \\
&= 0 & \text{otherwise}
\end{aligned}
$$

- lowpass filtering effect



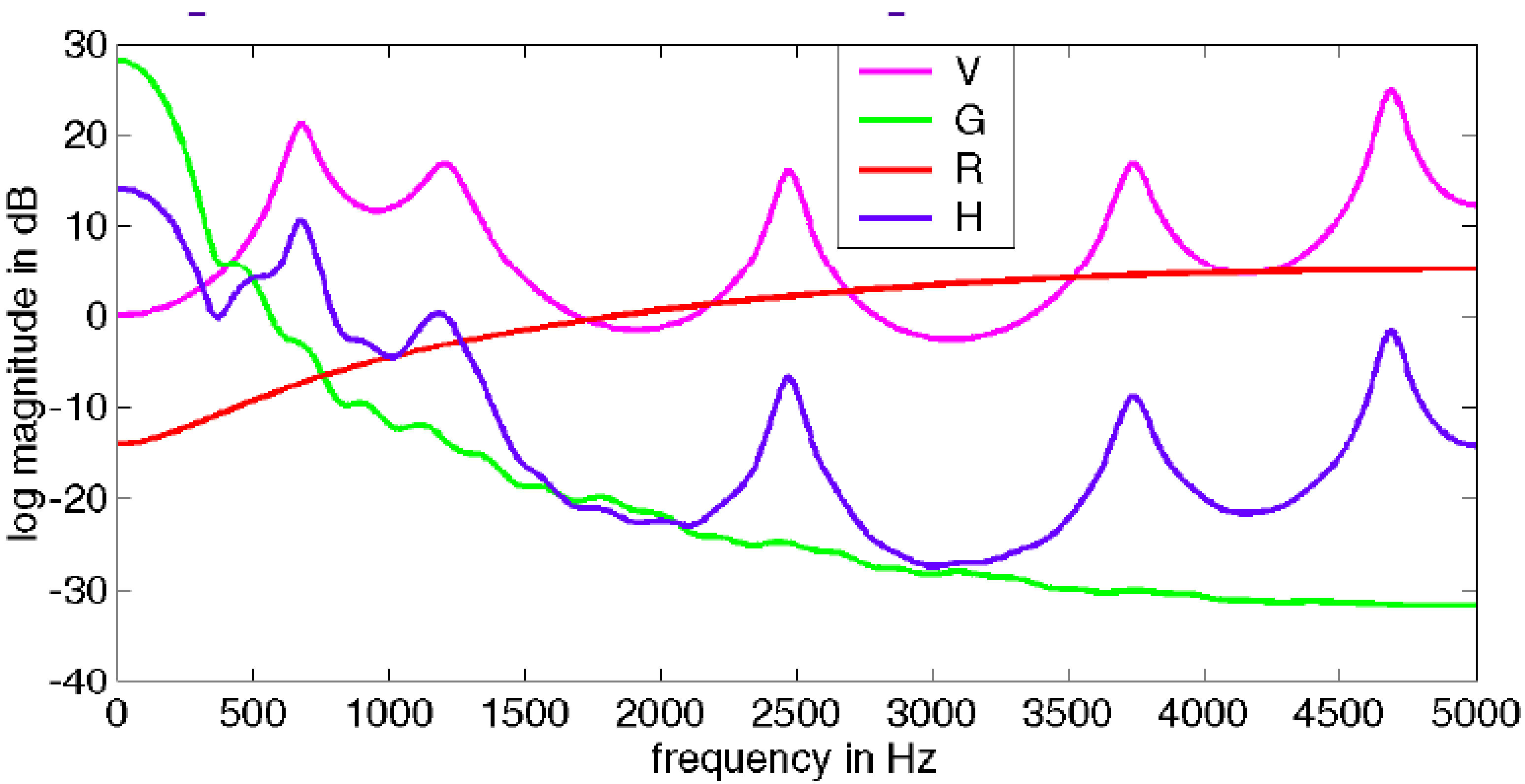
General Synthesis Model



$$R(z) = 1 - \alpha z^{-1}$$

Fig. 3.50 General discrete-time model for speech production.

Components of Speech Model



Summary

- Derived sound propagation equations for vocal tract
  - first considered uniform lossless tube
  - added simple models of loss
  - added model for radiation at lips
  - added source model at glottis
  - added nasal model for nasal tract
  - broadened the model to N-tube approximation—lossless case
  - looked at 2-tube models for simple vowels
  - digital speech production/synthesis models