CS578- Speech Signal Processing
Lecture 3: Acoustics of Speech Production

Yannis Stylianou

University of Crete, Computer Science Dept., Multimedia Informatics Lab
yannis@csd.uoc.gr

Univ. of Crete, 2008 Winter Period
Outline

1 Physics of sound

2 Uniform Tube Model

3 Concatenating N uniform tubes

4 Glottal Flow Derivative

5 Vocal Fold/Vocal Tract Interaction

6 Acknowledgments

7 References
COMPRESSION AND RAREFACTION OF AIR PARTICLES

(a) Infinitely Large Wall

(b) 

(c)
Some definitions

- **Sound wave**: propagation of disturbance (local changes in pressure, displacement, and velocity) of particles through a medium, creating the effect of compression or rarefaction.

- **Wavelength**: distance between two consecutive peak compressions, \( \lambda \)

- **Frequency**: number of cycles of compressions per second, \( f \)

- **Speed of sound**: \( c = f \lambda \) (at sea level and at 70° F, \( c = 344 m/sec \))

- **Isothermal process**: a slow variation of pressure where the temperature in the medium remains constant

- **Adiabatic process**: a fast variation of pressure where the temperature in the medium increases
Some definitions

- **Sound wave**: propagation of disturbance (local changes in pressure, displacement, and velocity) of particles through a medium, creating the effect of compression or rarefaction.

- **Wavelength**: distance between two consecutive peak compressions, $\lambda$

- **Frequency**: number of cycles of compressions per second, $f$

- **Speed of sound**: $c = f\lambda$ (at sea level and at $70^\circ F$, $c = 344 m/sec$)

- **Isothermal process**: a slow variation of pressure where the temperature in the medium remains constant

- **Adiabatic process**: a fast variation of pressure where the temperature in the medium increases
Some definitions

- **Sound wave**: propagation of disturbance (local changes in pressure, displacement, and velocity) of particles through a medium, creating the effect of compression or rarefaction.

- **Wavelength**: distance between two consecutive peak compressions, $\lambda$

- **Frequency**: number of cycles of compressions per second, $f$

- **Speed of sound**: $c = f \lambda$ (at sea level and at $70^\circ F$, $c = 344 m/sec$)

- **Isothermal process**: a slow variation of pressure where the temperature in the medium remains constant

- **Adiabatic process**: a fast variation of pressure where the temperature in the medium increases
Some definitions

- **Sound wave**: propagation of disturbance (local changes in pressure, displacement, and velocity) of particles through a medium, creating the effect of compression or rarefaction.

- **Wavelength**: distance between two consecutive peak compressions, $\lambda$

- **Frequency**: number of cycles of compressions per second, $f$

- **Speed of sound**: $c = f\lambda$ (at sea level and at $70^\circ F$, $c = 344 m/sec$)

- **Isothermal process**: a slow variation of pressure where the temperature in the medium remains constant

- **Adiabatic process**: a fast variation of pressure where the temperature in the medium increases
**Some definitions**

- **Sound wave:** propagation of disturbance (local changes in pressure, displacement, and velocity) of particles through a medium, creating the effect of compression or rarefaction.

- **Wavelength:** distance between two consecutive peak compressions, $\lambda$

- **Frequency:** number of cycles of compressions per second, $f$

- **Speed of sound:** $c = f \lambda$ (at sea level and at $70^\circ F$, $c = 344 m/sec$)

- **Isothermal process:** a slow variation of pressure where the temperature in the medium remains constant

- **Adiabatic process:** a fast variation of pressure where the temperature in the medium increases
Some definitions

- **Sound wave:** propagation of disturbance (local changes in pressure, displacement, and velocity) of particles through a medium, creating the effect of compression or rarefaction.

- **Wavelength:** distance between two consecutive peak compressions, \( \lambda \)

- **Frequency:** number of cycles of compressions per second, \( f \)

- **Speed of sound:** \( c = f \lambda \) (at sea level and at 70°F, \( c = 344 m/sec \))

- **Isothermal process:** a slow variation of pressure where the temperature in the medium remains constant

- **Adiabatic process:** a fast variation of pressure where the temperature in the medium increases
Cube configuration

Infinite Vibrating Wall

\[ p + \frac{\partial P}{\partial x} \Delta x \]
**Notation**

Assuming *planar propagation*, and within the cube:

- $p(x, t)$ fluctuation of pressure about an ambient or average pressure $P_0$.
  - Threshold of hearing: $2 \times 10^{-5}$ newtons/m$^2$
  - Threshold of pain: 20 newtons/m$^2$

- $\nu(x, t)$ fluctuation of particles’ velocity about zero average velocity.

- $\rho(x, t)$ fluctuation of particles’ density about an average density $\rho_0$. 

Assuming *planar propagation*, and within the cube:

- $p(x, t)$ fluctuation of pressure about an ambient or average pressure $P_0$.
  - Threshold of hearing: $2 \times 10^{-5}$ newtons/m$^2$
  - Threshold of pain: 20 newtons/m$^2$

- $\nu(x, t)$ fluctuation of particles’ velocity about zero average velocity.

- $\rho(x, t)$ fluctuation of particles’ density about an average density $\rho_0$. 

**Notation**
Assuming *planar propagation*, and within the cube:

- $p(x, t)$ fluctuation of pressure about an ambient or average pressure $P_0$.
  - Threshold of hearing: $2 \times 10^{-5}$ *newtons/m$^2*  
  - Threshold of pain: 20 *newtons/m$^2*
- $\nu(x, t)$ fluctuation of particles’ velocity about zero average velocity.
- $\rho(x, t)$ fluctuation of particles’ density about an average density $\rho_0$. 
The Wave Equation

Under the assumptions:

- If there is no friction of air particles in the cube with those outside the cube (no viscosity),
- Cube is very small,
- The density of air particles is constant in the cube (i.e., $\rho_0 = \rho$)

then, one form of the Wave Equation is given by:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial v}{\partial t}$$

$$-\frac{\partial p}{\partial t} = \rho c^2 \frac{\partial v}{\partial x}$$
The Wave Equation

Under the assumptions:

- If there is no friction of air particles in the cube with those outside the cube (no viscosity),
- Cube is very small,
- The density of air particles is constant in the cube (i.e., $\rho_0 = \rho$)

then, one form of the Wave Equation is given by:

$$-rac{\partial p}{\partial x} = \rho \frac{\partial v}{\partial t}$$

$$-rac{\partial p}{\partial t} = \rho c^2 \frac{\partial v}{\partial x}$$
Under the assumptions:

- If there is no friction of air particles in the cube with those outside the cube (no *viscosity*),
- Cube is very small,
- The density of air particles is constant in the cube (i.e., $\rho_0 = \rho$)

then, one form of the *Wave Equation* is given by:

\[
- \frac{\partial p}{\partial x} = \rho \frac{\partial v}{\partial t} \\
- \frac{\partial p}{\partial t} = \rho c^2 \frac{\partial v}{\partial x}
\]
The Wave Equation

Under the assumptions:

- If there is no friction of air particles in the cube with those outside the cube (no viscosity),
- Cube is very small,
- The density of air particles is constant in the cube (i.e., \( \rho_0 = \rho \))

then, one form of the Wave Equation is given by:

\[
- \frac{\partial p}{\partial x} = \rho \frac{\partial v}{\partial t} \\
- \frac{\partial p}{\partial t} = \rho c^2 \frac{\partial v}{\partial x}
\]
The Wave Equation

Under the assumptions:

- If there is no friction of air particles in the cube with those outside the cube (no viscosity),
- Cube is very small,
- The density of air particles is constant in the cube (i.e., \( \rho_0 = \rho \))

then, one form of the Wave Equation is given by:

\[
- \frac{\partial p}{\partial x} = \rho \frac{\partial v}{\partial t} \\
- \frac{\partial p}{\partial t} = \rho c^2 \frac{\partial v}{\partial x}
\]
Outline

1. Physics of sound
2. Uniform Tube Model
3. Concatenating N uniform tubes
4. Glottal Flow Derivative
5. Vocal Fold/Vocal Tract Interaction
6. Acknowledgments
7. References
Lossless Case of cross section $A$

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

where $u(x, t) = Av(x, t)$
Solution for a Lossless Tube

Under the assumptions/conditions:

- No friction along the walls of the tube
- At the open end of the tube, there are no variations in air pressure, i.e. \( p(l, t) = 0 \)
- Volume velocity at \( x = 0 \): \( u(0, t) = U_g(\Omega) e^{j\Omega t} \)

\[ u(x, t) = \frac{\cos \left[ \Omega \frac{(l - x)}{c} \right]}{\cos \left( \Omega \frac{l}{c} \right)} U_g(\Omega) e^{j\Omega t} \]

\[ p(x, t) = j \frac{\rho c}{A} \frac{\sin \left[ \Omega \frac{(l - x)}{c} \right]}{\cos \left( \Omega \frac{l}{c} \right)} U_g(\Omega) e^{j\Omega t} \]

where \( U_g(\Omega) e^{j\Omega t} \) denotes volume velocity at \( x = 0 \)
Solution for a Lossless Tube

Under the assumptions/conditions:

- No friction along the walls of the tube
- At the open end of the tube, there are no variations in air pressure, i.e. \( p(l, t) = 0 \)
- Volume velocity at \( x = 0 \): \( u(0, t) = U_g(\Omega)e^{j\Omega t} \)

\( \triangleright \) Volume velocity:

\[
 u(x, t) = \frac{\cos \left[\frac{\Omega(l - x)}{c}\right]}{\cos \left(\frac{\Omega \ l}{c}\right)} U_g(\Omega)e^{j\Omega t}
\]

\( \triangleright \) (Incremental) Pressure:

\[
 p(x, t) = j \frac{\rho c}{A} \frac{\sin \left[\frac{\Omega(l - x)}{c}\right]}{\cos \left(\frac{\Omega \ l}{c}\right)} U_g(\Omega)e^{j\Omega t}
\]

where \( U_g(\Omega)e^{j\Omega t} \) denotes volume velocity at \( x = 0 \)
Velocity and Pressure are "orthogonal"
At $x = l$

$$u(l, t) = \frac{1}{\cos(\Omega \ l/c)} U_g(\Omega) e^{j\Omega t}$$

Then, the frequency response $V(\Omega)$ is:

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

providing resonances of infinite amplitudes at frequencies:

$$\Omega_k = (2k + 1) \frac{\pi c}{2l}, \quad k = 0, 1, 2, \cdots$$

Example: if $l = 35$cm, $c = 350$ m/s, then $f_k = 250, 750, 1250, \cdots$ Hz.
At \( x = l \)

\[
u(l, t) = \frac{1}{\cos(\Omega l/c)} U_g(\Omega)e^{j\Omega t}
\]

Then, the frequency response \( V(\Omega) \) is:

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

providing resonances of infinite amplitudes at frequencies:

\[
\Omega_k = (2k + 1) \frac{\pi c}{2l}, \quad k = 0, 1, 2, \cdots
\]

Example: if \( l = 35 \text{cm}, c = 350 \text{ m/s} \), then \( f_k = 250, 750, 1250, \cdots \) Hz.
**Uniform Tube: Being Realistic**

Energy loss due to the wall vibration (left) and with viscous and thermal loss (right)[1]:

![Graphs showing energy loss](image)

<table>
<thead>
<tr>
<th>Formant</th>
<th>Frequency (Hz)</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>504.6</td>
<td>53.3</td>
</tr>
<tr>
<td>2nd</td>
<td>1512.3</td>
<td>40.8</td>
</tr>
<tr>
<td>3rd</td>
<td>2515.7</td>
<td>28.0</td>
</tr>
<tr>
<td>4th</td>
<td>3518.8</td>
<td>19.0</td>
</tr>
<tr>
<td>5th</td>
<td>4524.0</td>
<td>13.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Formant</th>
<th>Frequency (Hz)</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>502.5</td>
<td>59.3</td>
</tr>
<tr>
<td>2nd</td>
<td>1508.9</td>
<td>51.1</td>
</tr>
<tr>
<td>3rd</td>
<td>2511.2</td>
<td>41.1</td>
</tr>
<tr>
<td>4th</td>
<td>3513.5</td>
<td>34.5</td>
</tr>
<tr>
<td>5th</td>
<td>4518.0</td>
<td>30.8</td>
</tr>
</tbody>
</table>
Sound radiation at the lips, as an acoustic impedance:

\[ Z_r(\Omega) = \frac{P(l, \Omega)}{U(l, \Omega)} \]

All the previous losses, plus radiation loss[1]:
Since we measure pressure at the lips:

$$H(\Omega) = \frac{P(l, \Omega)}{U_g(\Omega)} = Z_r(\Omega)V(\Omega)$$
Numerical simulations for /o/[1]
Reflection coefficient:

\[ r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \]
Discretizing the continuous-space tube

- Impulse response of $N$ lossless concatenated tubes with total length $l$:

$$h(t) = b_0 \delta(t - N\tau) + \sum_{k=1}^{\infty} b_k \delta(t - N\tau - k2\tau)$$

where $\tau = \frac{\Delta x}{c}$ and $\Delta x = \frac{l}{N}$

- Frequency response:

$$H(\Omega) = \sum_{k=0}^{\infty} b_k e^{-j\Omega2k\tau}$$

- Observe that:

$$H(\Omega + \frac{2\pi}{2\tau}) = H(\Omega)$$
Impulse response of $N$ lossless concatenated tubes with total length $l$:

$$h(t) = b_0 \delta(t - N\tau) + \sum_{k=1}^{\infty} b_k \delta(t - N\tau - k2\tau)$$

where $\tau = \frac{\Delta x}{c}$ and $\Delta x = \frac{l}{N}$

Frequency response:

$$H(\Omega) = \sum_{k=0}^{\infty} b_k e^{-j\Omega 2k\tau}$$

Observe that:

$$H(\Omega + \frac{2\pi}{2\tau}) = H(\Omega)$$
Discretizing the continuous-space tube

- Impulse response of $N$ lossless concatenated tubes with total length $l$:

$$h(t) = b_0 \delta(t - N\tau) + \sum_{k=1}^{\infty} b_k \delta(t - N\tau - k2\tau)$$

where $\tau = \frac{\Delta x}{c}$ and $\Delta x = \frac{l}{N}$

- Frequency response:

$$H(\Omega) = \sum_{k=0}^{\infty} b_k e^{-j\Omega 2k\tau}$$

- Observe that:

$$H(\Omega + \frac{2\pi}{2\tau}) = H(\Omega)$$
(a) two concatenated tubes, (b) lip boundary condition, (c) glottal boundary condition
For a lossless two-tube model

Transfer function relating the volume velocity at the lips to the glottis:

\[ V(s) = \frac{b e^{-s2\tau}}{1 + a_1 e^{-s2\tau} + a_2 e^{-s4\tau}} \]

with \( a_1 = r_1 r_g + r_1 r_L \), \( a_2 = r_L r_g \) and \( b = 0.5(1 + r_g)(1 + r_L)(1 + r_1) \)
(Show me this)
Discrete-time lossless models

- **Two cubes**: By setting \( z = e^{s^2 \tau} \), then:

\[
V(z) = \frac{bz^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}}
\]

- **N cubes**:

\[
V(z) = \frac{Az^{-N/2}}{1 + \sum_{k=1}^{N} a_k z^{-k}}
\]
Two cubes: By setting $z = e^{s^2 \tau}$, then:

$$V(z) = \frac{b z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

N cubes:

$$V(z) = \frac{A z^{-N/2}}{1 + \sum_{k=1}^{N} a_k z^{-k}}$$
Choosing the number of tube elements

Question:
If a vocal tract has length \( l = 17.5 \, \text{cm} \) and the speed of sound \( c = 350 \, \text{m/s} \), how many tubes, \( N \), do we need to cover a bandwidth of 5000 \( \text{Hz} \)?

Answer: \( N = 10 \)
Choosing the number of tube elements

Question:
If a vocal tract has length \( l = 17.5 \, \text{cm} \) and the speed of sound \( c = 350 \, \text{m/s} \), how many tubes, \( N \), do we need to cover a bandwidth of 5000 \( \text{Hz} \)?

Answer: \( N = 10 \)
Complete discrete-time model from N Tubes

Discrete-time pressure-to-volume velocity frequency response:

\[ H(z) = R(z)V(Z) \]

where \( R(z) \approx 1 - \alpha z^{-1} \) and \( V(z) \) is an all-pole model.
And for the speech signal (voiced case):

\[ X(z) = A_v G(z)H(z) \]

with \( A_v \) to control loudness and \( G(z) \) being the z-transform of the glottal flow input.

or

\[ X(z) = A_v G(z) \frac{1 - \alpha z^{-1}}{1 + \sum_{k=1}^{N} a_k z^{-k}} \]
Complete discrete-time model from N Tubes

Discrete-time pressure-to-volume velocity frequency response:

\[ H(z) = R(z)V(Z) \]

where \( R(z) \approx 1 - \alpha z^{-1} \) and \( V(z) \) is an all-pole model.

And for the speech signal (voiced case):

\[ X(z) = A_v G(z)H(z) \]

with \( A_v \) to control loudness and \( G(z) \) being the z-transform of the glottal flow input.

or

\[ X(z) = A_v G(z) \frac{1 - \alpha z^{-1}}{1 + \sum_{k=1}^{N} a_k z^{-k}} \]
A typical glottal flow waveform over one cycle is modeled as:

\[ g[n] = (b^{-n} u[-n]) \ast (b^{-n} u[-n]) \]

which has as z-transform:

\[ G(z) = \frac{1}{(1 - \beta z)^2} \]

So for a voiced frame:

\[ X(z) = A_v \frac{(1 - az^{-1})}{(1 - bz)^2(1 + \sum_{k=1}^{N} a_k z^{-k})} \]
A typical glottal flow waveform over one cycle is modeled as:

\[ g[n] = (b^{-n}u[-n]) \ast (b^{-n}u[-n]) \]

which has as z-transform:

\[ G(z) = \frac{1}{(1 - \beta z)^2} \]

So for a \textit{voiced} frame:

\[ X(z) = A_v \frac{(1 - az^{-1})}{(1 - bz)^2(1 + \sum_{k=1}^{N} a_k z^{-k})} \]
Modeling other states

- For noisy inputs:

\[ X(z) = A_n U(z) V(z) R(z) \]

- For impulsive sounds:

\[ X(z) = A_i V(z) R(z) \]

- Being more general:

\[ X(z) = A \frac{(1 - az^{-1}) \prod_{k=1}^{M_i} (1 - c_kz^{-1}) \prod_{k=1}^{M_o} (1 - d_kz)}{(1 - bz)^2 \left(1 - \sum_{k=1}^{N} a_kz^{-k}\right)} \]
Modeling other states

- For noisy inputs:

\[ X(z) = A_n U(z) V(z) R(z) \]

- For impulsive sounds:

\[ X(z) = A_i V(z) R(z) \]

- Being more general:

\[ X(z) = A \frac{(1 - az^{-1}) \prod_{k=1}^{M_i} (1 - c_k z^{-1}) \prod_{k=1}^{M_o} (1 - d_k z)}{(1 - b z)^2 (1 - \sum_{k=1}^{N} a_k z^{-k})} \]
Modeling other states

- For noisy inputs:
  \[ X(z) = A_n U(z) V(z) R(z) \]

- For impulsive sounds:
  \[ X(z) = A_i V(z) R(z) \]

- being more general:
  \[ X(z) = A (1 - a z^{-1}) \prod_{k=1}^{M_i} (1 - c_k z^{-1}) \prod_{k=1}^{M_o} (1 - d_k z) \]
  \[ \quad \frac{1}{(1 - b z)^2 (1 - \sum_{k=1}^{N} a_k z^{-k})} \]
Outline

1 Physics of sound
2 Uniform Tube Model
3 Concatenating N uniform tubes
4 Glottal Flow Derivative
5 Vocal Fold/Vocal Tract Interaction
6 Acknowledgments
7 References
Glottal Flow Derivative

Since speech signals, $x(t)$ can be obtained in general by:

$$x(t) \approx A \frac{d}{dt} [u_g(t) \ast v(t)]$$

and because:

$$A \frac{d}{dt} [u_g(t) \ast v(t)] = A \left[ \frac{d}{dt} u_g(t) \right] \ast v(t)$$

we usually consider the derivative $\frac{d}{dt} u_g(t)$ as input to the system, which is referred to as Glottal Flow Derivative.
**Glottal flow and its derivative**

![Diagram of glottal flow and its derivative](image_url)
Ripple in the glottal flow derivative?
Regarding the first formant [2]
Most, if not all, figures in this lecture are coming from the book:

**T. F. Quatieri:** Discrete-Time Speech Signal Processing, principles and practice
2002, Prentice Hall

and have been used after permission from Prentice Hall

C. Jankowski, *Fine Structure Features for Speaker Identification.*