CS578- Speech Signal Processing

Lecture 3: Acoustics of Speech Production

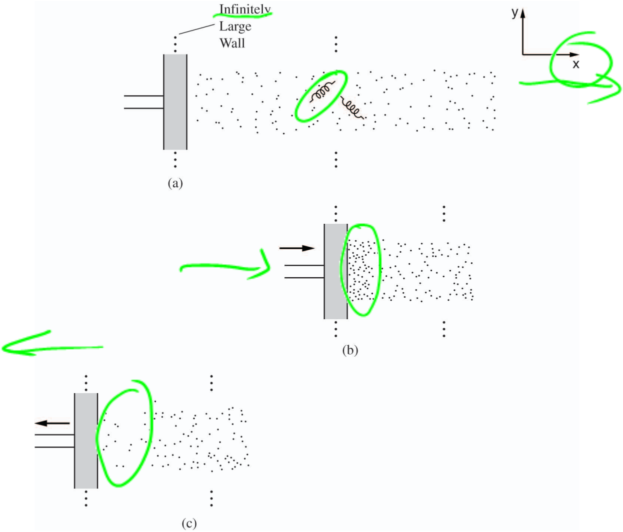
Yannis Stylianou

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

Univ. of Crete

Outline

Compression and rarefaction of air particles



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, λ
- **Frequency:** number of cycles of compressions per second, f
- **Speed of sound:** $c = f\lambda$ (at sea level and at $70°F$, $c = 344m/sec$)
- **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, λ
- **Frequency:** number of cycles of compressions per second, f
- **Speed of sound:** $c = f\lambda$ (at sea level and at $70°F$, $c = 344m/sec$)
- **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, λ
- **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, λ
- **Frequency:** number of cycles of compressions per second, f
- **Speed of sound:** $c = f\lambda$ (at sea level and at $70^\circ F$, $c = 344m/sec$)
- **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

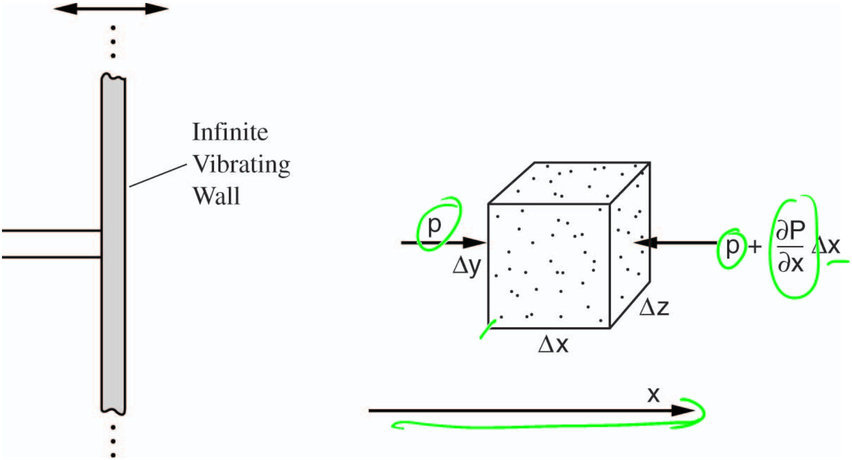
Some definitions

- **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, λ
- **Frequency:** number of cycles of compressions per second, f
- **Speed of sound:** $c = f\lambda$ (at sea level and at $70°F$, $c = 344m/sec$)
- **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, λ
- **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

Cube configuration



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: $2\ 10^{-5}$ *newtons*/m^2
  - ▷ Threshold of pain: 20 *newtons*/m^2
- $v(x, t)$ fluctuation of particles' velocity about zero average velocity.
- $\rho(x, t)$ fluctuation of particles' density about an average density ρ_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: $2\ 10^{-5}$ *newtons*/m^2
  ▷ Threshold of pain: 20 *newtons*/m^2
- $\upsilon(x, t)$ fluctuation of particles' velocity about zero average velocity.
- $\rho(x, t)$ fluctuation of particles' density about an average density ρ_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: $2\ 10^{-5}$ *newtons*/m^2
  - ▷ Threshold of pain: 20 *newtons*/m^2
- $\upsilon(x,t)$ fluctuation of particles' velocity about zero average velocity.
- $\rho(x,t)$ fluctuation of particles' density about an average density ρ_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:

$$
\begin{aligned}
-\frac{\partial p}{\partial x} &= \rho\frac{\partial v}{\partial t} \\
-\frac{\partial p}{\partial t} &= \rho c^2\frac{\partial v}{\partial x}
\end{aligned}
$$

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:

$$
\begin{aligned}
-\frac{\partial p}{\partial x} &= \rho \frac{\partial v}{\partial t} \\
-\frac{\partial p}{\partial t} &= \rho c^2 \frac{\partial v}{\partial x}
\end{aligned}
$$

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:

$$
\begin{aligned}
-\frac{\partial p}{\partial x} &= \rho \frac{\partial v}{\partial t} \\
-\frac{\partial p}{\partial t} &= \rho c^2 \frac{\partial v}{\partial x}
\end{aligned}
$$

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:

$$\begin{aligned} -\frac{\partial p}{\partial x} &= \rho \frac{\partial v}{\partial t} \\ -\frac{\partial p}{\partial t} &= \rho c^2 \frac{\partial v}{\partial x} \end{aligned}$$

The Wave Equation

Under the assumptions:

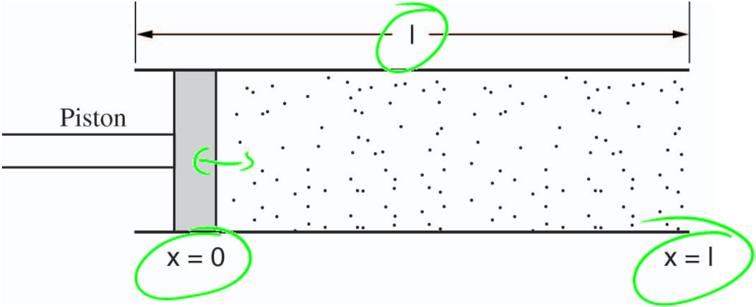
- If there is no friction of air particles in the cube with those outside the cube (no *viscosity*),
- Cube is very small,
- The density of air particles is constant in the cube (i.e.,$\rho_0 = \rho$)

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

$$\begin{aligned} -\frac{\partial p}{\partial x} &= \rho \frac{\partial v}{\partial t} \\ -\frac{\partial p}{\partial t} &= \rho c^2 \frac{\partial v}{\partial x} \end{aligned}$$

Outline

Lossless Case of cross section A



$$
\begin{aligned}
-\frac{\partial p}{\partial x} &= \frac{\rho}{A}\frac{\partial u}{\partial t} \\
-\frac{\partial p}{\partial t} &= \frac{\rho c^2}{A}\frac{\partial u}{\partial x}
\end{aligned}
$$

where $u(x,t) = A\upsilon(x,t)$

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 $x = 0$: $u(0, t) = U_g(\Omega)e^{j\Omega t}$

▷ Volume velocity:

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

▷ (Incremental) Pressure:

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

where $U_g(\Omega)e^{j\Omega t}$ denotes volume velocity at $x = 0$

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 $x = 0$: $u(0, t) = U_g(\Omega)e^{j\Omega t}$

▷ Volume velocity:

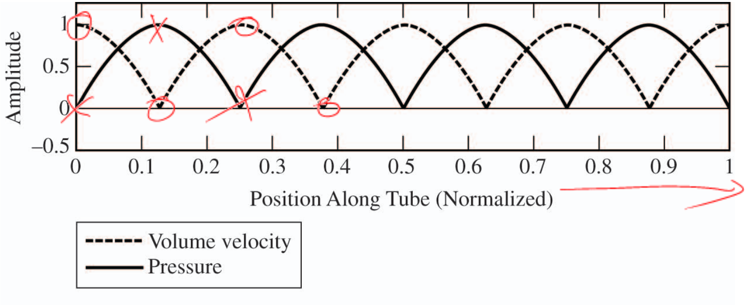
$$u(x, t) = \frac{\cos\left[\Omega(l - x)/c\right]}{\cos\left(\Omega\, l/c\right)} U_g(\Omega)e^{j\Omega t}$$

▷ (Incremental) Pressure:

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

where $U_g(\Omega)e^{j\Omega t}$ denotes volume velocity at $x = 0$

Velocity and Pressure are "orthogonal"



Input/Output volume velocity

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= 35cm, c = 350 m/s, then $f_k = 250, 750, 1250, \cdots$ Hz.

Input/Output volume velocity

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:

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

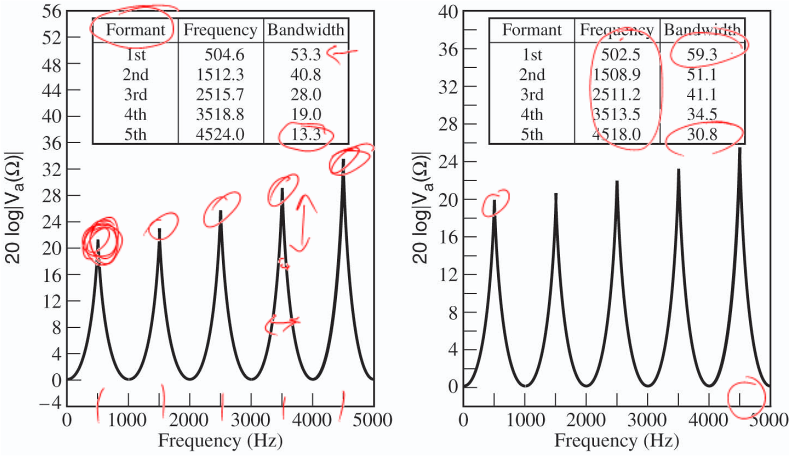
providing resonances of infinite amplitudes at frequencies:

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

Example: if l= 35cm, c = 350 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]:



Formant	Frequency	Bandwidth
1st	504.6	53.3
2nd	1512.3	40.8
3rd	2515.7	28.0
4th	3518.8	19.0
5th	4524.0	13.3

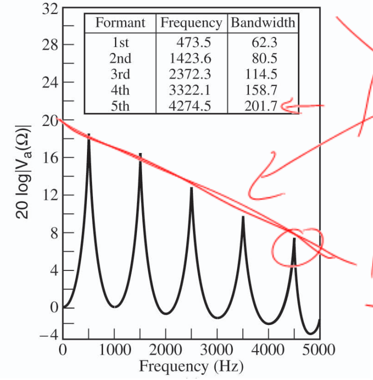
Formant	Frequency	Bandwidth
1st	502.5	59.3
2nd	1508.9	51.1
3rd	2511.2	41.1
4th	3513.5	34.5
5th	4518.0	30.8

Uniform tube: being more realistic

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]:

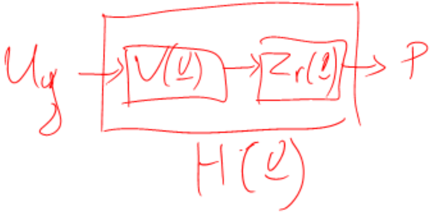


Formant	Frequency	Bandwidth
1st	473.5	62.3
2nd	1423.6	80.5
3rd	2372.3	114.5
4th	3322.1	158.7
5th	4274.5	201.7

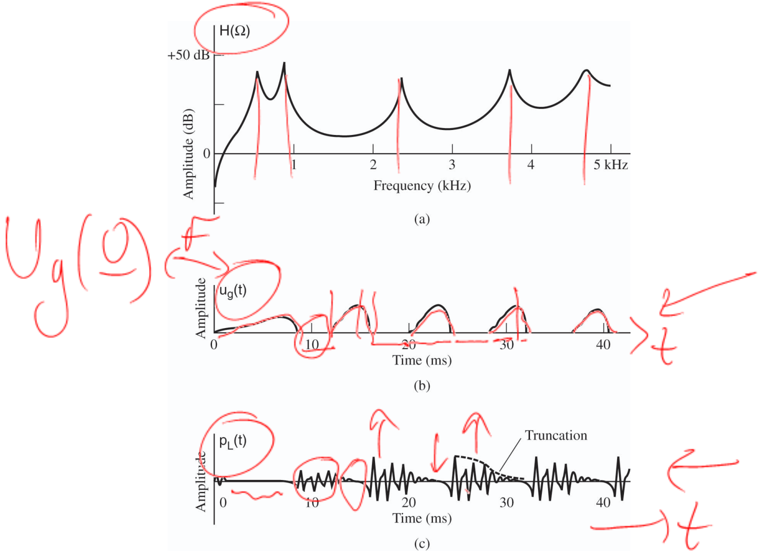
Pressure-to-volume velocity frequency response

Since we measure pressure at the lips:

$$H(\Omega) = \frac{P(l,\Omega)}{U_g(\Omega)} = Z_r(\Omega)V(\Omega)$$

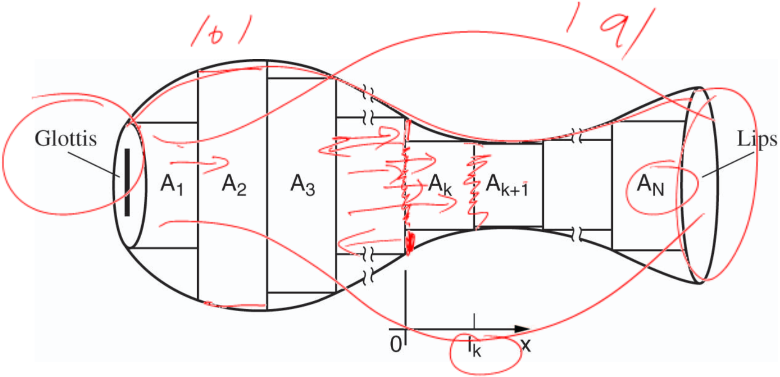


Numerical simulations for /o/[1]



Outline

Concatenating lossless Uniform Tubes



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:

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

- 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:

$$H(\Omega + \frac{2\pi}{2\tau}) = H(\Omega)$$

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 $\Delta x = \frac{l}{N}$

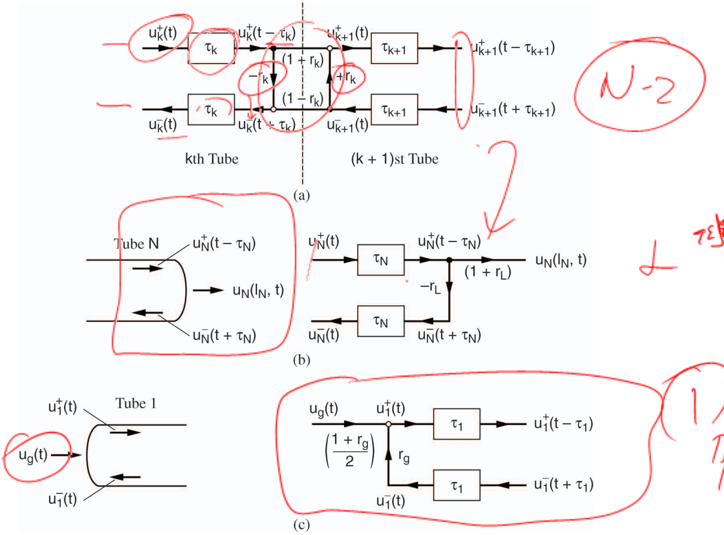
- Frequency response:

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

- Observe that:

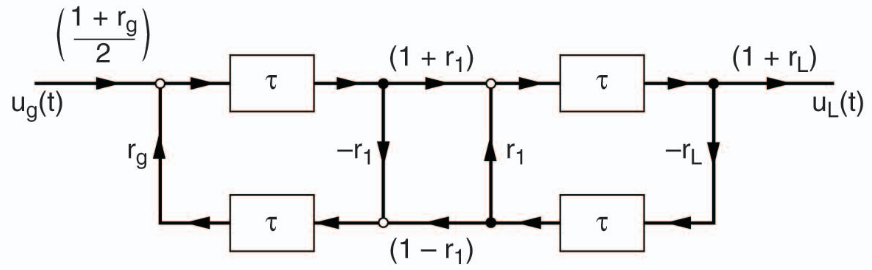
$$H(\Omega + \frac{2\pi}{2\tau}) = H(\Omega)$$

Signal flow graphs



(a) two concatenated tubes, (b) lip boundary condition, (c) glottal boundary condition

For a lossless two-tube model



(a)

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

$$V(s) = \frac{be^{-s2\tau}}{1 + a_1 e^{-s2\tau} + a_2 e^{-s4\tau}}$$

Laplace

$s = \sigma + j\Omega$

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 $z = e^{s2\tau}$, then:

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

- **N cubes:**

$$V(z) = \frac{Az^{-N/2}}{1 + \sum_{k=1}^{N} a_k z^{-k}}$$

Discrete-time lossless models

- **Two cubes:** By setting $z = e^{s2\tau}$, then:

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

- **N cubes:**

$$V(z) = \frac{Az^{-N/2}}{1 + \sum_{k=1}^{N} a_k z^{-k}}$$

Choosing the number of tube elements

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

Answer: N = 10

Choosing the number of tube elements

Question:
If a vocal tract has length $l = 17.5$ cm and the speed of sound $c = 350$ m/s, how many tubes, N, do we need to cover a bandwidth of 5000 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

$$X(z) = A_v G(z) \frac{1 - \alpha z^{-1}}{1 + \sum_{k=1}^{N} a_k z^{-k}}$$

Glottal waveform model

A typical glottal flow waveform over one cycle is modeled as:

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

which has as z-transform:

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

So for a *voiced* frame:

$$X(z) = A_v \frac{(1-az^{-1})}{(1-bz)^2(1+\sum_{k=1}^{N} a_k z^{-k})}$$

Glottal waveform model

A typical glottal flow waveform over one cycle is modeled as:

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

which has as z-transform:

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

So for a *voiced* frame:

$$X(z) = A_v \frac{(1-az^{-1})}{(1-bz)^2(1+\sum_{k=1}^{N} a_k z^{-k})}$$

$$A_v\, G(z) \cdot \frac{R(z)}{V(z)}$$

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:**

$$X(z) = A \frac{(1 - az^{-1}) \prod_{k=1}^{M_i} (1 - c_k z^{-1}) \prod_{k=1}^{M_o} (1 - d_k z)}{(1 - bz)^2 \left(1 - \sum_{k=1}^{N} a_k z^{-k}\right)}$$

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:**

$$X(z) = A \frac{(1 - az^{-1}) \prod_{k=1}^{M_i} (1 - c_k z^{-1}) \prod_{k=1}^{M_o} (1 - d_k z)}{(1 - bz)^2 \left(1 - \sum_{k=1}^{N} a_k z^{-k}\right)}$$

Modeling other states

- **For noisy inputs:**

$$X(z) = A_n U(z) V(z) R(z)$$

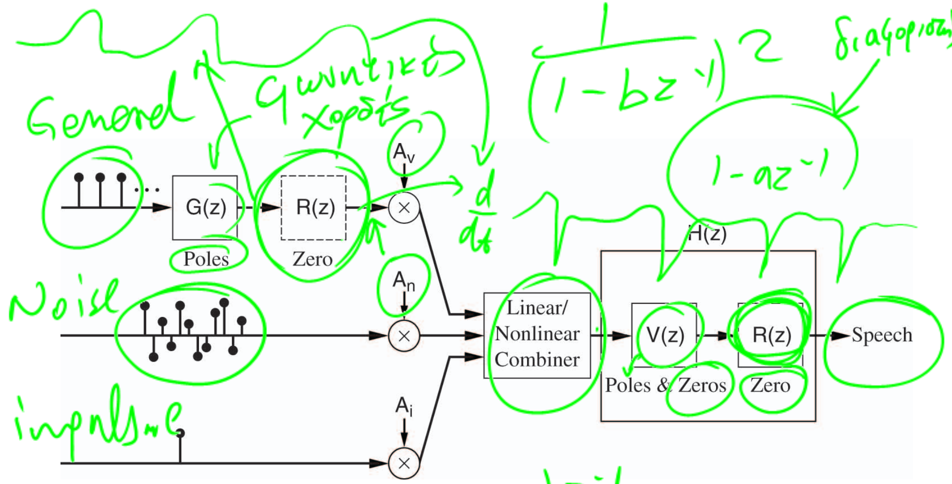
- **For impulsive sounds:**

$$X(z) = A_i V(z) R(z)$$

- **being more general:**

$$X(z) = A \frac{(1 - a z^{-1}) \prod_{k=1}^{M_i} (1 - c_k z^{-1}) \prod_{k=1}^{M_o} (1 - d_k z)}{(1 - bz)^2 \left(1 - \sum_{k=1}^{N} a_k z^{-k}\right)}$$

An Overview then



$X(z)\cdot(1-az^{-1}) \xrightarrow{z^{-1}} x[n]-ax[n-1]$

$X(z)(1-z^{-1}) \xrightarrow{z^{-1}} x[n]-x[n-1]$

$\frac{dx}{dt}$

Outline

Glottal Flow Derivative

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

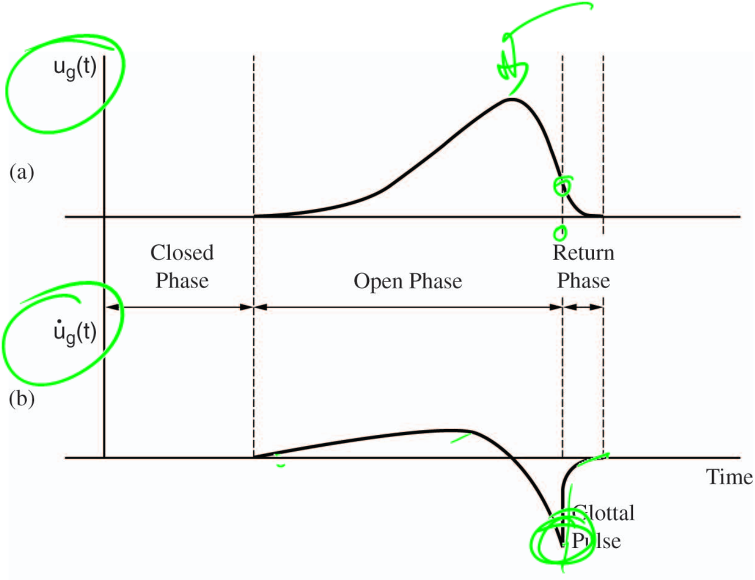
$$x(t) \approx A \frac{d}{dt} \left[u_g(t) \star v(t) \right]$$

and because:

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

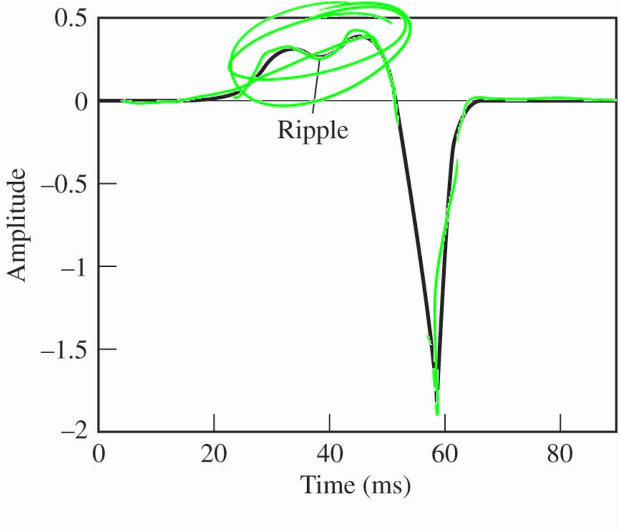
we usually consider the derivative $\frac{d}{dt} u_g(t)$ as input to the system, which is referred to as *Glottal Flow Derivative*

Glottal flow and its derivative

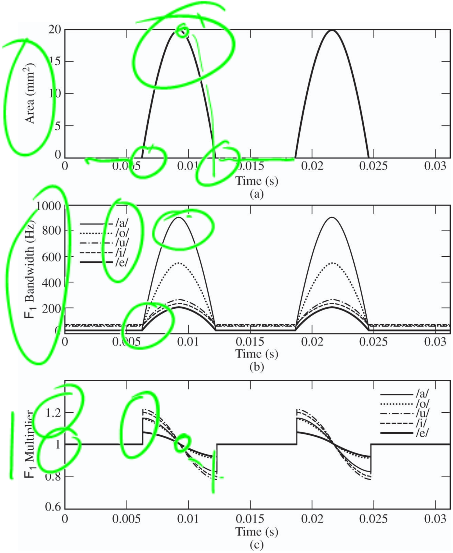


Outline

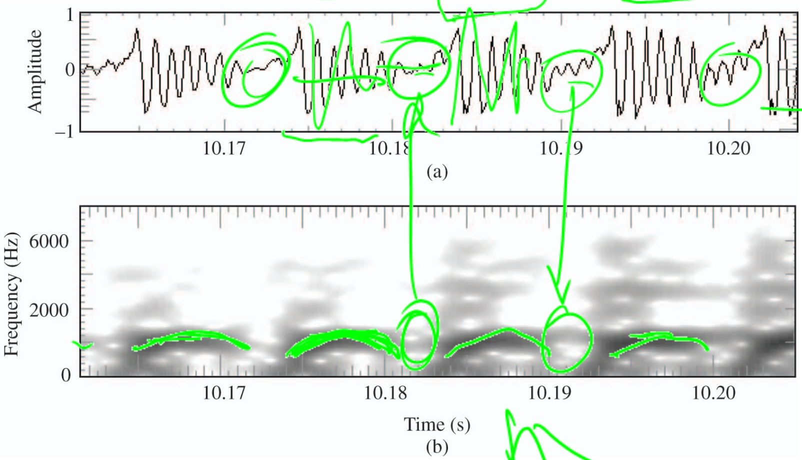
Ripple in the glottal flow derivative?



Regarding the first formant [2]



Truncation effect - again



Outline

Acknowledgments

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

Outline

M. Portnoff, *A Quasi-One-Dimensional Digital Simulation for the Time-Varying Vocal Tract*.
PhD thesis, Massachusetts Institue of Technology, May 1973.

C. Jankowski, *Fine Structure Features for Speaker Identification*.
PhD thesis, Massachusetts Institute of Technology, Dept. of EE and CS, June 1996.