Lecture 2

Sound Waves in a Tube

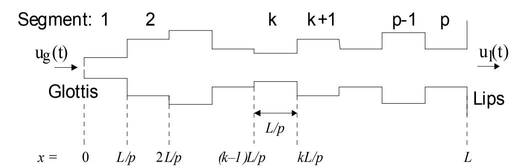
– Derive a theoretical model of how sound waves are affected by the vocal tract
– Describe a model for lip radiation
– Describe a model for the pulsating glottal waveform during voiced speech
– Assemble the components of a simple speech synthesiser

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

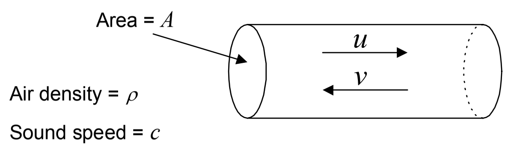
$$\approx 20\sqrt{\text{Absolute Temperature}} \approx 340 \text{ m/s}$$

– 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:

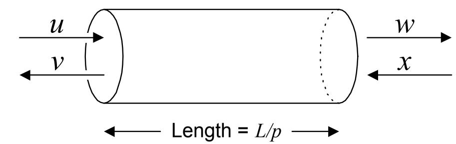
- Volume flow ≈ Current, Pressure ≈ Voltage
- Acoustic Impedance of tube = $\rho c/A$

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 $u(t) = w\left(t + \frac{L}{cp}\right)$

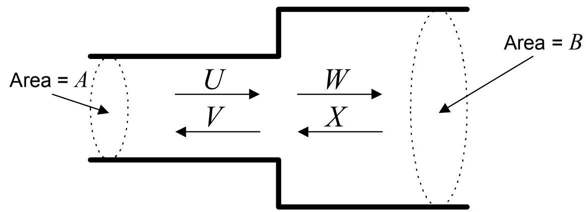
Segment length chosen to correspond to half a sample period. If we take *z*-transforms, this time delay corresponds to multiplying by $z^{-1/2}$:

$$V(z) = z^{-1/2} X(z) \quad \text{and} \quad U(z) = z^{+1/2} W(z)$$

In matrix form:

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} z^{+1/2} & 0 \\ 0 & z^{-1/2} \end{pmatrix} \begin{pmatrix} W \\ X \end{pmatrix} = z^{+1/2} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \begin{pmatrix} W \\ X \end{pmatrix}$$

Segment Junction



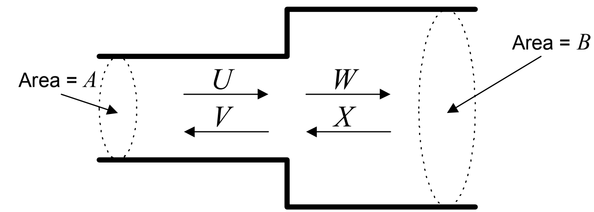
Flow Continuity: $$(U-V)=(W-X)$$

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

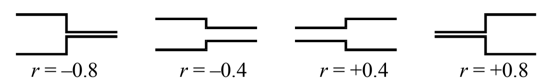
Reflection Coefficients



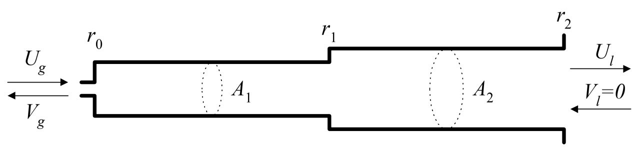
Define the reflection coefficient to be $r=\dfrac{B-A}{B+A}$

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

Reflection coefficients always lie in the range ±1:



2-Segment Vocal Tract

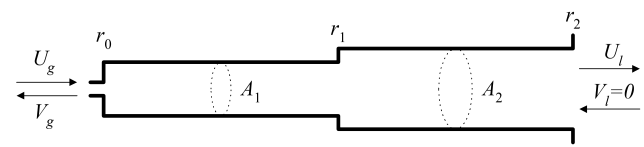


$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}$$

$$\begin{pmatrix} U_g \\ V_g \end{pmatrix} = \frac{1}{1+r_0}\begin{pmatrix} 1 & -r_0 \\ -r_0 & 1 \end{pmatrix} \times z^{1/2}\begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \times \frac{1}{1+r_1}\begin{pmatrix} 1 & -r_1 \\ -r_1 & 1 \end{pmatrix} \times z^{1/2}\begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \times \frac{1}{1+r_2}\begin{pmatrix} 1 & -r_2 \\ -r_2 & 1 \end{pmatrix}\begin{pmatrix} U_l \\ 0 \end{pmatrix}$$

- 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:

$$\begin{pmatrix} U_g \\ V_g \end{pmatrix} = \frac{z^{+1}}{\prod_{k=0}^{2}(1+r_k)}\begin{pmatrix} 1 + (r_0 r_1 + r_1 r_2)z^{-1} + r_0 r_2 z^{-2} \\ -r_0 - (r_1 + r_0 r_1 r_2)z^{-1} - r_2 z^{-2} \end{pmatrix} U_l$$

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:

$$\frac{U_l}{U_g} = \frac{\prod_{k=0}^{2}(1+r_k) \times z^{-1}}{1 + (r_0 r_1 + r_1 r_2)z^{-1} + r_0 r_2 z^{-2}} = \frac{Gz^{-1}}{1 + (r_0 r_1 + r_1 r_2)z^{-1} + r_0 r_2 z^{-2}} = \frac{Gz^{-1}}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

p-segment Vocal Tract

Note that: $\frac{1}{1+r}\begin{pmatrix}1 & -r\\ -r & 1\end{pmatrix}\times z^{1/2}\begin{pmatrix}1 & 0\\ 0 & z^{-1}\end{pmatrix}=\frac{z^{1/2}}{1+r}\begin{pmatrix}1 & -rz^{-1}\\ -r & z^{-1}\end{pmatrix}$

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

$$\begin{pmatrix}U_g\\ V_g\end{pmatrix}=\frac{z^{1/2 p}}{\prod_{k=0}^{p}(1+r_k)}\prod_{k=0}^{p-1}\begin{pmatrix}1 & -r_k z^{-1}\\ -r_k & z^{-1}\end{pmatrix}\times\begin{pmatrix}1\\ -r_p\end{pmatrix}U_l$$

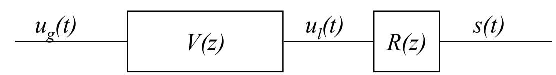
This results in a transfer function of the form:

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

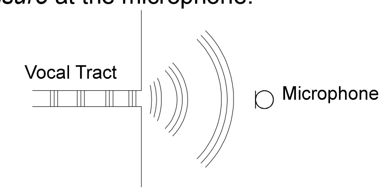
Where:

- G is a gain term
- $z^{-1/2 p}$ 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.



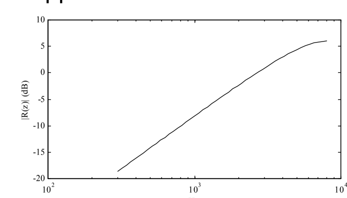
For a lip-opening area of A, acoustic theory predicts a 1st-order high-pass response with a corner frequency of:

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

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

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

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

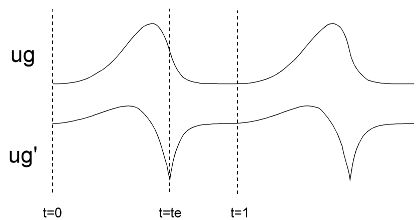


Spectrum of Glottal Waveform

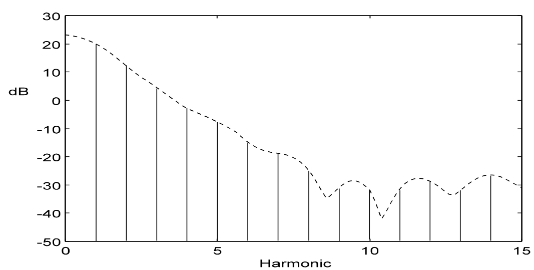
"LF Model" (Liljencrants & Fant)

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

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

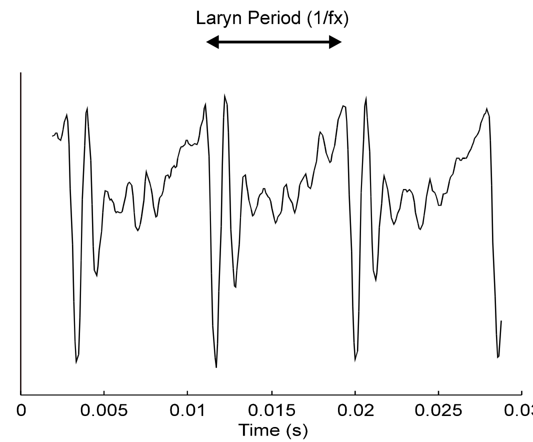


Line Spectrum of ug (approx –12 dB/octave):



Vowel Waveform

Vowel /ɑ/ from "part"



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

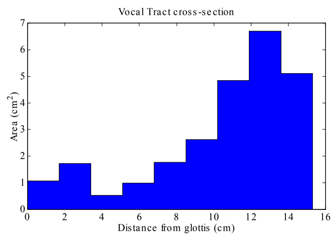
There is not necessarily any relation between the larynx frequency and the vocal tract resonances.

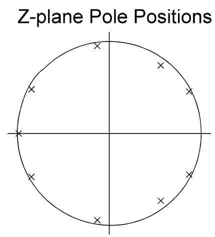
Resonances at a multiple of the larynx frequency will be louder (good for singers)

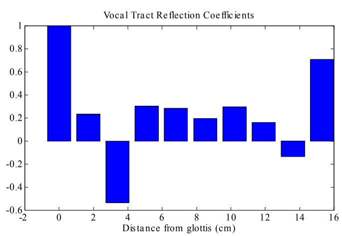
TUBE.PPT(4/15/2002) 2.13

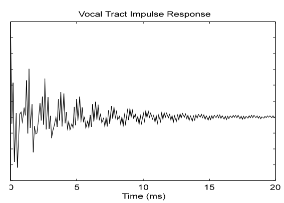
Vocal Tract Shape and Response

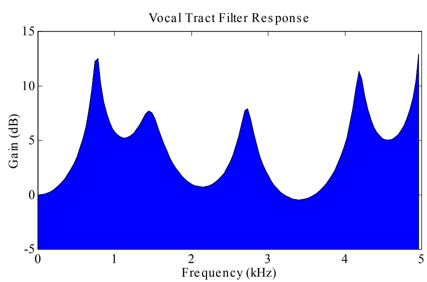
Example: /ɑ/ vowel ("part")











TUBE.PPT(4/15/2002) 2.14

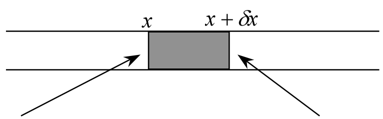
Appendix

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:



Pressure $= p$

Velocity $= v = \dfrac{u}{A}$

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

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> $V = A \times \delta x$

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

<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 ÷ (6×10^{23})
R = gas constant = 8.314 J / (K · mol)
T = Temperature (° K)
ρ = density (≈ 1.225 kg / m^3)
M = molecular weight of air = 0.029 kg / mol
γ = specific heat ratio of air = 1.4

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

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

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

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.