Return to Main

Objectives

Acoustic Models:

- Acoustic Theory
- Lossless Tubes
- Resonances
- Losses
- Lip Radiation
- Nasal Cavity

Lossless Tubes:

- Concatenated Tubes
- Excitation Models
- Two Tube Models
- Three Tube Models
- Transfer Functions

Digital Models:

- Digital Equivalents
- Digital Transfer Functions
- Excitation Models
- Vocoder Model

On-Line Resources:

- Sound Waves in Tubes
- Tube Models
- Cool Edit
- CD-ROM Tutorial

LECTURE 03: SOUND PROPAGATION

- Objectives:
  - Basic properties of lossless tubes
  - Resonant structure of the vocal tract
  - Articulator positions (basic speech sounds) translate to predictable spectral signatures
  - Digital filter-based models of the vocal tract (linear acoustics)
  - Relationship of the parameters of these digital models to speech recognition.

Note that this lecture is based on material in this textbook:

J. Deller, et. al., *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., ISBN: 0-7803-5386-2, 2000.

ECE 8463: FUNDAMENTALS OF SPEECH RECOGNITION

Professor Joseph Picone
Department of Electrical and Computer Engineering
Mississippi State University

email: picone@isip.msstate.edu
phone/fax: 601-325-3149; office: 413 Simrall
URL: http://www.isip.msstate.edu/resources/courses/ece_8463

Modern speech understanding systems merge interdisciplinary technologies from Signal Processing, Pattern Recognition, Natural Language, and Linguistics into a unified statistical framework. These systems, which have applications in a wide range of signal processing problems, represent a revolution in Digital Signal Processing (DSP). Once a field dominated by vector-oriented processors and linear algebra-based mathematics, the current generation of DSP-based systems rely on sophisticated statistical models implemented using a complex software paradigm. Such systems are now capable of understanding continuous speech input for vocabularies of hundreds of thousands of words in operational environments.

In this course, we will explore the core components of modern statistically-based speech recognition systems. We will view speech recognition problem in terms of three tasks: signal modeling, network searching, and language understanding. We will conclude our discussion with an overview of state-of-the-art systems, and a review of available resources to support further research and technology development.

Tar files containing a compilation of all the notes are available. However, these files are large and will require a substantial amount of time to download. A tar file of the html version of the notes is available here. These were generated using wget:

```
wget -np -k -m http://www.isip.msstate.edu/publications/courses/ece_8463/lectures/current
```

A pdf file containing the entire set of lecture notes is available here. These were generated using Adobe Acrobat.

Questions or comments about the material presented here can be directed to help@isip.msstate.edu.

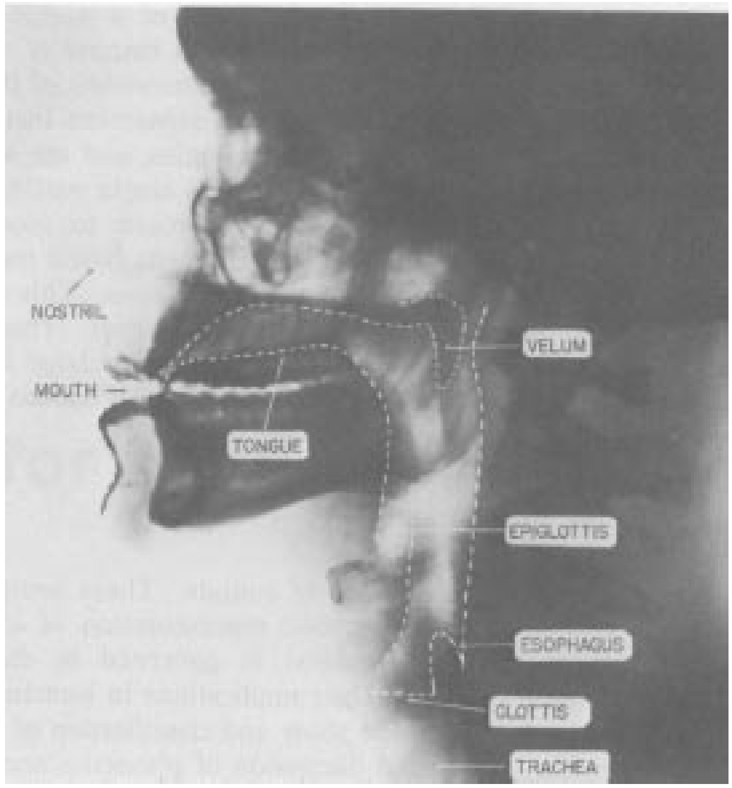
LECTURE 03: SOUND PROPAGATION

- Objectives:
  - Basic properties of lossless tubes
  - Resonant structure of the vocal tract
  - Articulator positions (basic speech sounds) translate to predictable spectral signatures
  - Digital filter-based models of the vocal tract (linear acoustics)
  - Relationship of the parameters of these digital models to speech recognition.

Note that this lecture is based on material in this textbook:

J. Deller, et. al., *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., ISBN: 0-7803-5386-2, 2000.

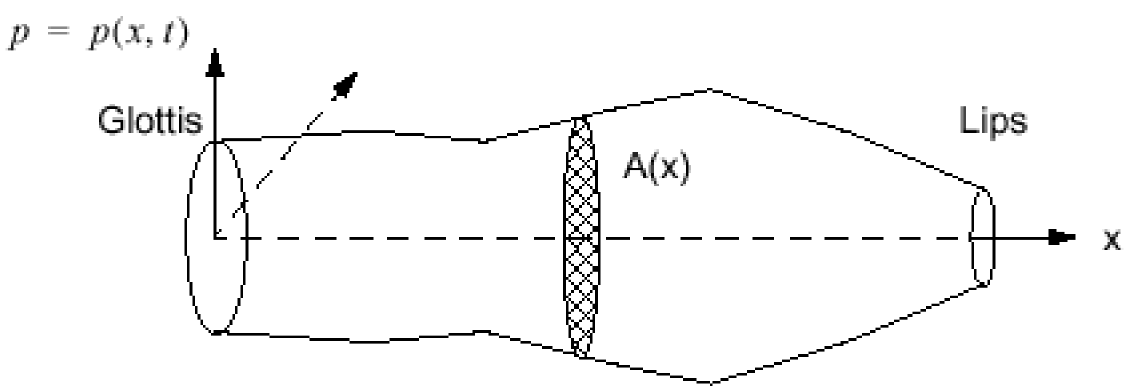
SOUND PROPAGATION



A detailed acoustic theory must consider the effects of the following:

- Time variation of the vocal tract shape
- Losses due to heat conduction and viscous friction at the vocal tract walls
- Softness of the vocal tract walls
- Radiation of sound at the lips
- Nasal coupling
- Excitation of sound in the vocal tract

Let us begin by considering a simple case of a lossless tube:



UNIFORM LOSSLESS TUBE

For frequencies that are long compared to the dimensions of the vocal tract (less than about 4000 Hz, which implies a wavelength of 8.5 cm), sound waves satisfy the following pair of equations:

$$\rho\frac{\partial(u/A)}{\partial t} + grad'''' p = 0 \qquad \text{or} \qquad -\frac{\partial p}{\partial x} = \rho\frac{\partial(u/A)}{\partial t}$$

$$\frac{1}{\rho c^2}\frac{\partial p}{\partial t} + \frac{\partial A}{\partial t} + div\ u = 0 \qquad \qquad -\frac{\partial u}{\partial x} = \frac{1}{\rho c^2}\frac{\partial(pA)}{\partial t} + \frac{\partial A}{\partial t}$$

where

- $p = p(x, t)$ is the variation of the sound pressure in the tube
- $u = u(x, t)$ is the variation in the volume velocity
- ρ is the density of air in the tube (1.2 mg/cc)
- c is the velocity of sound (35000 cm/s)
- $A = A(x, t)$ is the area function (about 17.5 cm long)

Uniform Lossless Tube

If $A(x, t) = A$, then the above equations reduce to:

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

The solution is a traveling wave:

$$u(x, t) = u^{+}(t - x/c) - u^{-}(t + x/c)$$

$$p(x, t) = \frac{\rho c}{A}[u^{+}(t - x/c) + u^{-}(t + x/c)]$$

which is analogous to a transmission line:

$$-\frac{\partial v}{\partial x} = L\frac{\partial i}{\partial t} \qquad -\frac{\partial i}{\partial x} = C\frac{\partial v}{\partial t}$$

What are the salient features of the lossless transmission line model?

where

Acoustic Quantity	*Analogous Electric Quantity*
p - pressure	v - voltage
u - volume velocity	i - current
ρ/A - acoustic inductance	L - inductance
$A/(\rho c^2)$ - acoustic capacitance	C - capacitance

The sinusoisdal steady state solutions are:

$$p(x,t) = jZ_o \frac{\sin[\Omega(l-x)/c]}{\cos[\Omega l/c]} U_G(\Omega) e^{j\Omega t}$$

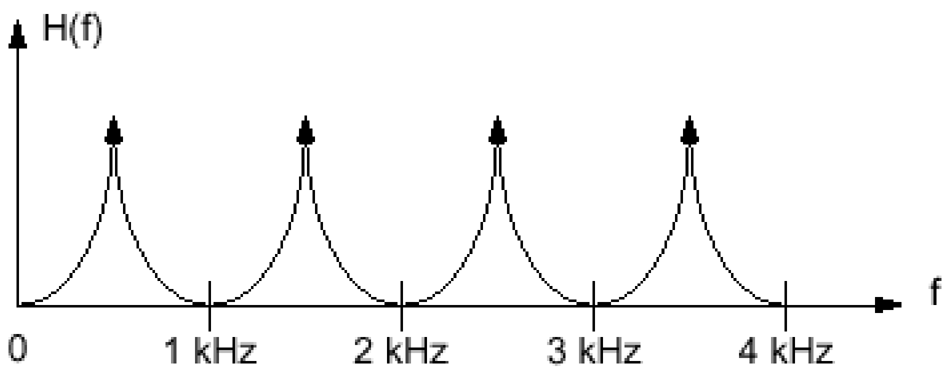
$$u(x,t) = \frac{\cos[\Omega(l-x)/c]}{\cos[\Omega l/c]} U_G(\Omega) e^{j\Omega t}$$

where $Z_0 = \frac{\rho c}{A}$ is the characteristic impedance.

The transfer function is given by:

$$\frac{U(l,\Omega)}{U(0,\Omega)} = \frac{1}{\cos(\Omega l/c)}$$

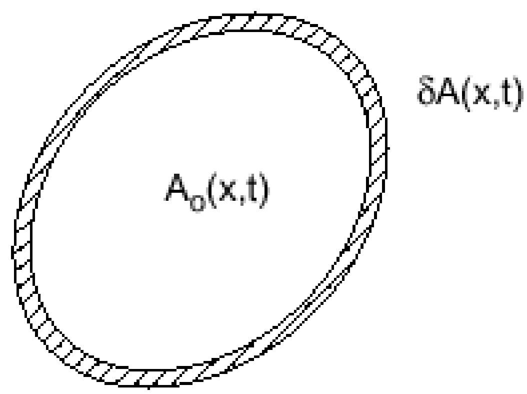
This function has poles located at every $\frac{(2n+1)\pi c}{2l}$. Note that these correspond to the frequencies at which the tube becomes a quarter wavelength: $\left(\frac{\Omega l}{c} = \frac{\pi}{2}\right) \Rightarrow \left(\Omega = \frac{c}{4l}\right)$.



Is this model realistic?

EFFECTS OF LOSSES

What do we predict the effects of yielding walls to be?



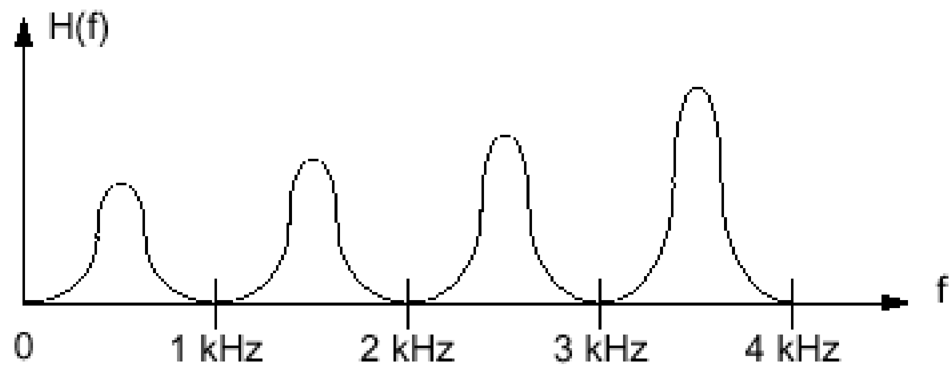
Use perturbation analysis:

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

We can develop a model that relates δA(x,t) to pressure:

$$\frac{m_w d^2(\delta A)}{dt^2} + b_w \frac{d(\delta A)}{dt} + k_w(\delta A) = p(x,t)$$

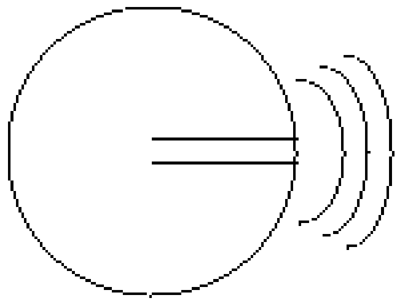
and solve for the new transfer function. But we can easily predict the effect of this:



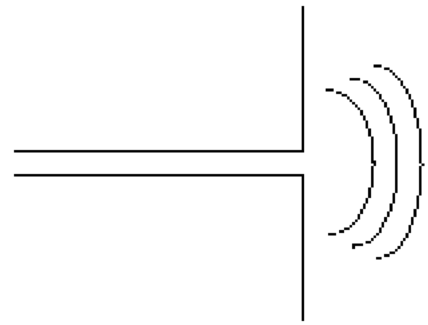
What would you expect to be the effect of friction and thermal losses?

LIP RADIATION

How is the sound pressure wave within the vocal tract coupled into the air?



Radiation from a spherical baffle



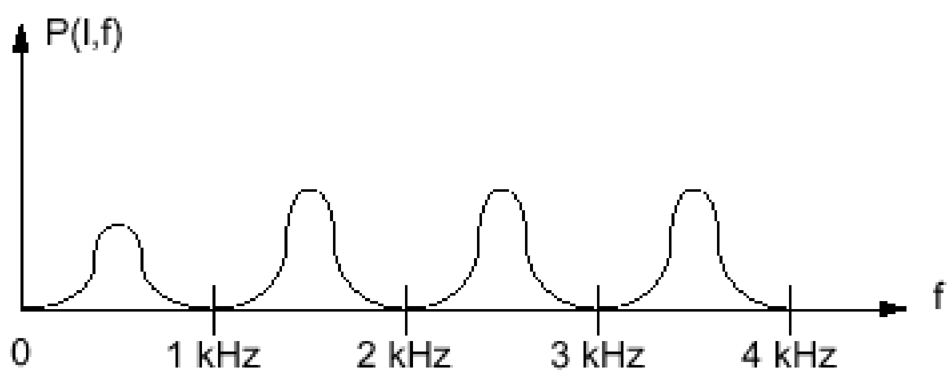
Radiation from an infinite plane baffle

Net effect is to place a complex load on the system:

$$Z_L(\Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r} \quad \text{and} \quad P(l, \Omega) = Z_L(\Omega) U(l, \Omega)$$

where $R_r = 128/9\pi^2$ and $L_r = 8a/3\pi c$, and a is the radius of the opening.

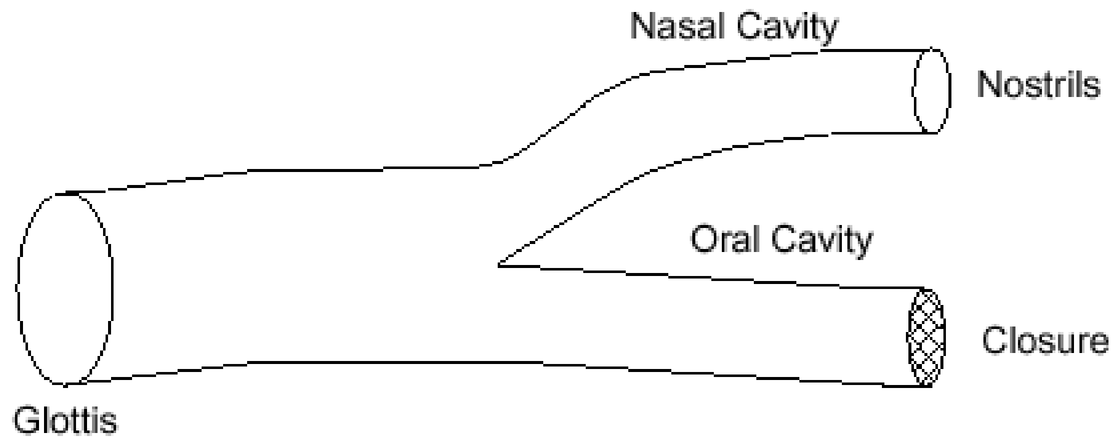
This impedance acts as a short circuit at low frequencies, and an imaginary impedance at high frequencies. The next effect on the volume velocity is to act as a highpass filter and to attenuate low frequencies. Lip radiation introduces a zero in the spectrum at DC and broadens the bandwidths at higher frequencies.



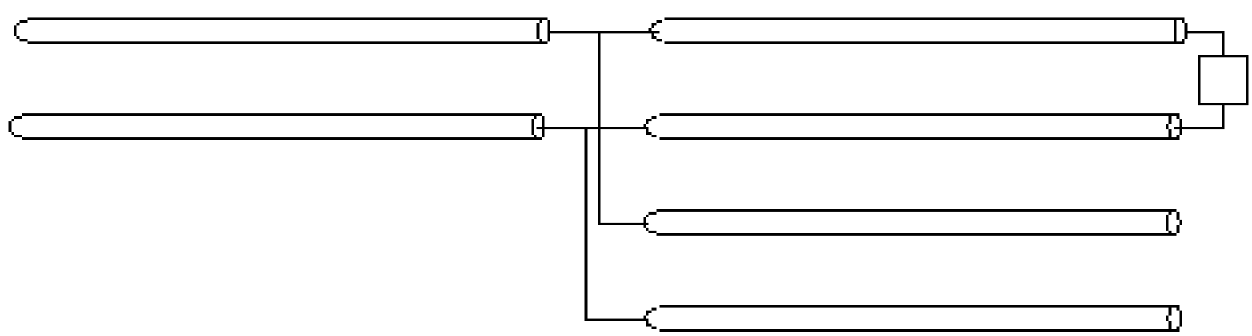
NASAL COUPLING

How is the sound pressure wave within the vocal tract coupled into the air?

We also must worry about the nasal cavity, especially for labial sounds for which the mouth is closed during sound production.



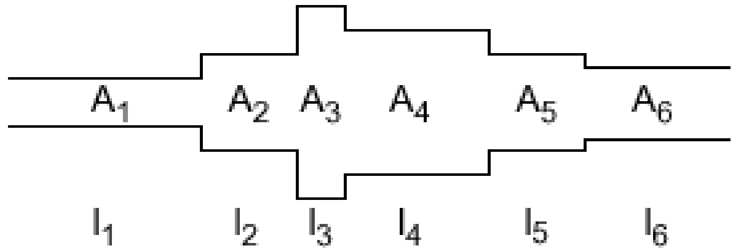
This is the equivalent of placing a transmission line in parallel with the vocal tract (oral cavity). What will the effect be?



The net effect is to produce a zero in the spectrum at about 1 kHz. As a result, nasal sounds (such as “m” and “n” in American English) have very little high frequency energy.

PIECEWISE LINEAR APPROXIMATIONS
FOR THE VOCAL TRACT

Consider the following approximation to the vocal tract area function:



Recall,

$$p_k(x,t) = \frac{\rho c}{A_k}[u_k^+(t - x/c) + u_k^-(t + x/c)]$$

$$u(x,t) = u_k^+(t - x/c) - u_k^-(t + x/c)$$

For the k[th] section, if we apply the boundary conditions:

$$p_k(l_k, t) = p_{k+1}(0, t)$$

$$u_k(l_k, t) = u_{k+1}(0, t)$$

We can combine these two equations to show:

$$u_{k+1}^+(t) = \left[\frac{2A_{k+1}}{A_{k+1} + A_k}\right]u_k^+(t - \tau_k) + \left[\frac{A_{k+1} - A_k}{A_{k+1} + A_k}\right]u_{k+1}^-(t)$$

where $\tau_k = l_k / c$.

We can define a reflection coefficient for the k[th] junction:

$$r_k = \frac{u_{k+1}^+(t)}{u_{k+1}^-(t)} = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}$$

It is easy to show that the reflection coefficients are bounded: $-1 \le r_k \le 1$.

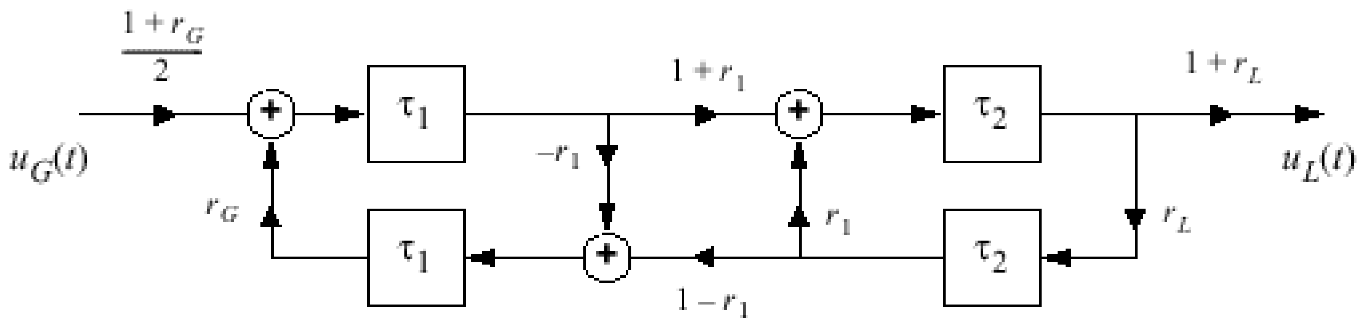
The velocity can be expressed in terms of the reflection coefficients:

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

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

Ultimately, we will relate $\{r_k\}$ to a discrete model of the velocity profile.

Consider a two tube approximation to the vocal tract:

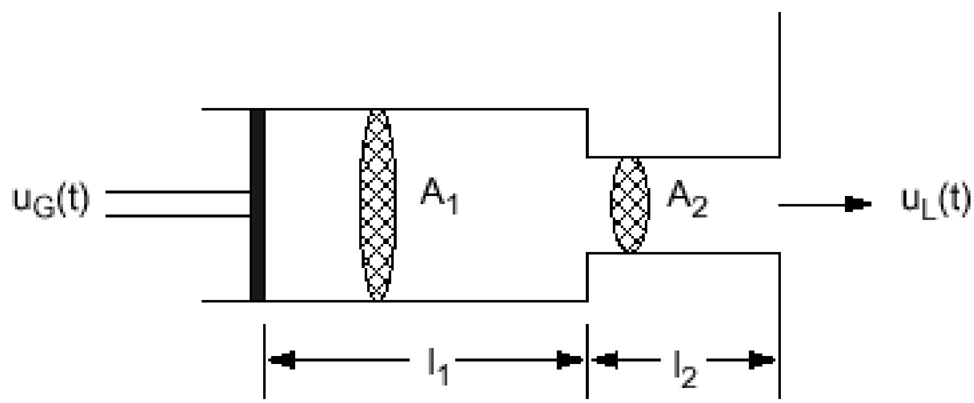


The frequency response of this system is:

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

What does this tell us about the frequency response?

If we consider the case $r_G = r_L = 1$:



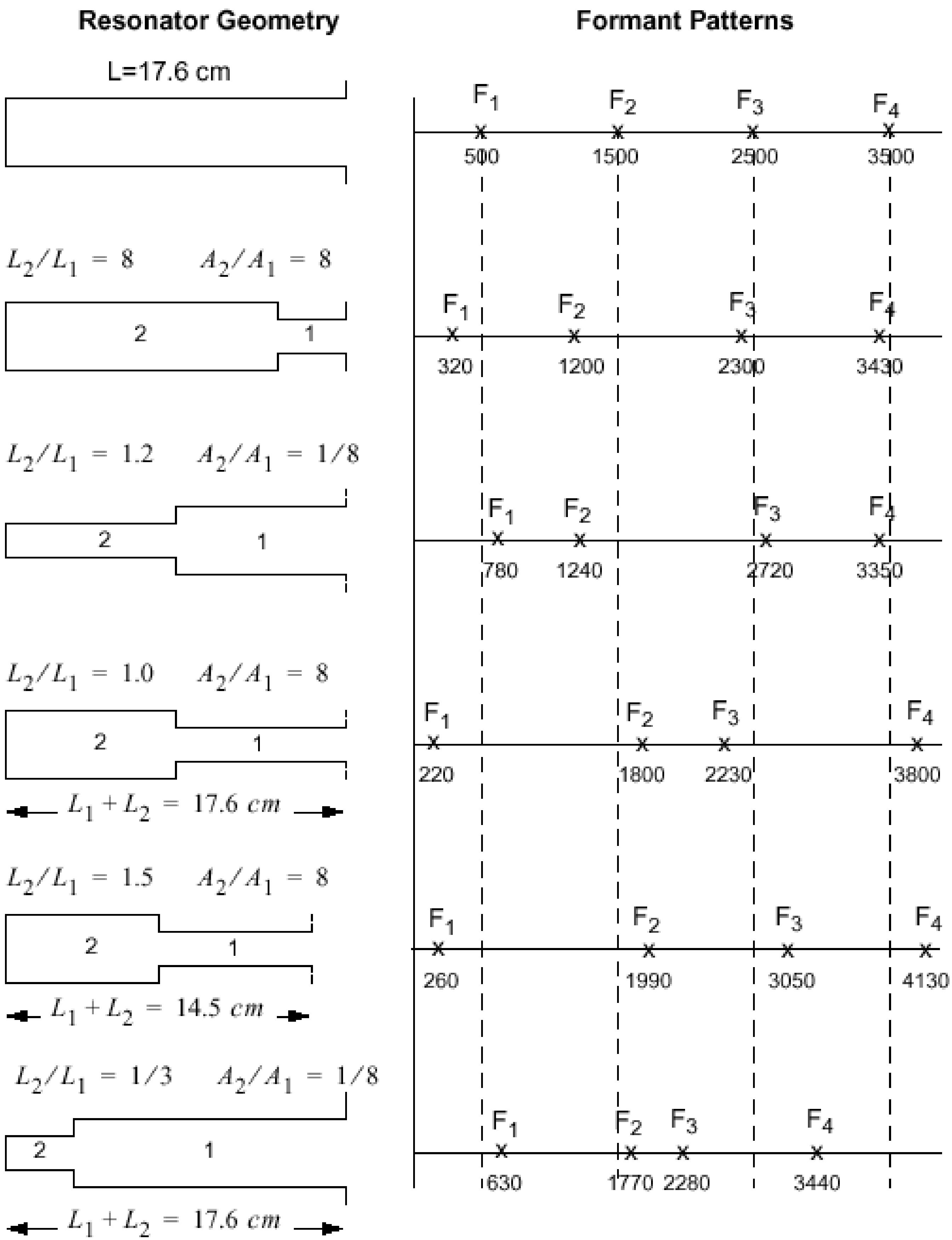
For this system, the poles are located at values that satisfy the equation:

$$\frac{A_1}{A_2}\tan(\Omega\tau_2) = \cot(\Omega\tau_1)$$

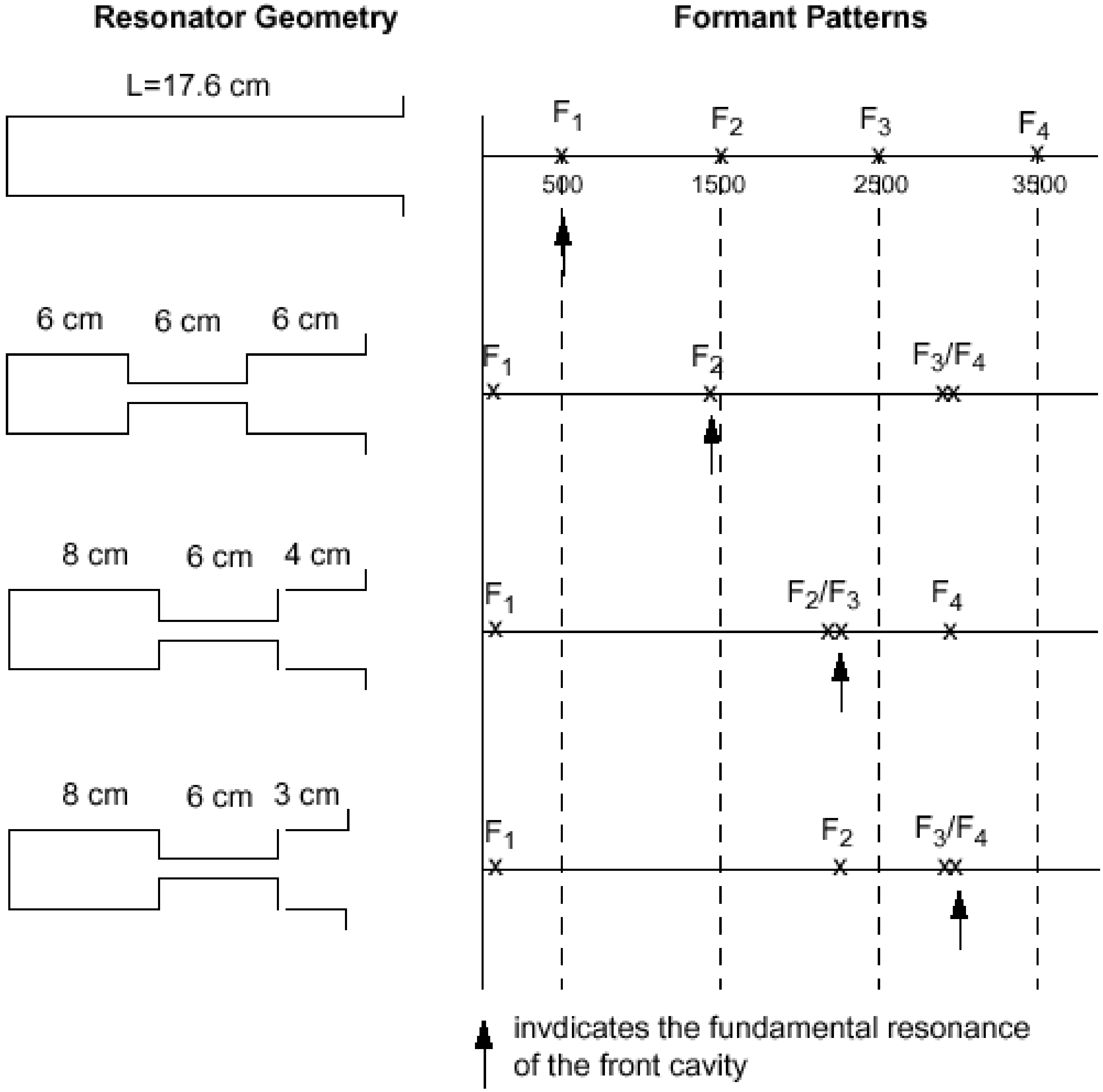
How does this compare to a single lossless tube?

Poles must be found through numerical analysis - nonlinear equation.

TWO TUBE MODELS



THREE TUBE MODELS



TRANSFER FUNCTION OF THE LOSSLESS TUBE MODEL

Recall, $V(\Omega) = \frac{U_L(\Omega)}{U_G(\Omega)}$. In the discrete domain, we can write: $V(z) = \frac{U_L(z)}{U_G(z)}$.

Following our derivation of the wave equation, we can express the transfer function for a lossless tube as follows:

$$\boldsymbol{U}_k = \boldsymbol{Q}_k \boldsymbol{U}_{k-1}$$

where

$$\boldsymbol{U}_k = \begin{bmatrix} U_k^+(z) \\ U_k^-(z) \end{bmatrix} \quad \text{and} \quad \boldsymbol{Q}_k = \begin{bmatrix} \frac{z^{1/2}}{1+r_k} & \frac{-r_k z^{1/2}}{1+r_k} \\ \frac{-r_k z^{1/2}}{1+r_k} & \frac{z^{-1/2}}{1+r_k} \end{bmatrix}$$

The combined transfer function is a product of these matrices. The net result is a transfer function that can be expressed as:

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

where

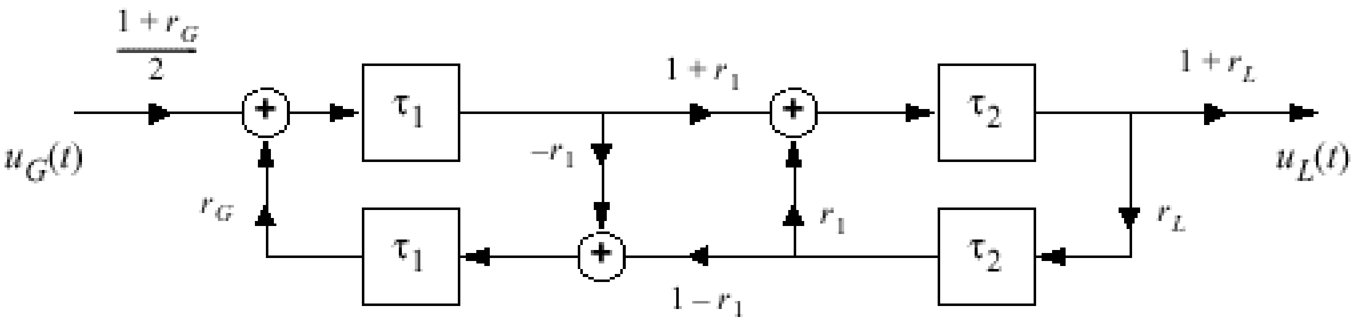
$$D(z) = \begin{bmatrix} 1 & -r_G \end{bmatrix} \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_N \\ -r_N z^{-N} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We can write $D(z)$ in a simpler form:

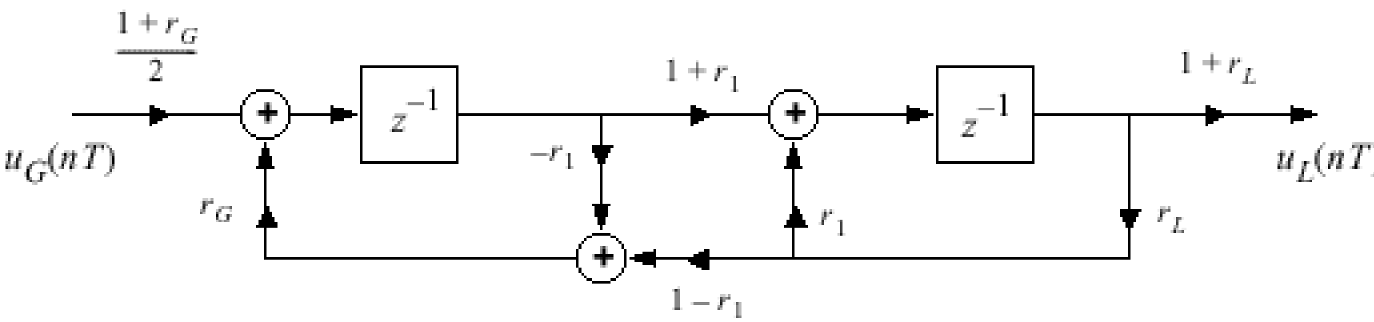
$$D(z) = 1 - \sum_{k=1}^{N} \alpha_k z^{-k}$$

Why is this important?

Recall our concatenated lossless tube model:



We can approximate this as a digital filter using the sampling theorem:



The transfer function of an N-tube model is:

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

where

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

We can compute $D(z)$ recursively:

$$D_o(z) = 1$$

$$D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}) \qquad k = 1, 2, \ldots, N$$

$$D(z) = D_N(z)$$

ALTERNATE DIGITAL FILTER IMPLEMENTATIONS USING DIGITAL RESONATORS

Note that for $D(z)$ to have real coefficients, zeros must occur in complex conjugate pairs. We can transform zeros in the Laplace domain:

$$s_k, s_k^* = -\sigma \pm j2\pi F_k$$

The corresponding complex conjugate poles in the discrete-domain are:

$$\begin{aligned} z_k, z_k^* &= e^{-\sigma_k T} e^{\pm j2\pi F_k T} \\ &= e^{-\sigma_k T}\cos(2\pi F_k T) \pm j e^{-\sigma_k T}\sin(2\pi F_k T) \end{aligned}$$

Note that magnitude of the pole in the z-plane is related to the bandwidth.

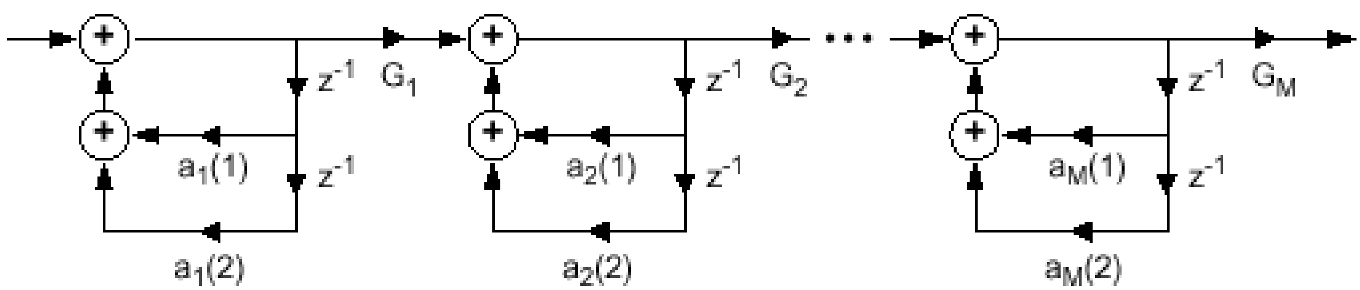
We can write a transfer function as a product of these poles:

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

where

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

This is an all-pole filter. It can be realized using a number of structures:
Under what conditions is this filter stable?



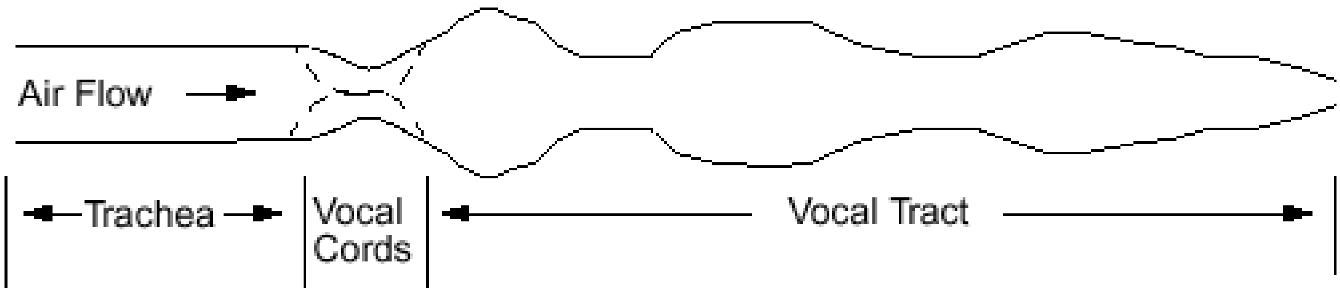
where,

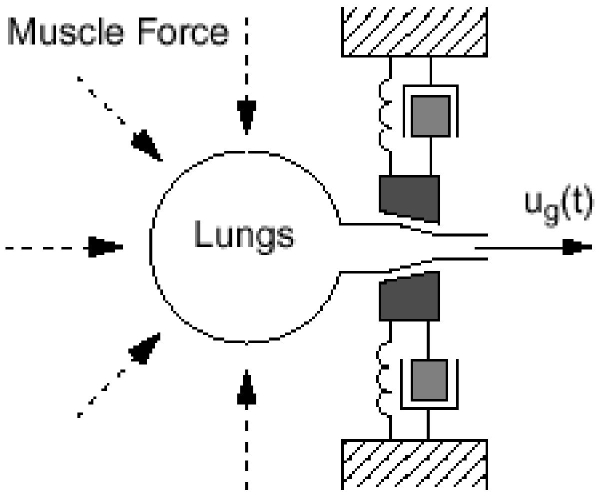
$$V_k(z) = \frac{G_M}{1 - a_k(1)z^{-1} - a_k(2)z^{-2}}$$

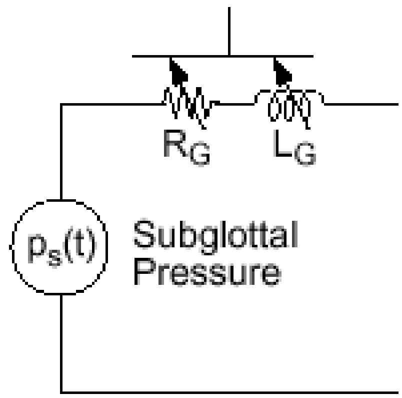
$$a_k(1) = 2|z_k|\cos(2\pi F_k T) \quad a_k(2) = -|z_k|^2 \quad G_k = 1 - 2|z_k|\cos(2\pi F_k T) + |z_k|^2$$

EXCITATION MODELS

How do we couple energy into the vocal tract?







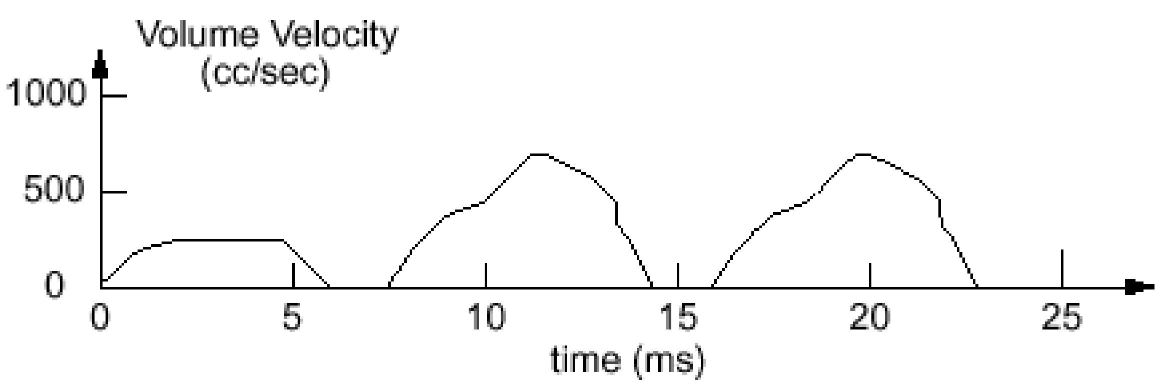
The glottal impedance can be approximated by:

$$Z_G = R_G + j\Omega L_G$$

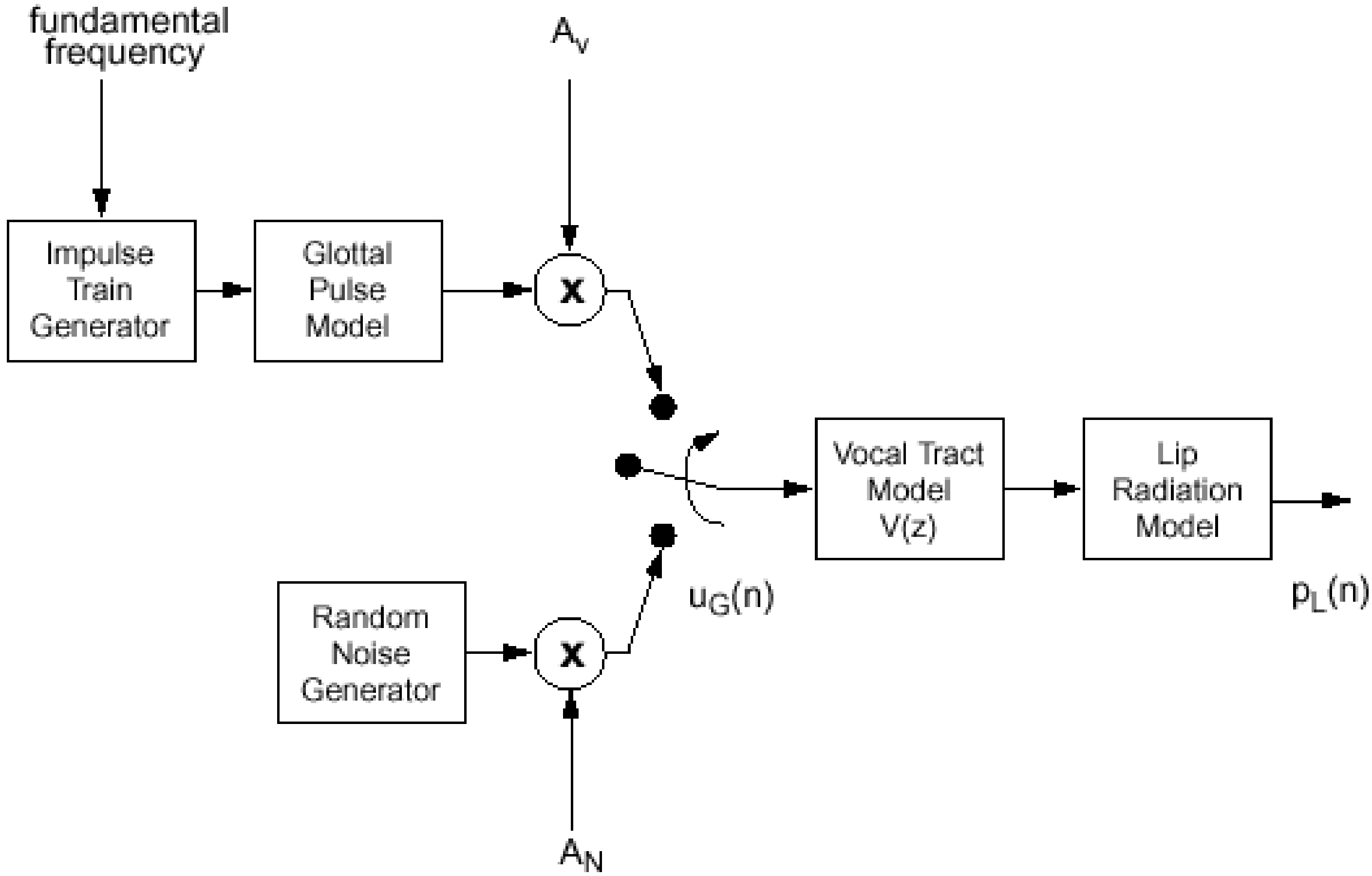
The boundary condition for the volume velocity is:

$$U(0, \Omega) = U_G(\Omega) - P(0, \Omega)/Z_G(\Omega)$$

For voiced sounds, the glottal volume velocity looks something like this:



THE VOCODER (COMPLETE) DIGITAL MODEL



Notes:

- Sample frequency is typically 8 kHz to 16 kHz
- Frame duration is typically 10 msec to 20 msec
- Window duration is typically 30 msec
- Fundamental frequency ranges from 50 Hz to 500 Hz
- Three resonant frequencies are usually found within 4 kHz bandwidth
- Some sounds, such as sibilants ("s") have extremely high bandwidths

Questions:

What does the overall spectrum look like?

What happened to the nasal cavity?

What is the form of V(z)?

Lecture 2

Sound Waves in a Tube

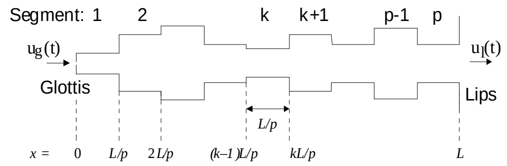
- Derive a theoretical model of how sound waves are affected by the vocal tract
- Describe a model for lip radiation
- Describe a model for the pulsating glottal waveform during voiced speech
- Assemble the components of a simple speech synthesiser

Appendix (not examinable)

- The physics of 1-dimensional sound waves

Multi-Tube Model of Vocal Tract

We model the vocal tract as a tube that has p segments:



u_g and u_l are the volume flows of air at the glottis and lips respectively (measured in litres per second).

Vocal tract is of length L (typically 15-17 cm in adults)

Length of each segment is the distance sound travels in half a sample period = $0.5cT$: 1.5 cm @ 11 kHz

- c = speed of sound in air

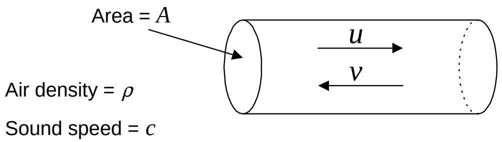
  $\approx 20\sqrt{\text{Absolute Temperature}} \approx 340$ m/s

- T = sample period = $1/f_{samp}$

Number of tube segments needed = $2L/cT \approx 0.001\, f_{samp}$

Sound Waves in a Tube

Acoustic signal is the superposition of two waves: u in the forward direction and v in the reverse direction:



Total volume flow = $u–v$

Total acoustic pressure = $(u+v) \times \rho c/A$

Exactly analogous to transmission lines:

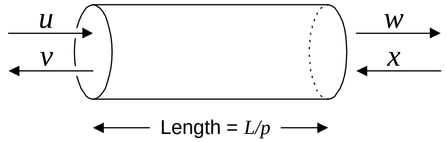
- Volume flow ≈ Current, Pressure ≈ Voltage
- Acoustic Impedance of tube = $\rho c/A$

Assumptions:

- Sound waves are 1-dimensional: true for frequencies < 3 kHz whose wavelengths are long compared to the tube width
- No frictional or wall-vibration energy losses

See appendix for a non-examinable derivation.

Segment Delays



Time for sound to travel along segment = L/cp

Hence: $v(t) = x\left(t - \frac{L}{cp}\right)$ and $u(t) = w\left(t + \frac{L}{cp}\right)$

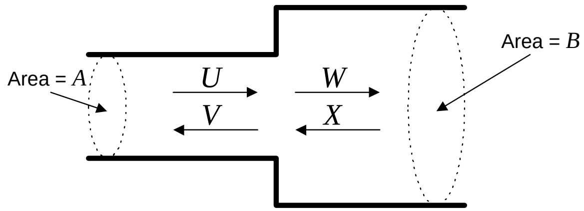
Segment length chosen to correspond to half a sample period. If we take z-transforms, this time delay corresponds to multiplying by $z^{-½}$:

$$V(z) = z^{-½}X(z) \quad \text{and} \quad U(z) = z^{+½}W(z)$$

In matrix form:

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} z^{+½} & 0 \\ 0 & z^{-½} \end{pmatrix}\begin{pmatrix} W \\ X \end{pmatrix} = z^{+½}\begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix}\begin{pmatrix} W \\ X \end{pmatrix}$$

Segment Junction



Flow Continuity: $(U - V) = (W - X)$

Pressure Continuity: $\frac{\rho c}{A}(U + V) = \frac{\rho c}{B}(W + X)$

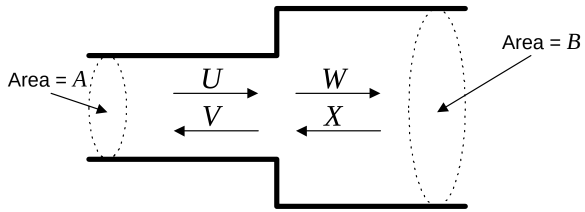
In matrix form: $\begin{pmatrix} 1 & -1 \\ B & B \end{pmatrix}\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ A & A \end{pmatrix}\begin{pmatrix} W \\ X \end{pmatrix}$

Hence:

$$\begin{pmatrix} U \\ V \end{pmatrix} = \frac{1}{2B}\begin{pmatrix} B & 1 \\ -B & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ A & A \end{pmatrix}\begin{pmatrix} W \\ X \end{pmatrix}$$

$$= \frac{1}{2B}\begin{pmatrix} A+B & A-B \\ A-B & A+B \end{pmatrix}\begin{pmatrix} W \\ X \end{pmatrix}$$

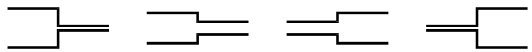
Reflection Coefficients



Define the reflection coefficient to be $r = \frac{B - A}{B + A}$

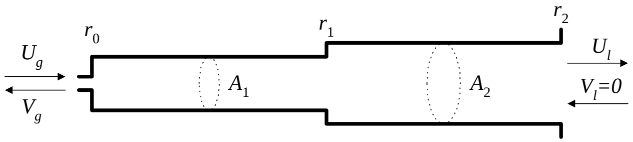
$$\begin{pmatrix} U \\ V \end{pmatrix} = \frac{1}{2B}\begin{pmatrix} A+B & A-B \\ A-B & A+B \end{pmatrix}\begin{pmatrix} W \\ X \end{pmatrix} = \frac{1}{1+r}\begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}\begin{pmatrix} W \\ X \end{pmatrix}$$

Reflection coefficients always lie in the range ±1:



$r = -0.8$ $r = -0.4$ $r = +0.4$ $r = +0.8$

2-Segment Vocal Tract

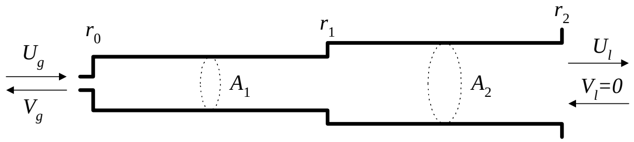


$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}$$

$$\begin{pmatrix} U_g \\ V_g \end{pmatrix} = \frac{1}{1+r_0}\begin{pmatrix} 1 & -r_0 \\ -r_0 & 1 \end{pmatrix} \times z^{½} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \times \frac{1}{1+r_1}\begin{pmatrix} 1 & -r_1 \\ -r_1 & 1 \end{pmatrix} \times z^{½} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \times \frac{1}{1+r_2}\begin{pmatrix} 1 & -r_2 \\ -r_2 & 1 \end{pmatrix}\begin{pmatrix} U_l \\ 0 \end{pmatrix}$$

- Assume $V_l = 0$: no sound reflected back into mouth
- Work backwards from lips towards glottis:
  - Junction: use the reflection matrix
  - Tube segment: use the delay matrix
- A_3 is large but not infinite: assumption of narrow tube breaks down at this point
- A_0 is approximately zero: area of glottis opening

Vocal Tract Transfer Function



Multiplying out the matrices gives:

$$\begin{pmatrix} U_g \\ V_g \end{pmatrix} = \frac{z^{+1}}{\prod_{k=0}^{2}(1+r_k)}\begin{pmatrix} 1 + (r_0r_1 + r_1r_2)z^{-1} + r_0r_2z^{-2} \\ -r_0 - (r_1 + r_0r_1r_2)z^{-1} - r_2z^{-2} \end{pmatrix} U_l$$

We can ignore V_g: it gets absorbed in the lungs.

The vocal tract transfer function is given by the ratio of U_l to U_g:

$$\frac{U_l}{U_g} = \frac{\prod_{k=0}^{2}(1+r_k) \times z^{-1}}{1 + (r_0r_1 + r_1r_2)z^{-1} + r_0r_2z^{-2}}$$

$$= \frac{Gz^{-1}}{1 + (r_0r_1 + r_1r_2)z^{-1} + r_0r_2z^{-2}}$$

$$= \frac{Gz^{-1}}{1 - a_1z^{-1} - a_2z^{-2}}$$

p-segment Vocal Tract

Note that: $\frac{1}{1+r}\begin{pmatrix}1 & -r\\ -r & 1\end{pmatrix}\times z^{½}\begin{pmatrix}1 & 0\\ 0 & z^{-1}\end{pmatrix}=\frac{z^{½}}{1+r}\begin{pmatrix}1 & -rz^{-1}\\ -r & z^{-1}\end{pmatrix}$

Multiplying together all the matrices for a *p*-segment vocal tract gives:

$$\begin{pmatrix}U_g\\ V_g\end{pmatrix}=\frac{z^{½ p}}{\prod_{k=0}^{p}(1+r_k)}\prod_{k=0}^{p-1}\begin{pmatrix}1 & -r_k z^{-1}\\ -r_k & z^{-1}\end{pmatrix}\times\begin{pmatrix}1\\ -r_p\end{pmatrix}U_l$$

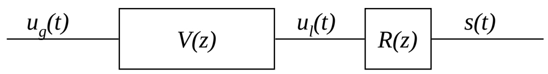
This results in a transfer function of the form:

$$V(z)=\frac{U_l}{U_g}=\frac{Gz^{-½ p}}{1-a_1z^{-1}-a_2z^{-2}-\ldots-a_pz^{-p}}$$

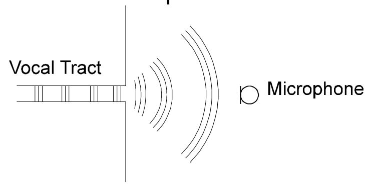
Where:

- G is a gain term
- $z^{-½p}$ is the acoustic time delay along the vocal tract
- The denominator represents a p^{th} order all-pole filter

Lip Radiation



R(z) is the transfer function between *airflow* at the lips and *pressure* at the microphone.



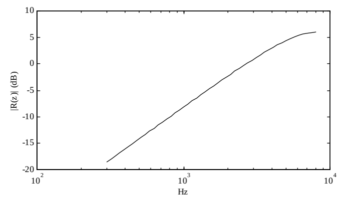
For a lip-opening area of A, acoustic theory predicts a 1st-order high-pass response with a corner frequency of:

$$\frac{c}{\sqrt{4A}}\,\text{Hz}\approx 5\,\text{kHz}$$

For f_{samp} < 20 kHz, a good approximation is:

$$R(z)=\frac{S(z)}{U_l(z)}=1-z^{-1}$$

$$\Rightarrow |R(z)|=2\sin\left(\frac{\omega T}{2}\right)$$

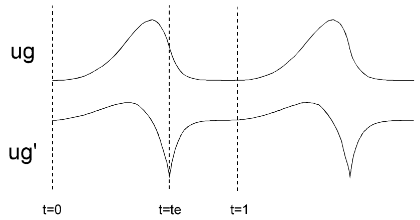


Spectrum of Glottal Waveform

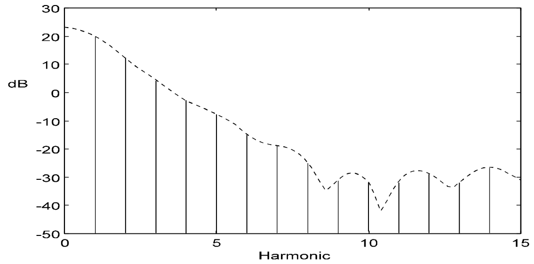
"LF Model" (Liljencrants & Fant)

$$u'_g(t) = \begin{cases} e^{at}\sin(bt) & 0 \le t < t_e \\ c + de^{-ft} & t_e \le t < 1 \end{cases}$$

with $u_g(0) = u_g(1) = 0$; $u_g(t)$ and $u'_g(t)$ continuous at t_e

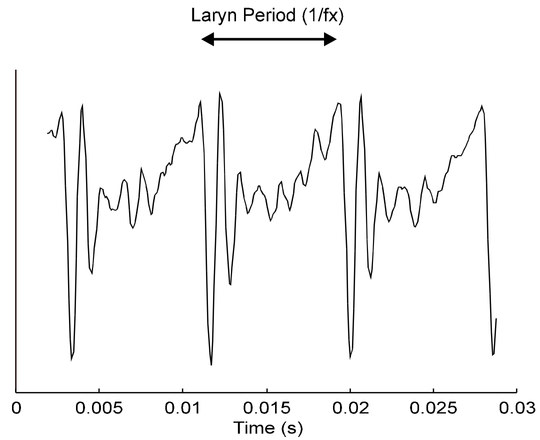


Line Spectrum of ug (approx –12 dB/octave):



Vowel Waveform

Vowel /ɑ/ from "part"



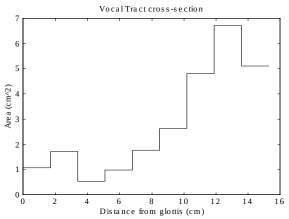
- Larynx Frequency ≈ 130 Hz
- First Vocal tract resonance (formant) ≈ 1 kHz

There is not necessarily any relation between the larynx frequency and the vocal tract resonances.

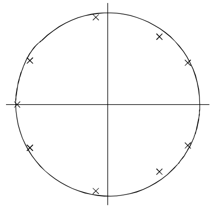
Resonances at a multiple of the larynx frequency will be louder (good for singers)

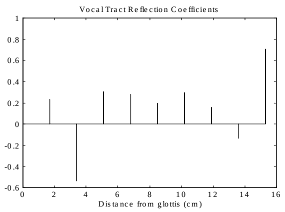
Vocal Tract Shape and Response

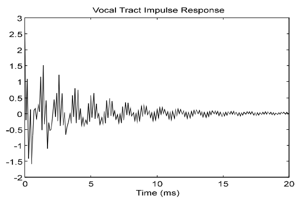
Example: /ɑ/ vowel (“part”)



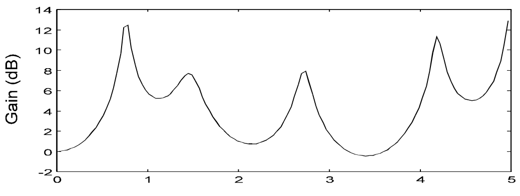
Z-plane Pole Positions







Vocal Tract Filter Response



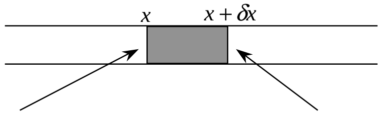
Appendix

Theoretical Derivation of Sound Waves

This section is non-examinable

1-Dimensional Sound Waves

Consider a small chunk of air in a tube with a uniform cross-sectional area A:



$\text{Pressure} = p$ $\qquad$ $\text{Pressure} = p + \delta x \dfrac{\partial p}{\partial x}$

$\text{Velocity} = v = \dfrac{u}{A}$ $\qquad$ $\text{Velocity} = v + \delta v = \dfrac{1}{A}\left(u + \delta x \dfrac{\partial u}{\partial x}\right)$

$$\Rightarrow \quad \delta v = \frac{\delta x}{A}\frac{\partial u}{\partial x}$$

<u>Volume of air chunk:</u> $\quad V = A \times \delta x$

Hence: $\quad \dfrac{\partial V}{\partial t} = A \times \delta v = A \times \dfrac{\delta x}{A}\dfrac{\partial u}{\partial x} = \dfrac{V}{A} \times \dfrac{\partial u}{\partial x}$ $\quad$ ①

<u>Net force on air chunk:</u>

$$F = Ap - A\left(p + \delta x \frac{\partial p}{\partial x}\right) = -A\,\delta x \frac{\partial p}{\partial x}$$

Gas Laws

<u>Ideal Gas Law :</u>

We can express the pressure in terms of the density:

$$pV = nRT$$
$$= \frac{\rho V}{M} RT$$
$$\Rightarrow \quad p = \rho \times \frac{RT}{M}$$

n = moles of air = molecules $\div (6 \times 10^{23})$
R = gas constant = 8.314 J / (K · mol)
T = Temperature (° K)
ρ = density ($\approx$ 1.225 kg / m^3)
M = molecular weight of air = 0.029 kg / mol
γ = specific heat ratio of air = 1.4

We define $\quad c^2 = \dfrac{\gamma RT}{M} \approx (340 \text{ m/s})^2 \quad \Rightarrow p\gamma = \rho c^2$ $\quad$ ②

c will turn out to be the speed of sound and depends only on T.

<u>Adiabatic Gas Law:</u> For pressure changes too rapid for heat conduction to occur (e.g. sound vibrations):

$$\frac{d}{dt}\left(pV^{\gamma}\right) = 0 \quad \Rightarrow \quad V^{\gamma}\frac{\partial p}{\partial t} + p\gamma V^{\gamma-1}\frac{\partial V}{\partial t} = 0$$

using ① and ② $\quad \Rightarrow \quad V^{\gamma}\dfrac{\partial p}{\partial t} = -\rho c^2 \times \dfrac{V^{\gamma}}{A} \times \dfrac{\partial u}{\partial x}$

$$\Rightarrow \quad A\frac{\partial p}{\partial t} = -\rho c^2 \frac{\partial u}{\partial x}$$ ③

Wave Equations

Mass × Acceleration = Force:

$$\rho V \times \frac{1}{A}\frac{\partial u}{\partial t} = -A\delta x \frac{\partial p}{\partial x} = -V\frac{\partial p}{\partial x} \quad \Rightarrow \quad \rho \frac{\partial u}{\partial t} = -A\frac{\partial p}{\partial x} \qquad ④$$

Wave Equations:

Equations ③ and ④ are known as the *wave equations*:

$$\rho \frac{\partial u}{\partial t} = -A\frac{\partial p}{\partial x} \quad \text{and} \quad A\frac{\partial p}{\partial t} = -\rho c^2 \frac{\partial u}{\partial x}$$

Solution:

$$u(x,t) = u^+(t - x/c) - u^-(t + x/c)$$

$$p(x,t) = \frac{\rho c}{A} \times \left\{u^+(t - x/c) + u^-(t + x/c)\right\}$$

It is easily verified that this solution satisfies the wave equations for any differentiable functions u^+ and u^-.

The two functions u^+ and u^- represent waves travelling in +ve and –ve x directions at velocity c. The actual values of the waves are determined by the boundary conditions at the end of the tube section.

The equations are the same as for a transmission line with $u \approx$ current, $p \approx$ voltage and $\rho c/A \approx$ impedance.

University of California
Berkeley

College of Engineering
Department of Electrical Engineering
and Computer Sciences

Professors : N.Morgan / B.Gold
EE225D Spring,1999

Acoustic Tube Models

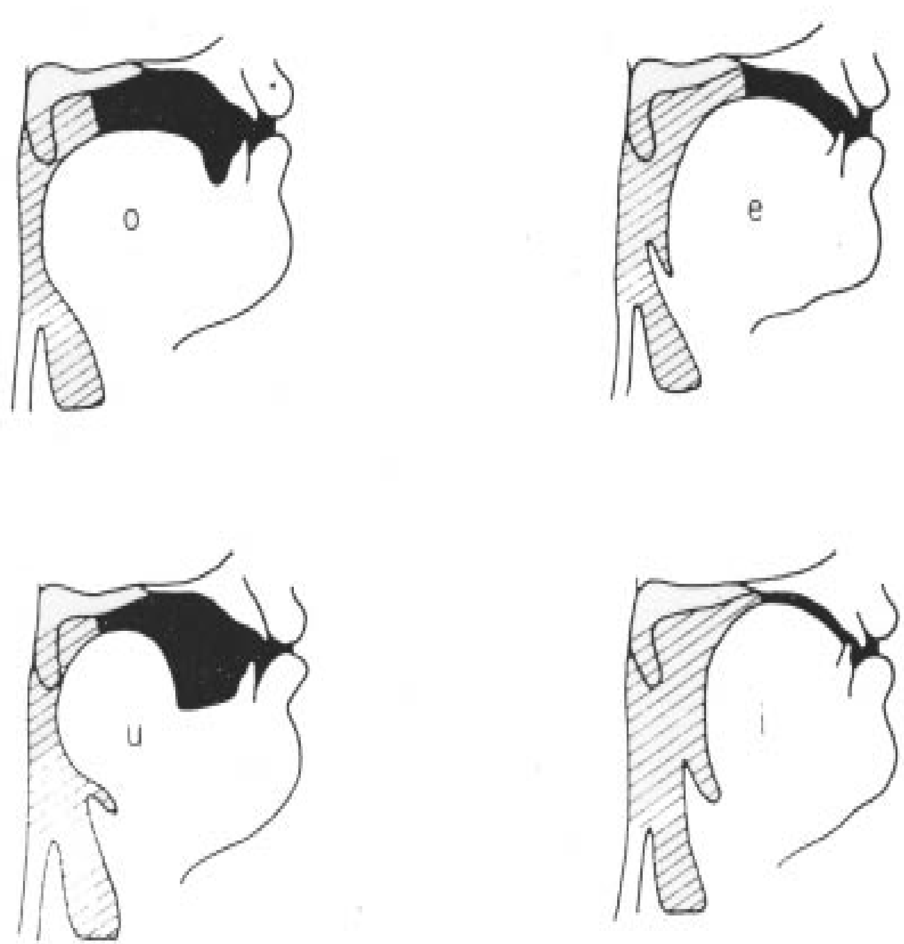
Lecture 13

Introduction :

Acoustic Tube Models of English Phonemes → 2 tube model.

Assumptions :

- Lossless tubes
- Plane waves
- Rigid walls
- Friction
- Thermal effect



Vocal tract area for four vowel sounds

Vocal tract areas for four vowel sounds.

i - Tongue is High.

e - Tongue is a little Lower.

u - Tongue is very Low.

o - Togue is somewhat low.

1. Tube response vs. area function.
2. Discrete-time-space version.
3. Example - 2 tube representation of vowels.

Problem for Today :

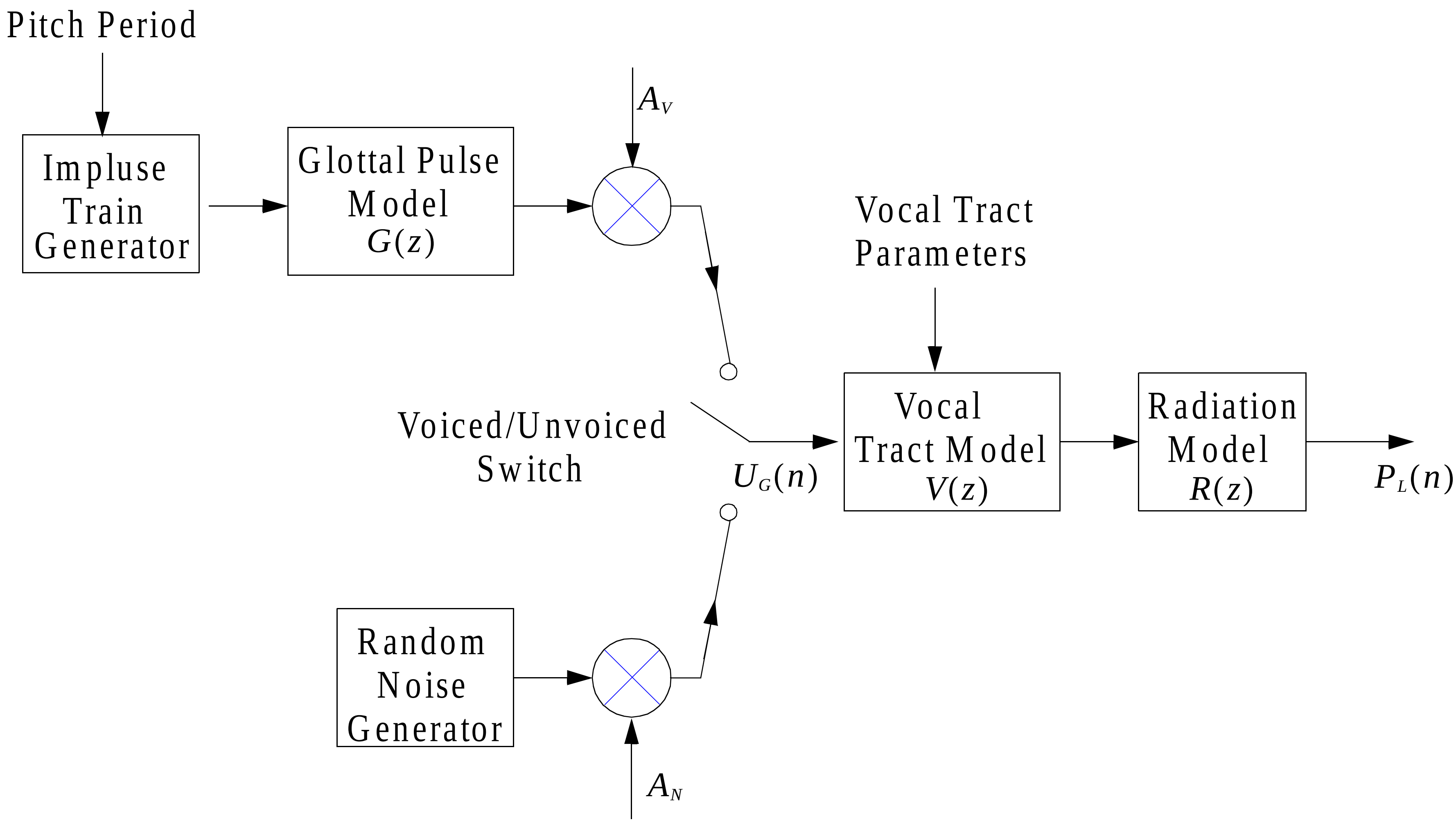
Develop a 2 tube model to derive a frequency response that approximates some vowels.

By solving a complicated wave equation, the frequency response can be found.

Look up equation in R & S.

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

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



Viewgraph 2 : General Discrete-Time Model for Speech Production.

Assumption in this Model :

Vocal Tract Model - Time varying

Radiation Model - May be time varying

Glottal Pulse Model - Usually considered independent of vocal tract model, but later we'll examine this wave closely

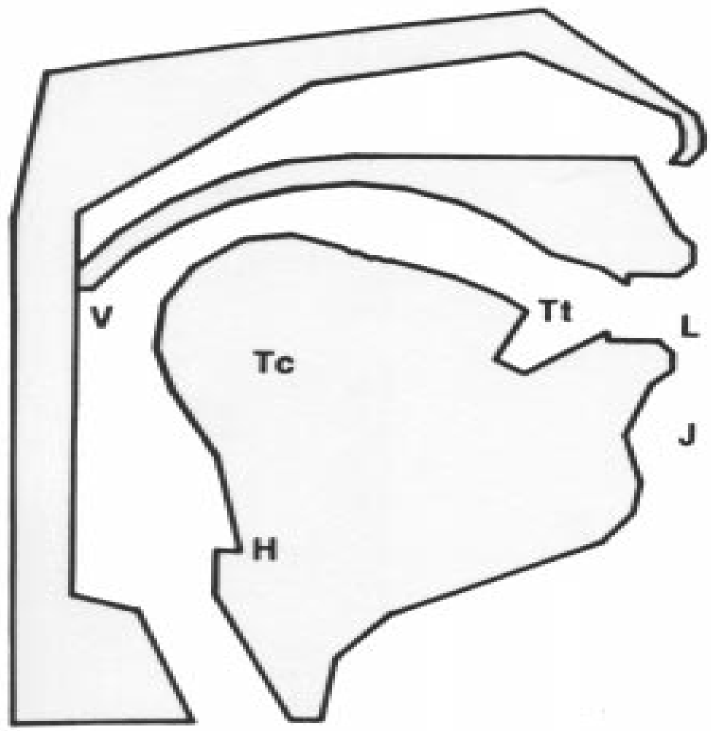
$$u(x, t) = u^{+}\left(t - \frac{x}{C}\right) - u^{-}\left(t + \frac{x}{C}\right)$$

$$p(x, t) = Z_o\left[u^{+}\left(t - \frac{x}{C}\right) + u^{-}\left(t + \frac{x}{C}\right)\right]$$

$p(l, t) = 0$: open tube

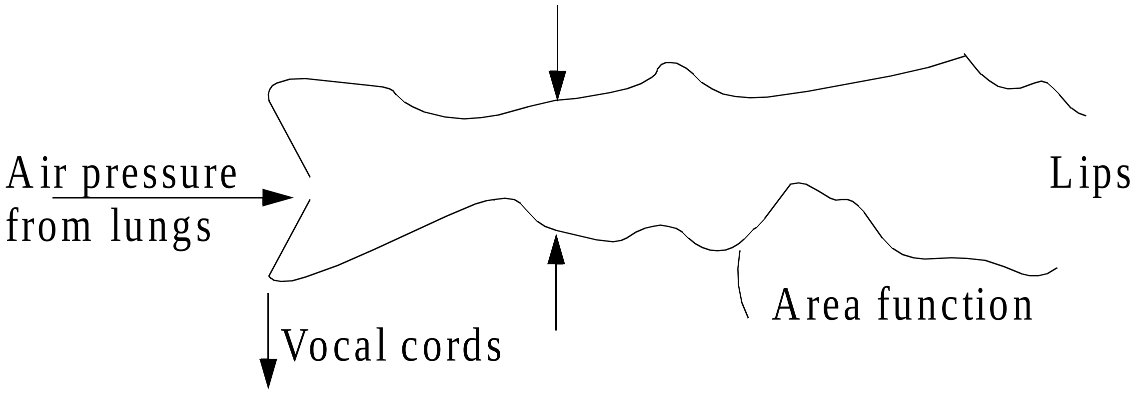
$$u^{+}\left(t - \frac{l}{C}\right) = -u^{-}\left(t + \frac{l}{C}\right)$$

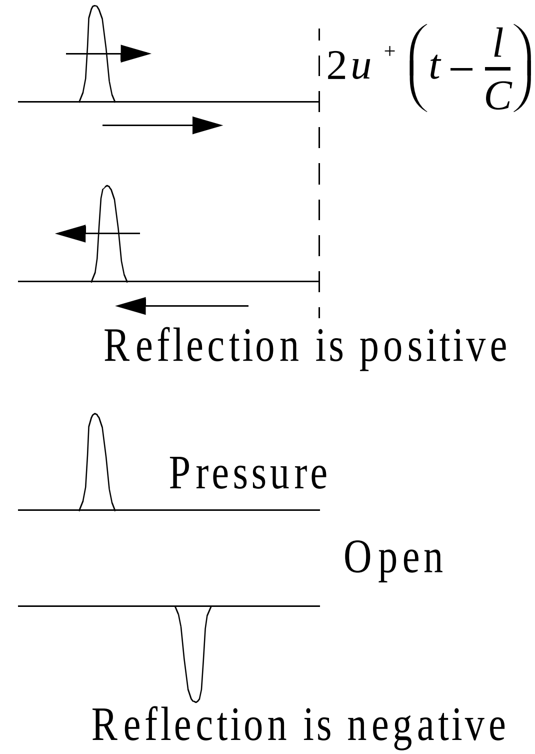
$$u(l, t) = 2u^{+}\left(t - \frac{l}{C}\right)$$



H = HYOID POSITION
J = ANGLE OF JAW OPENING
L = LIP PROTRUSION AND ELEVATION
Tc = TONGUE CENTER
Tt = POSITION OF TONGUE TIP
V = VELUM OPENING

Model of Vocal Tract





Closed Tube $u(l, t) = 0$

so $$u^{+}\left(t - \frac{l}{C}\right) = -u^{-}\left(t + \frac{l}{C}\right)$$

- Given Area Function we can compute Spectrum

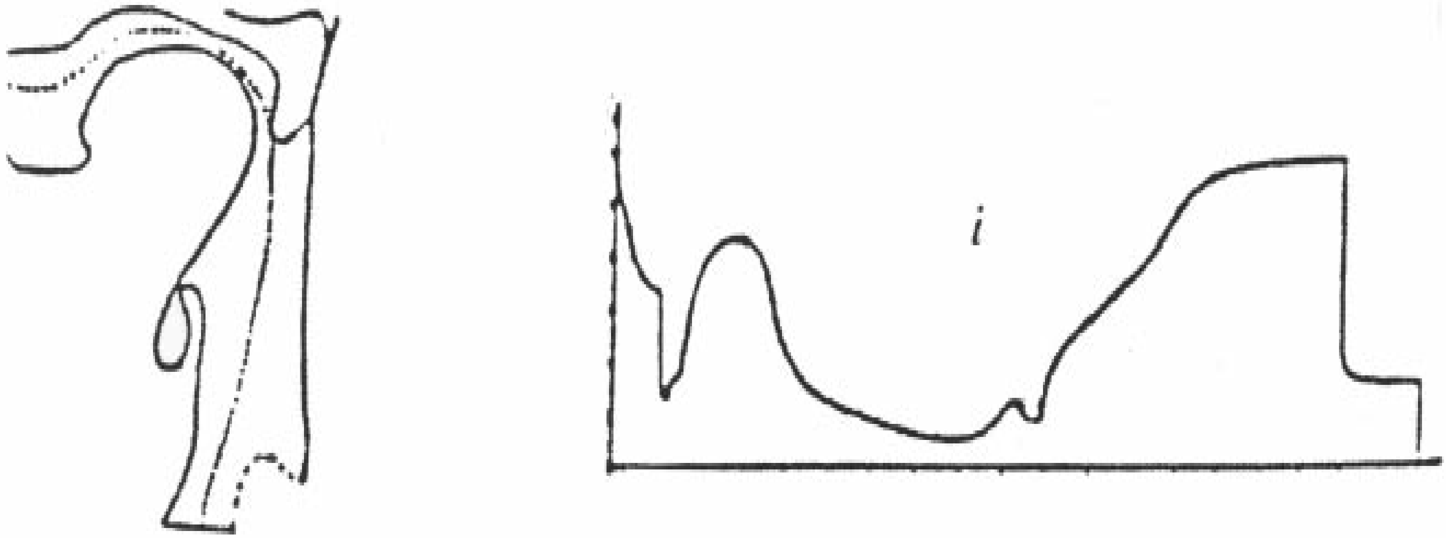


Figure 11.1: X-ray tracing and area function for phoneme /i/

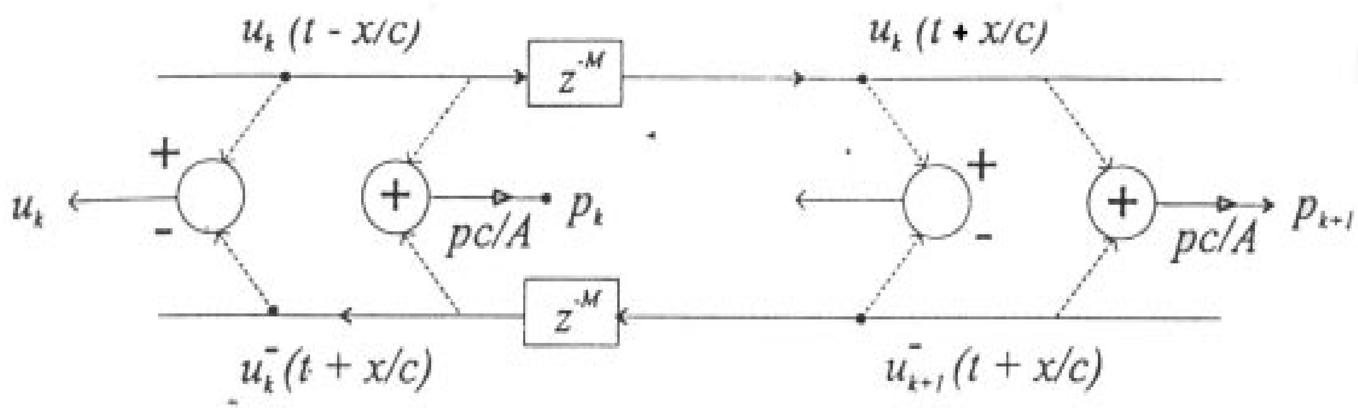


Figure 11.2: Single section of digital wave guide

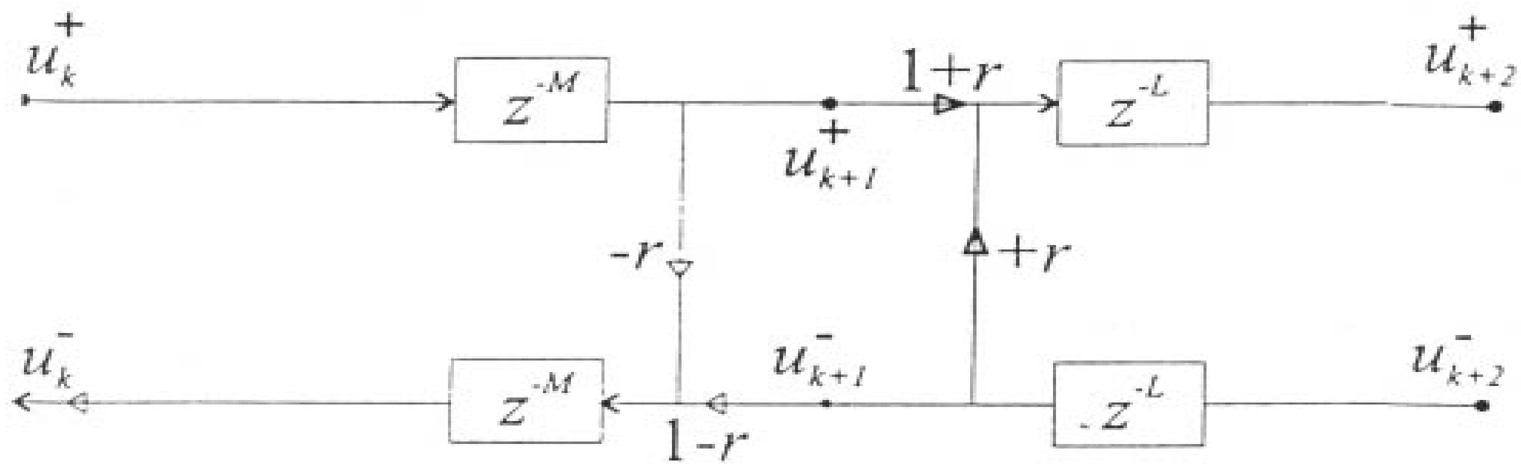


Figure 11.3: Two section digital wave guide

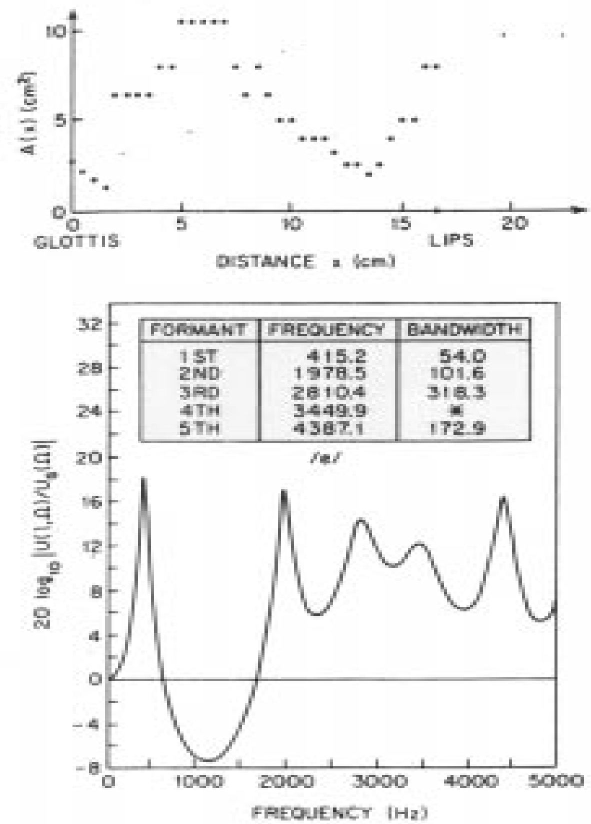


Figure 3.24 Area function and frequency response for the Russian vowel /*e*/

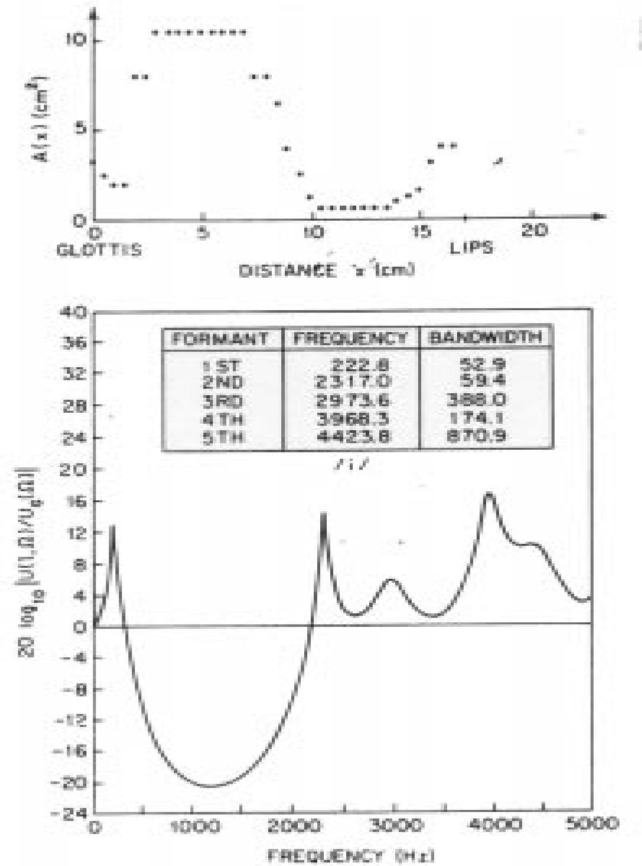
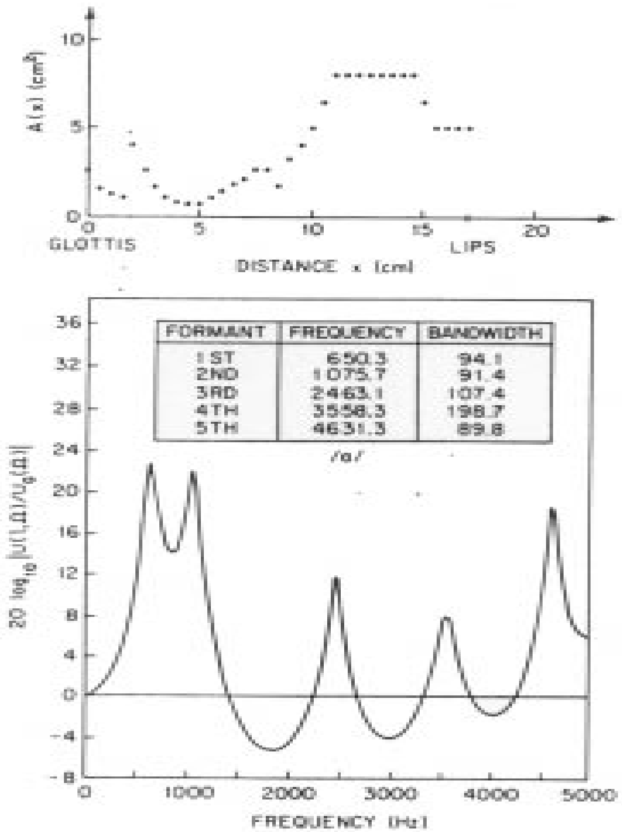
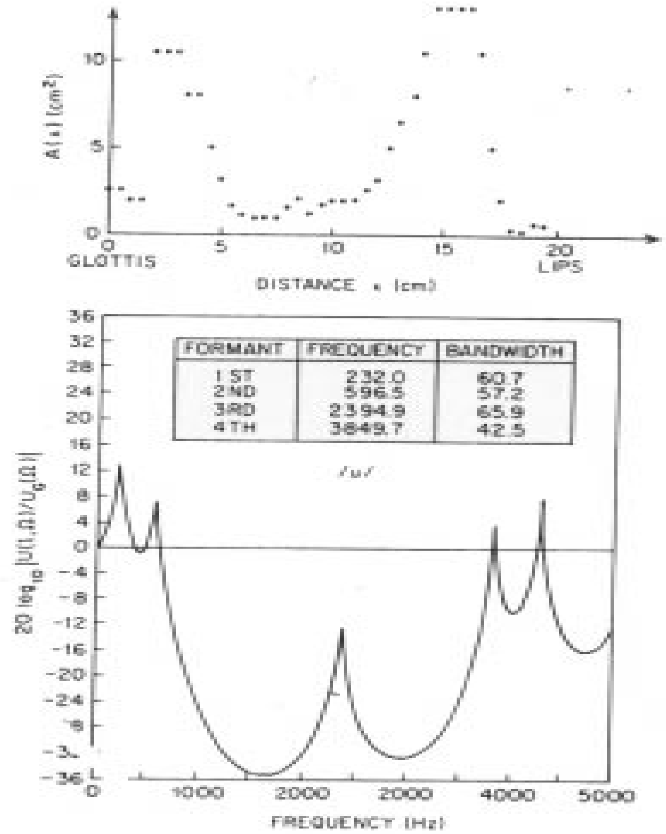


Figure 3.25 Area function and frequency response for the Russian vowel /*i*/



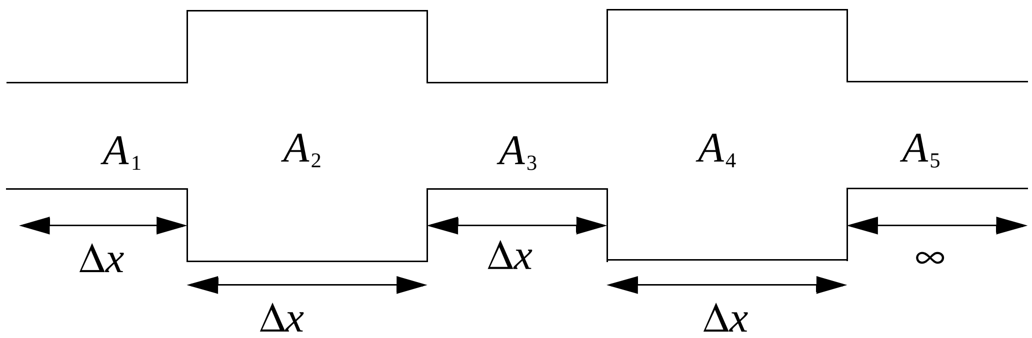
FORMANT	FREQUENCY	BANDWIDTH
1ST	650.3	94.1
2ND	1075.7	91.4
3RD	2463.1	107.4
4TH	3558.3	198.7
5TH	4631.3	89.8



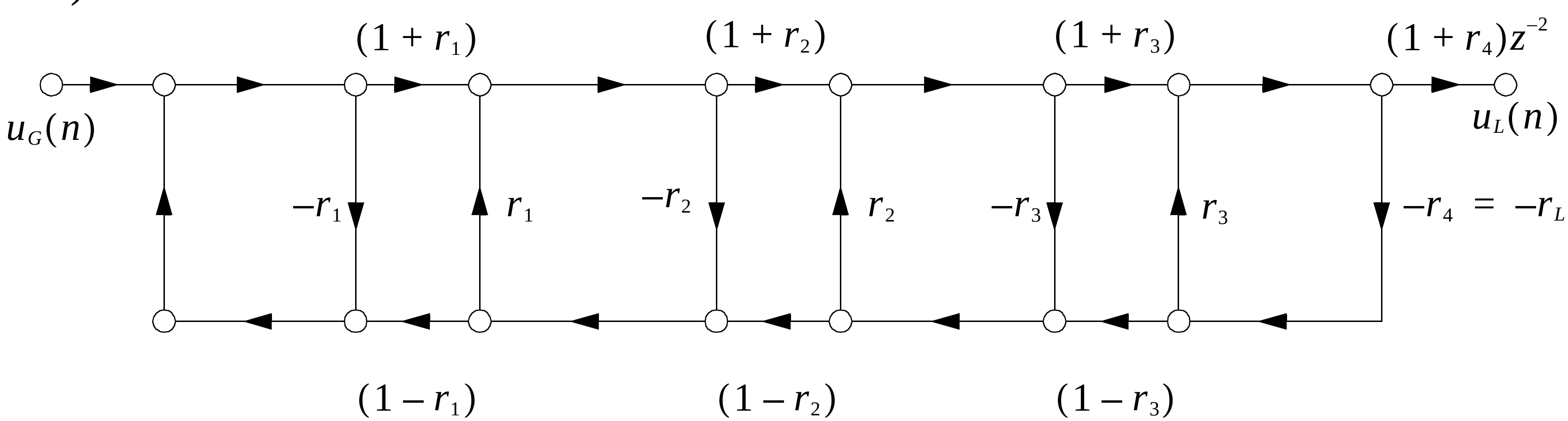
FORMANT	FREQUENCY	BANDWIDTH
1ST	232.0	60.7
2ND	596.5	57.2
3RD	2394.9	65.9
4TH	3849.7	42.5

More Complex Tube Structure

a)



b)



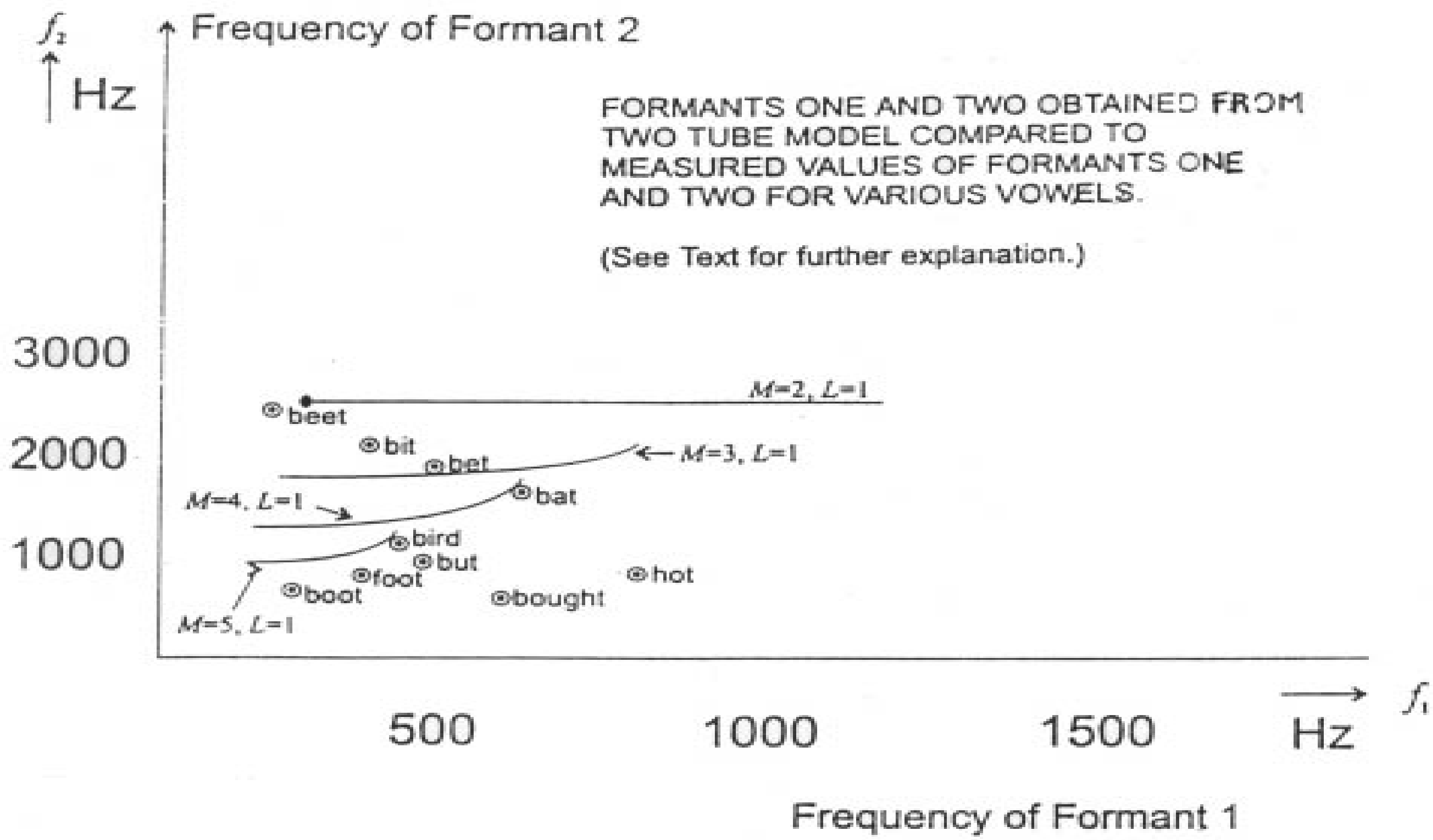


Figure 11.4 Formants 1 and 2 obtained from two tube model

Welcome!

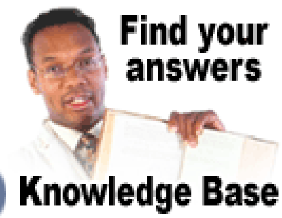
to Syntrillium Software home of Cool Edit 2000 and Pro

...Good Sound Stuff! COOL EDIT Digital Audio Editing

This web site is enhanced with Macromedia Flash. Get the latest version here:

GET macromedia FLASH PLAYER

COOL EDIT PRO



Cool Edit 2000

Did you know you have a recording studio in your computer? Record, clean up, mix, master, and export to MP3 and other formats with our **EXPANDABLE** audio editor.

Cool Edit Pro

Record, mix, and master up to 64 tracks with our most powerful professional audio software. Supports 24/96 and even 32/192 recording and offers more than 40 powerful effects!

More software!
Let your computer play for a while with...

Snoqualmie

Award-winning screen saver like none you've seen before!

Kaleidoscope

The screen saver that responds to music!

Wind Chimes

Bring in the music of the wind!

Knowledge Base

Search our Knowledge Base of technical support information and self-help tools for Syntrillium products.

User Forums

Consult with your counterparts and peers in online open forums. Join discussions about common issues and solutions.

Tutorials and "How-To's"

New Cool Edit tutorials are available! Download tutorials on Scripts and Batch Processing, Preparing Audio for CD, Using Crossfades, and more!

Software Updates

Go here to download and install the latest updates for your Syntrillium Software.

Contact Help

Get answers and help directly from Syntrillium's Technical Support Team.

View NAMM 2002

Pictures and video from our trip to the NAMM 2002 convention!

New Web Site!
Let us know what you think! Cast your vote or email us.

Close-ups

Featured musicians, artists, and radio and audio engineers. Find out how the experts get the most from Cool Edit!
NEW!
Interview with Audley Freed

Stickmen

They're skinny, funny, and full of effects! Check out the latest wacky antics of our Stickmen!

T-Shirts

Get your own Syntrillium T-Shirt! Look like a million, for cheap.

FoosCam

LIVE WEBCAM! Okay, it's just of our Foosball table. But if you're lucky, you may catch us taking a break by foosing around.

NEW IN 2002

cool edit | pro VERSION 2.0

AND

ROVER

MORE INFO

VIDEO & PICTURES

NAMM International Music Products Association

WATCH NOW

Get news in your inbox!

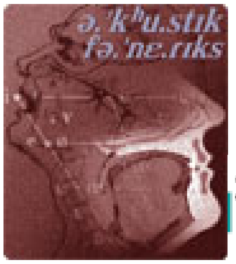
Your Email

More information

News

- Syntrillium debuts free and comprehensive loop library at winter NAMM
- Cool Edit Pro 2.0 sets new standard for affordability vs. functionality
- New book on Cool Edit provides start-to-finish instruction and tips on PC-based recording
- Syntrillium offers Cool Edit Pro users remote control option with debute of new Red Rover interface
- Syntrillium launches new web site. You're on it right now!

More news...



SPP1

What people are saying about SPP1

"This product is one of the very few that lives up to the huge pedagogical potential of interactive CD-ROM technology. It is unusually well-crafted."
- Dr. M. Owren, Department of Psychology, Cornell University

Speech Production and Perception I

A New Multimedia Course on CD-ROM - **$75**

- Developed over three years by teachers of speech science, experts in phonetics and speech research, working with a team of experienced writers, artists and programmers
- Designed to enable students to acquire an intuitive understanding of the correspondence between sound, spectrum and articulation.
- Interactive use of the computer for dynamic, experience-based learning with self- or teacher-guided instruction.
- Uses your Windows-compatible audio card
- Runs on Windows® 95/98/NT and Windows® 3.1
- In use at more than 150 universities and institutions internationally
- Volume discounts are available

Speech Production and Perception I: An Interactive Multimedia Course provides a non-mathematical introduction to the basic concepts of acoustic phonetics and speech science. The course cultivates genuine understanding of these concepts through personal interaction and experience, using hundreds of interactive models and simulations. Course development was carried out with major support from the National Institutes of Health (MH51970-SBIR).

The course has been created for undergraduate students studying Speech and Hearing Science, Communication Disorders, Linguistics, and Phonetics. Computer Science and Electrical Engineering students interested in speech transmission and processing will also find the course stimulating and useful. Its cost is about the same as that of a good textbook.

Test-teaching at universities and colleges in the United States and Europe has shown that the course is an effective adjunct to lecture- and demonstration-based teaching, as well as a resource for independent learning by students.

The course incorporates its own state-of-the-art graphic interface, including custom-designed high-speed digital signal processing and full color visual displays. The combination of efficient, innovative code and a sophisticated user interface allows students easy access to the full range of course activities and resources.

The present course consists of units on:

- Spectrograms
- Vowel Acoustics
- Consonant Acoustics
- Speech Perception
- Vowel Perception

In addition, students have access to:

- A Library with IPA consonant and vowel charts. The library also contains more than 100 new digitally recorded examples of the consonants and vowels in the charts, with spectrograms of each one, an interactive glossary with definitions of more than 100 technical words and phrases, and cross-references to textbooks;
- A Lab in which they can make and compare wide- and narrow-band spectrograms of their own utterances, the speech of others, or any other sounds they wish to record.

The course contains more than 200 interactive demonstrations and a dozen interactive exercises on CD-ROM, as well as separate student worksheets, an Installation Guide, and a User's Guide. Typical interactive demonstrations include adjustable filtering of synthetic voicing sources; plotting the vowel spaces of adult and child speakers; identification and discrimination experiments with speech and non-speech stimuli; creating and analyzing conventional and 3-dimensional spectrograms; and, examining animated vocal tracts synchronized with audio playback and spectrogram displays.

A student workbook with more than 40 pages of questions complements the interactive exercises. The worksheets for the course provide students with a permanent written record of the material covered, and the instructor with a convenient way of evaluating the student's comprehension of the course content. The worksheets are keyed to the topics in the course; each worksheet contains from 5 to 15 questions, generally requiring paragraph-length answers. The questions cover the course material at various levels, from what might be expected of students in introductory courses to questions suitable for consideration by advanced graduate students.

An Instructor's Pack, including a Teacher's Guide with sample answers to worksheet problems, is also available.

top