### Digital Speech Processing—Lectures 5-6

#### Sound Propagation in the Vocal Tract

**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
- some complicating factors:
  - time variation of the vocal tract shape (we will look mainly at fixed shapes)
  - losses in flow at vocal tract 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)

**Bottom Line:** simplify as much as possible and see what we can learn about the mechanics of sound propagation in the human vocal tract

---

#### Sound in the Vocal Tract

- Issues in creating a detailed physical model
  - time varying acoustic system
  - losses due to heat conduction and friction in the walls.
  - radiation of sound at the lips and nostrils
  - softness of the walls
  - nasal coupling
  - excitation of sound in the vocal tract

---

#### Sound Wave Propagation

- using the laws of conservation of mass, momentum and energy, it can be shown that sound wave propagation in a lossless tube satisfies the equations:

\[
\frac{\partial p}{\partial x} = \rho \left( \frac{\partial (uA)}{\partial t} \right)
\]

\[
\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \left( \frac{\partial (pA)}{\partial t} \right)
\]

where:

- \(p = p(x,t)\) = sound pressure in the tube at position \(x\) and time \(t\)
- \(u = u(x,t)\) = 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, 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 \(A(x,t) \Rightarrow\) 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)

\[
\begin{align*}
\frac{\partial p}{\partial x} &= \frac{p \partial u}{\partial t} \\
- \frac{\partial u}{\partial x} &= A \frac{\partial p}{\partial t} \\
\end{align*}
\]

**Acoustic-Electrical Analogs**

- Acoustic
  - \( p = \text{pressure} \)
  - \( u = \text{volume velocity} \)
- Electrical
  - \( v = \text{voltage} \)
  - \( i = \text{current} \)
  - \( L = \text{inductance} \)
  - \( C = \text{capacitance} \)

**Acoustic inductance**

- \( L = \frac{1}{i} \)

**Acoustic capacitance**

- \( U_{ac} = \frac{1}{C} \)

**Overall Transfer Function**

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

\[
U(l, t) = U_0(\Omega)e^{i\omega t}
\]

- \( \Omega = \frac{2\pi f}{c} \)

\[
\frac{U(l, \Omega)}{U_0(\Omega)} = \frac{1}{\cos(\Omega l / c)}
\]

**Frequency response of uniform tube in terms of volume velocities**

\[
\text{formants of uniform tube}
\]

\[
\Omega = \frac{2\pi f}{c} \quad \text{in} \quad \text{Hz}
\]

\[
\text{formants of uniform tube}
\]

\[
\text{formants of uniform tube}
\]

**Traveling Wave Solution**

- Assume traveling wave solution

\[
\begin{align*}
&u(x,t) = u'(t-\frac{x}{c}) - u'(t+\frac{x}{c}) \\
&p(x,t) = \frac{\rho c}{A} u'(t-\frac{x}{c}) + u'(t+\frac{x}{c}) \\
&- u'(t-\frac{x}{c}) \text{wave travelling forward} \\
&- u'(t+\frac{x}{c}) \text{wave travelling backward}
\end{align*}
\]

- \( u'(t-\frac{x}{c}) \) boundary conditions at the glottis and at the lips gives:

\[
\begin{align*}
u(0,t) &= U_0(\Omega)e^{i\omega t} \\
p(0,t) &= 0
\end{align*}
\]

- \( p(t,t) = 0 \)

- \( u'(t-\frac{x}{c}) = K' e^{i(\omega t - \frac{x}{c})} \)

- \( u'(t+\frac{x}{c}) = K' e^{-i(\omega t + \frac{x}{c})} \)

**Traveling Wave Solution**

- Solve for \( K' \) and \( K'' \)

\[
\begin{align*}
u(0,t) = U_0(\Omega)e^{i\omega t} = K' e^{i\omega t} - K'' e^{-i\omega t} \\
p(0,t) = 0 = \frac{\rho c}{A} \left[ K' e^{i(\omega t + \frac{x}{c})} + K'' e^{-i(\omega t - \frac{x}{c})} \right]
\end{align*}
\]

\[
\begin{align*}
K' &= U_0(\Omega) \left[ e^{i(\omega t - \frac{x}{c})} + e^{-i\omega t} \right] \\
K'' &= \frac{U_0(\Omega)}{1 + e^{i\frac{2x}{c}}}
\end{align*}
\]

**Overall Transfer Function**

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

\[
U(l, t) = U_0(\Omega)e^{i\omega t}
\]

\[
\frac{U(l, \Omega)}{U_0(\Omega)} = \frac{1}{\cos(\Omega l / c)}
\]

- Formants of uniform tube

\[
\Omega = 2\pi f \quad \text{in} \quad \text{Hz}
\]

- Formants of uniform tube

\[
\Omega = 2\pi f \quad \text{in} \quad \text{Hz}
\]

- Formants of uniform tube

\[
\Omega = 2\pi f \quad \text{in} \quad \text{Hz}
\]

- Formants of uniform tube

\[
\Omega = 2\pi f \quad \text{in} \quad \text{Hz}
\]
Hardwalled Tube and Buzzer

Spectrum Slices

The formants are not at the frequencies 500, 1500, 2500, ... Hz. What are some possible sources of error in this experiment?

Summary of Solution of Sound Propagation Equations in the Vocal Tract

Step 1--Basic sound wave propagation equations

\[ \frac{\partial^2 p}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 \]

Step 2--Boundary conditions

At \( x = 0 \) (glottis), sound source

\( p(x=0,t) \rightarrow \) no pressure at lips

\[ \rho \frac{\partial v}{\partial x} = 0 \]

Step 3--Assume travelling waves form of solution

\[ p(x,t) = U_x e^{\lambda x - \omega t} \]

Step 4--Simplified boundary conditions

\[ p(x=0,t) = 0 \]

Step 5--Simplified forward and backward waves

\[ u(x,t) = U_x e^{\lambda x - \omega t} \]

Step 6--Simplified forward and backward waves

\[ A_x = \frac{\omega^2 - c^2 \lambda^2}{c^2} \]

Step 7--Determine \( A_x \) and \( B_x \), and solve for \( u(x,t) \)

\[ u(x,t) = \frac{\omega^2 - c^2 \lambda^2}{c^2} \]

Step 8--Solve for \( u(x,t) \)

\[ u(x,t) = \frac{\omega^2 - c^2 \lambda^2}{c^2} \]

Step 9--Solve for transfer function of volume velocity

\[ \frac{U_x}{p(x,t)} = \frac{1}{c^2} \frac{\omega^2 - c^2 \lambda^2}{c^2} \]

Step 10--Determine tube resonance

\[ \omega = n \cdot \frac{c}{l} \]

Step 11--Interpretation of travelling wave solution

\[ A_x = \frac{\omega^2 - c^2 \lambda^2}{c^2} \]

Frequency Domain Representation

- we can alternatively express \( p(x,t) \) and \( u(x,t) \) as

\[ p(x,t) = B_x \sin(\Omega(\frac{x}{c}) - \omega t) \]

\[ u(x,t) = \cos(\Omega(\frac{x}{c}) - \omega t) \]

where

\[ Z = \frac{pc}{A} \]

is characteristic acoustic impedance of tube

Traveling Wave Solution

- using s-transform notation \( s = \sigma + j\Omega \) we get

\[ V_s(s) = \frac{U_l(s)}{U_0(s)} = \frac{2e^{-\sigma l/c}}{1 + e^{2\sigma l/c}} \]

Alternative Wave Equation Solution (avoids solution for forward and backward travelling waves)

- express \( p(x,t) \) and \( u(x,t) \) as complex transfer functions of the form:

\[ p(x,t) = P(x) e^{\sigma x} \]

\[ u(x,t) = U(x) e^{\sigma x} \]

- inserting these representations back into the wave equation gives:

\[ \frac{d^2 P}{dx^2} - Z U = 0 \]

\[ \frac{dU}{dx} = \frac{P}{Z} \]

- can show that solutions of wave equation have form:

\[ P(x) = A e^{-\sigma x} + B e^{\sigma x} \]

\[ U(x) = C e^{-\sigma x} + D e^{\sigma x} \]

- by using boundary conditions, \( P(l,0) = 0, U(l,0) = U_l(0) \), can solve for \( A, B, C, D \)
Effects of Losses in VT

- several types of losses to be considered
  - viscous friction at the walls of the tube
  - heat conduction through the walls of the tube
  - vibration of the tube walls
- loss will change the frequency response of the tube
- consider first wall vibrations
- assumption: walls are elastic => cross-sectional area of the tube will change with pressure in the tube
- assume walls are locally reacting => vibration of the tube walls
- heat conduction through the walls of the tube
- viscous friction at the walls of the tube

\[ A(x,t) = A_0(x,t) + \delta A(x,t) \]

Fig. 3.16 Illustration of the effects of wall vibration.

Losses in Frequency Domain

- consider a time-invariant constant area tube excited by a complex volume velocity source
- 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:
- \( V_j(\Omega) = \frac{U(\Omega)}{\bar{U}_0(\Omega)} \)

Effects of Loss on FR

- there is a differential equation relationship between area perturbation \( \delta A(x,t) \) and the pressure variation, \( p(x,t) \) of the form:
  \[ m_w \frac{d^2(\delta A)}{dt^2} + b_w \frac{d(\delta A)}{dt} + k_w (\delta A) = p(x,t) \]
  where
  - \( m_w(x) \) - mass/unit length of the vocal tract wall
  - \( b_w(x) \) - damping/unit length of the vocal tract wall
  - \( k_w(x) \) - stiffness/unit length of the vocal tract wall
- neglecting second order terms in \( u/A \) and \( pA \), the basic wave equations become
  \[ \frac{\partial^2 u}{\partial t^2} + \frac{1}{\rho} \frac{\partial (pA)}{\partial x} = \frac{\partial}{\partial x} \left( \frac{\partial u}{\partial x} \right) \]

Effects of Loss

- there is a differential equation relationship between area perturbation \( \delta A(x,t) \) and the pressure variation, \( p(x,t) \) of the form:
- \( m_w \frac{d^2(\delta A)}{dt^2} + b_w \frac{d(\delta A)}{dt} + k_w (\delta A) = p(x,t) \)
  where
- \( m_w \) - mass/unit length of the vocal tract wall
- \( b_w \) - damping/unit length of the vocal tract wall
- \( k_w \) - stiffness/unit length of the vocal tract wall
- neglecting second order terms in \( u/A \) and \( pA \), the basic wave equations become
- \[ \frac{\partial^2 u}{\partial t^2} + \frac{1}{\rho} \frac{\partial (pA)}{\partial x} = \frac{\partial}{\partial x} \left( \frac{\partial u}{\partial x} \right) \]

Friction and Thermal Conduction Losses

- Viscous friction can be accounted for in the frequency domain by including a real, frequency dependent term in the expression for the acoustic impedance, \( Z \), of the form:
  \[ Z(\Omega) = \frac{Z_0(\Omega)}{\rho A(x) \sqrt{\gamma \rho/\rho_{\text{air}}} + \rho_{\text{air}} A(x)} \]
  \( S(x) \) is the circumference of the tube in cm
  \( \rho \) is the density of air (0.00114 gm/cm^3)
  \( \mu \) is the coefficient of friction (0.000186)
- Heat conduction accounted for by adding a real frequency dependent term to the acoustic admittance, of the form:
  \[ Y(\Omega) = \frac{Z(\Omega)(\gamma-1)}{\rho c_s \sqrt{\gamma \rho/\rho_{\text{air}}} + \rho_{\text{air}} A(x)} \]
  \( c_s \) is the specific heat at constant pressure (0.24)
  \( \gamma \) is the ratio of specific heat at constant pressure to that at constant volume (1.4)
  \( \beta \) is the coefficient of heat conduction (0.000055)
Effects of Radiation at 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.](image)

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:

$$P(l,t) = Z_l(\Omega) U(l,t)$$

where

$$Z_l(\Omega) = \frac{j\Omega L \rho_s}{R_s + j\Omega L}$$

- this radiation load is the equivalent of a parallel connection of a radiation resistance, $R_s$, and a radiation inductance, $L_s$. Suitable values for these components are:

$$R_s = \frac{128}{9\pi^2} \quad L_s = \frac{8\pi^2}{3\pi c}$$

Behavior of Radiation Load

- radiation losses most significant at higher frequencies

$$Z_l(\Omega) = \frac{j\Omega L \rho_s}{R_s + j\Omega L}$$

- at low frequencies, $Z_l(\Omega) = 0$ (short circuit termination) => old solution
- at mid-range frequencies, $Z_l(\Omega) = j\Omega L$ (inductive load) => $R_s \gg \Omega L$
- at higher frequencies, $Z_l(\Omega) = R_s$ (resistive load) => $\Omega L \gg R_s$

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:

$$V_f(\Omega) = \frac{U_f(\Omega)}{U_g(\Omega)}$$

- assuming input at glottis of form:

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

Vocal Tract Transfer Function

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

$$H_f(\Omega) = \frac{P(l,t)}{U_g(\Omega)} = \frac{P(l,t)}{U(l,t)} \frac{U(l,t)}{U_g(\Omega)} = Z_l(\Omega) \cdot V_f(\Omega)$$

- 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 ($\frac{dA}{dt}=0$)
  - measure $U(l,\Omega)/U_g(\Omega)$ for the vowels /AA/, /EH/, /IY/, /UW/
Area Function from X-Ray Photographs


---

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

---

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, thermal losses, and radiation losses.

---

Nasal Coupling Effects

- 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 frequencies 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

---

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).
Vocal Cord Simulation

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

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.

Glottal Excitation Model

- Vocal tract acts as a load on the vocal cord oscillator.
- Time varying glottal resistance and inductance both functions of $1/\sqrt{\Delta t}$ when $A_g=0$ (total closure), impedance is infinite and volume velocity is zero.

Rosenberg Glottal Pulse and Spectrum

- $g(t) = 0.5 \left[ 1 - \cos(\pi n/N_g) \right]$ for $0 \leq n \leq N_g - 1$
- $g(t) = 0$ otherwise

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.

Source-System Model

- Pitch, Voiced/Unvoiced, Amplitude
- Formants, Vocal Tract Area Functions, Articulatory Parameters
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 time-varying linear system

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), the approximation error for modeling the vocal tract goes to zero

How do we use the lossless tube model to solve for various vocal tract transfer functions

Concatenated Tube Models

Wave Propagation in Lossless Tubes

• since each individual tube is lossless, can solve the basic wave equation for each individual tube, giving:
• for \( k^{th} \) tube:
  \[
  \rho_i(x,t) = \frac{\rho}{A} \left[ u_i^+(t-x/c) + u_i^-(t+x/c) \right], \quad 0 \leq x \leq \ell_i
  \]
  \[
  u_i(x,t) = u_i^-(t-x/c) - u_i^+(t+x/c), \quad 0 \leq x \leq \ell_i
  \]
  where \( x \) is the distance measured from the left-hand end of the \( k^{th} \) tube \((0 \leq x \leq \ell_i)\) and \( u_i^+() \) and \( u_i^-() \) are positive-going and negative-going traveling waves in the \( k^{th} \) tube

Wave Propagation in Lossless Tubes

• boundary conditions at edges of adjacent tubes state that both pressure and volume velocity must be continuous in both time and space
• consider junction between \( k^{th} \) and \((k+1)^{th}\) tubes

Lossless Tube Junction

• at the junction between \( k^{th} \) and \((k+1)^{th}\) tubes, we get:
  \[
  \rho_i(t,\ell_i) - \rho_{i+1}(0,0)
  \]
  \[
  u_i(t,\ell_i) - u_{i+1}(0,0)
  \]
• substituting from the previous set of equations, we get:
  \[
  \frac{A_{i+1}}{A_i} \left[ u_i^+(t-\ell_i) + u_i^-(t+\ell_i) \right] - u_{i+1}^- + u_{i+1}^+(t)
  \]
  \[
  u_i^+(t-\ell_i) - u_i^-(t-\ell_i) = u_{i+1}^- - u_{i+1}^+(t)
  \]
  where \( \ell_i = \ell / c \) is the time for a wave to travel the length of the \( k^{th} \) tube
• at the junction between tubes, part of the positive going wave is propagated to the right while part is reflected back to the left
• similarly part of the negative going wave is propagated to the left while part is reflected back to the right
Lossless Tube Model

- solve for $u_{in}(t)$ and $u_{out}(t)$ in terms of $u_{in}(t)$ and $u_{out}(t)$ to see how forward and reverse travelling waves propagate

\[
\begin{align*}
\frac{dA}{A} - \frac{dA}{A} & \quad = \frac{A_0 - A}{A_0 + A} \quad dA(t) \\
\frac{dA}{A} + \frac{dA}{A} & \quad = \frac{A_0 + A}{A_0 - A} \quad dA(t)
\end{align*}
\]

- the quantity

\[
\tau = \left[ \frac{A_0 - A}{A_0 + A} \right]
\]

- amount of $u_{out}(t)$ that is reflected at the junction

- $\tau$ is called the reflection coefficient for the $k$th junction, with

\[
\begin{align*}
u_{in}(t) &= (1 + \rho \tau) u_{in}(t) - \rho \tau u_{out}(t) \\
u_{out}(t) &= -(1 - \rho \tau) u_{in}(t) + \rho \tau u_{out}(t)
\end{align*}
\]

Boundary Conditions

- for a 5-tube model there are 5 sets of forward and backward delays, 4 junctions (each characterized by a reflection coefficient), and a set of boundary conditions at the lips and glottis

- assume $N$ sections (indexed 1 to $N$) starting at the glottis

- want to relate pressure ($p_N(l_N,t)$) and volume velocity ($u_N(l_N,t)$) at output of $N$th tube to the radiated pressure and volume velocity (at the lips)

\[
\begin{align*}
p_N(l_N,t) &= u_N(l_N,t) \\
u_N(l_N,t) &= \frac{A_0 - A}{A_0 + A} u_N(l_N,t)
\end{align*}
\]

Termination at Lips

\[
u_N(t + \tau_N) = -\rho \tau u_N(t - \tau_N)
\]

- output volume velocity at the lips is

\[
u_N(t) = u_N(t - \tau_N) - u_N(t + \tau_N) = (1 + \rho \tau) u_N(t - \tau_N)
\]

Adding the Excitation Source

- assume that the excitation source and the vocal tract are linearly separable, then at the first tube we have the boundary relation:

\[
u_1(t) = \frac{1}{Z_0} \left[ u_1(t) + u_1(t) \right] = \frac{1}{Z_0} \left[ u_1(t) + u_1(t) \right]
\]

\[
u_1(t) = -\frac{1}{Z_0} \left[ \rho \tau u_1(t) + \tau \rho u_1(t) \right]
\]

- where the reflection coefficient, $\tau$, is

\[
\begin{align*}
\tau & = \left[ \frac{Z_0 - \rho}{Z_0 + \rho} \right] \\
\tau & \leq 1
\end{align*}
\]
Lossless Two Tube Model

- Volume velocity at lips is \( u_L(t) = u_L(t) \)
- Transfer function from glottis to lips is

\[
\frac{V_G(t)}{u_L(t)} = \frac{0.5[1 + \tau_g] + \tau_\ell}{1 + \tau_g e^{-\sigma G/\ell} + \tau_\ell e^{-\sigma L/\ell}}
\]

- Note total delay of \((\tau_g + \tau_\ell)\) 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 VT Model

Fig. 3.9F: Complete flow diagram of a two-tube model.
Lossless VT Sections

\[ r = \frac{k}{A_{k+1} + A_k} \]

Glottis:

\[ Z_{G} = \frac{1+rG}{2} \]

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 \( \Delta x/N \) where \( \Delta x \) is the overall length of the vocal tract

\[ V_s(t) = e^{-\alpha N t} \sum_{k=0}^{\infty} a_k e^{-2k\tau} \]

Signal Flow Graph Model

• \( \alpha \) is the delay of \( N \) cycles, the time to traverse the length of one tube

\[ V_s(t) = e^{-\alpha N t} \sum_{k=0}^{\infty} a_k e^{-2k\tau} \]

\[ V_s(t) = e^{-\alpha N t} \sum_{k=0}^{\infty} a_k e^{-2k\tau} \]

\[ V_s(t) = e^{-\alpha N t} \sum_{k=0}^{\infty} a_k e^{-2k\tau} \]
Signal Flow Graphs

Analog model

D-T equivalent for bandlimited inputs
D-T equivalent without half-sample delays

Transfer Function of Lossless Tube Model

\[ V(z) = \frac{D(z)}{D(z) + 1} \]

Example:
\[ D(z) = 1 + z^{-1} + z^{-2} \]

\[ r_n = 1 \Rightarrow A_{in} = \infty \] (infinite tube at lips)
\[ r_n = 0.714 \Rightarrow A_{in} = 28 \text{ cm}^2 \]
**Summary of Lossless Tube Models**

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

2. Solve the wave equation for the $k^{th}$ tube:
   \[
   p_k(x,t) = \frac{pc}{A_k} \left[ u_k'(t-x/c) + u_k(t+x/c) \right], \quad 0 \leq x \leq L_k
   \]
   \[
   u_k(x,t) = \left[ u_k'(t-x/c) - u_k(t+x/c) \right], \quad 0 \leq x \leq L_k
   \]

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

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

6. For an $N$-tube model there are $(N-1)$ junctions with reflection coefficients
   - there are boundary conditions at the lips and glottis

7. Need to relate $(p_l, u_l)$ and $(p_u, u_u)$ to pressure and volume velocity at the lips via:
   \[
   r_{GL} = \frac{pc}{A_l} - \frac{z_l}{A_l + z_l}
   \]
   where $z_l$ is lip impedance, i.e.,
   \[
   P_l(t) = r_{GL} U_l(t) + \frac{z_l}{A_l + z_l} U_l(t)
   \]

8. Boundary condition at the glottis:
   \[
   r_G = \frac{Z_G - \rho c / A_k}{Z_G + \rho c / A_k}
   \]
   where $Z_G$ is glottal impedance

9. For the special case of a 2-tube model we get (Figure 3.37):
Digital Models for Speech
- sound is generated in 3 ways—each yielding a distinctive output
- vocal tract imposes resonance (and anti-resonance) structure so as to produce different speech sounds
- this leads to a "terminal analog" speech model—correct at the terminal points, but without mimicking the physics of speech production

Digital Model Features
- poles of V(z) correspond to vocal tract resonances
- missing zeros of V(z) for nasals
  - modify V(z)
  - increase N to large number to model zeros
- roots of V(z) are at:
  \[ z_k = e^{-\alpha_j T}, \theta_k = 2\pi F_k T \]

Other Synthesis Implementations
- direct form difference equation
- cascade of second order systems

Excitation Model

Glottal Pulse Model
\[ g[n] = 0.5 \left[ 1 - \cos(\pi n / N_1) \right], \quad 0 \leq n \leq N_1 - 1 \]
\[ = \cos(\pi (n - N_1) / N_2), \quad N_1 \leq n \leq N_1 + N_2 - 1 \]
\[ = 0 \quad \text{otherwise} \]
- lowpass filtering effect
Summary of Lecture

- 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
  - examined range of digital speech production/synthesis models