ECE 598: The Speech Chain

Lecture 8: Formant Transitions;
Vocal Tract Transfer Function
Today

- **Perturbation Theory:**
  - A different way to estimate vocal tract resonant frequencies, useful for consonant transitions

- **Syllable-Final Consonants:** Formant Transitions

- **Vocal Tract Transfer Function**
  - Uniform Tube (Quarter-Wave Resonator)
  - During Vowels: All-Pole Spectrum
    - Q
    - Bandwidth

- **Nasal Vowels:** Sum of two transfer functions gives spectral zeros
Topic #1: Perturbation Theory
Perturbation Theory
(Chiba and Kajiyama, *The Vowel*, 1940)

A(x) is constant everywhere, except for one small perturbation.

Method:
1. Compute formants of the “unperturbed” vocal tract.
2. Perturb the formant frequencies to match the area perturbation.
Conservation of Energy Under Perturbation

Perturbation of Area

\[ A(x) = A_0 + \alpha(x), \quad |\alpha(x)| \ll A_0 \]

Potential Energy at \( \Omega_n \):

\[ E_P = \int_0^L A_0 \frac{1}{\rho c^2} |P_n(x)|^2 \, dx \]

Increasing \( A(x) \) Increases \( E_P \):

\[ E_P = \int_0^L (A_0 + \alpha(x)) \frac{1}{\rho c^2} |P_n(x)|^2 \, dx \]
Conservation of Energy Under Perturbation

Kinetic Energy at $\Omega_n$:

$$E_K = \int_0^L A_0 \rho |U_n(x)|^2 dx$$

Increasing $A(x)$ Decreases $E_K$:

$$(A_0 + \alpha(x)) \tilde{U}_n(x) = A_0 U_n(x)$$

$$E_K = \int_0^L (A_0 + \alpha(x)) \rho |\tilde{U}_n(x)|^2 dx$$

$$= \int_0^L \frac{A_0^2}{A_0 + \alpha(x)} \rho |U_n(x)|^2 dx$$

$\delta_n$ Preserves the Balance of Kinetic and Potential Energies:

$$\Omega_n = \Omega_{n,0} + \delta_n, \quad |\delta_n| \ll \Omega_{n,0}$$

$$\delta_n \approx \frac{\pi c}{2L} \int_0^L \frac{\alpha(x)}{A_0} \left( |\rho c U_n(x)|^2 - |P_n(x)|^2 \right) dx$$
“Sensitivity” Functions

Localized Area Change:

\[ \alpha(x) = \alpha_\xi \delta(x - \xi) \]

Sensitivity Function Measures Dependence of Formant on an Area Change:

\[ S_n(\xi) = \frac{\partial \Omega_n}{\partial \log A(\xi)} \]

\[ S_n(\xi) \approx \frac{\delta_n}{\alpha_\xi / A_0} \]
Sensitivity Functions for the Quarter-Wave Resonator (Lips Open)

- Note: low F3 of /er/ is caused in part by a side branch under the tongue – perturbation alone is not enough to explain it.
Sensitivity Functions for the Half-Wave Resonator (Lips Rounded)

- Note: high F3 of /l/ is caused in part by a side branch above the tongue – perturbation alone is not enough to explain it.
Formant Frequencies of Vowels

Inter-Subject Average Formants
Topic #2: Formant Transitions, Syllable-Final Consonant
Events in the Closure of a Nasal Consonant

Formant Transitions

Vowel Nasalization

Nasal Murmur
Formant Transitions: A Perturbation Theory Model

Unperturbed Standing Wave Functions:

\[ p(x, t) = \sum_{n=1}^{\infty} P_n(x) \cos(\Omega_n t + \theta_n) \]

\[ u(x, t) = \sum_{n=1}^{\infty} U_n(x) \cos(\Omega_n t + \theta_n) \]

Perturbation Theory:

\[ \frac{\partial \Omega_n}{\partial t} \approx \frac{\pi c}{2L} \int_{0}^{L} \left( \frac{\partial \log A(x, t)}{\partial t} \right) \left( |\rho c U_n(x)|^2 - |P_n(x)|^2 \right) \, dx \]

Sensitivity Function:

\[ A(x) \approx A_0 + \alpha_\xi \delta(x - \xi) \]

\[ \frac{\partial \Omega_n}{\partial t} \approx -s_n(\xi) \left( \frac{\partial \log A(\xi)}{\partial t} \right) \]
Formant Transitions: Labial Consonants

First Sensitivity Function: $-dF_1/d \log A(x)$

Second Sensitivity Function: $-dF_2/d \log A(x)$

Third Sensitivity Function: $-dF_3/d \log A(x)$
Formant Transitions: Alveolar Consonants

First Sensitivity Function: $-dF_1/d \log A(x)$

Second Sensitivity Function: $-dF_2/d \log A(x)$

Third Sensitivity Function: $-dF_3/d \log A(x)$
Formant Transitions: Post-alveolar Consonants

First Sensitivity Function: $-\frac{dF1}{d \log A(x)}$

Second Sensitivity Function: $-\frac{dF2}{d \log A(x)}$

Third Sensitivity Function: $-\frac{dF3}{d \log A(x)}$

"the shoe"

"zsazsa"
Formant Transitions: Velar Consonants

First Sensitivity Function: $-\frac{dF1}{d \log A(x)}$

Second Sensitivity Function: $-\frac{dF2}{d \log A(x)}$

Third Sensitivity Function: $-\frac{dF3}{d \log A(x)}$
Topic #3: Vocal Tract Transfer Functions
Transfer Function

- “Transfer Function” $T(\omega) = \text{Output}(\omega)/\text{Input}(\omega)$
- In speech, it’s convenient to write $T(\omega) = U_L(\omega)/U_G(\omega)$
  - $U_L(\omega) = \text{volume velocity at the lips}$
  - $U_G(\omega) = \text{volume velocity at the glottis}$
  - $T(0) = 1$
- Speech recorded at a microphone = pressure
  - $P_R(\omega) = R(\omega)T(\omega)U_G(\omega)$
  - $R(\omega) = j\rho f/r = \text{“radiation characteristic”}$
    - $\rho = \text{density of air}$
    - $r = \text{distance to the microphone}$
    - $f = \text{frequency in Hertz}$
Transfer Function of an Ideal Uniform Tube

- Ideal Terminations:
  - Reflection coefficient at glottis: zero velocity, $\gamma = 1$
  - Reflection coefficient at lips: zero pressure, $\gamma = -1$
  - Obviously, this is an approximation, but it gives...

$$T(\omega) = 1/\cos(\omega L/c)$$

$$\frac{\omega_1^2 \omega_2^2 \omega_3^2 \ldots}{\ldots(\omega+\omega_3)(\omega+\omega_2)(\omega+\omega_1)(\omega-\omega_1)(\omega-\omega_2)(\omega-\omega_3)\ldots}$$

$$\omega_n = n\pi c/L - \pi c/2L$$

$$F_n = nc/2L - c/4L$$
Transfer Function of an Ideal Uniform Tube

$20 \log_{10} |T(f)|$, Ideal Quarter-Wave Resonator

Peaks are actually infinite in height (figure is clipped to fit the display)
Transfer Function of a Non-Ideal Uniform Tube

- Almost ideal terminations:
  - At glottis: velocity almost zero, \( \gamma \approx 1 \)
  - At lips: pressure almost zero, \( \gamma \approx -1 \)

\[
T(\omega) = \frac{1}{(j/Q + \cos(\omega L/c))}
\]

... at \( F_n = nc/2L - c/4L, \ldots \)

\[
T(2\pi F_n) = -jQ
\]

\[
20\log_{10}|T(2\pi F_n)| = 20\log_{10}Q
\]
Transfer Function of a Non-Ideal Uniform Tube

$20 \log_{10} |T(f)|$, Quarter-Wave Resonator, $Q=10$
Transfer Function of a Vowel: Height of First Peak is $Q_1 = F_1/B_1$

$$T(\omega) = \prod_{n=1}^{\infty} \frac{(2\pi F_n)^2 + (\pi B_n)^2}{(j\omega + j2\pi F_n + \pi B_n)(j\omega - j2\pi F_n + \pi B_n)}$$

$$T(2\pi F_1) \approx \frac{(2\pi F_1)^2}{(j4\pi F_1 \pi B_1)} = -jF_1/B_1$$

Call $Q_n = F_n/B_n$

$$T(2\pi F_1) \approx -jQ_1$$

$$20\log_{10}|T(2\pi F_1)| \approx 20\log_{10}Q_1$$
Transfer Function of a Vowel: Bandwidth of a Peak is $B_n$

$$T(\omega) = \prod_{n=1}^{\infty} \frac{(2\pi F_n)^2 + (\pi B_n)^2}{(j\omega + j2\pi F_n + \pi B_n)(j\omega - j2\pi F_n + \pi B_n)}$$

$$T(2\pi F_1 + \pi B_1) \approx \frac{(2\pi F_1)^2}{(j4\pi F_1)(\pi B_1 + \pi B_1)} = -jQ_1/2$$

At $f = F_1 + 0.5B_n$, 

$$|T(\omega)| = 0.5Q_n$$

$$20\log_{10}|T(\omega)| = 20\log_{10}Q_1 - 3\text{dB}$$
Amplitudes of Higher Formants: Include the Rolloff

\[ T(\omega) = \prod_{n=1}^{\infty} \frac{(2\pi F_n)^2 + (\pi B_n)^2}{(j \omega + j2\pi F_n + \pi B_n)(j \omega - j2\pi F_n + \pi B_n)} \]

At \( f \) above \( F_1 \)
\[ T(2\pi f) \approx \frac{F_1}{f} \]
\[ T(2\pi F_2) \approx (-jF_2/B_2)(F_1/F_2) \]
\[ 20\log_{10}|T(2\pi F_2)| \approx 20\log_{10}Q_2 - 20\log_{10}(F_2/F_1) \]

1/f Rolloff: 6 dB per octave (per doubling of frequency)
Vowel Transfer Function: Synthetic Example

\[ L_1 = 20 \log_{10} \left( \frac{500}{80} \right) = 16 \text{dB} \]

\[ L_2 = 20 \log_{10} \left( \frac{1500}{240} \right) - 20 \log_{10} \left( \frac{F_2}{F_1} \right) = 16 \text{dB} - 9.5 \text{dB} \]

\[ L_3 = 20 \log_{10} \left( \frac{2500}{600} \right) - 20 \log_{10} \left( \frac{F_3}{F_1} \right) - 20 \log_{10} \left( \frac{F_3}{F_2} \right) \]

\[ B_1 = 80 \text{Hz} \]

\[ B_2 = 240 \text{Hz} \]

\[ B_3 = 600 \text{Hz?} \]

(hard to measure because rolloff from \( F_1 \), \( F_2 \) turns the \( F_3 \) peak into a plateau)

\[ F_4 \text{ peak completely swamped by rolloff from lower formants} \]
Shorthand Notation for the Spectrum of a Vowel

\[ T(s) = \prod_{n=1}^{\infty} \frac{s_n s_n^*}{(s-s_n)(s-s_n^*)} \]

\[ s = j\omega \]
\[ s_n = -\pi B_n + j2\pi F_n \]
\[ s_n^* = -\pi B_n - j2\pi F_n \]
\[ s_n s_n^* = |s_n|^2 = (2\pi F_n)^2 + (\pi B_n)^2 \]

\[ T(0) = 1 \]
\[ 20 \log_{10} |T(0)| = 0 \text{dB} \]
Another Shorthand Notation for the Spectrum of a Vowel

\[ T(s) = \prod_{n=1}^{\infty} \frac{1}{(1-s/s_n)(1-s/s_n^*)} \]
Topic #4: Nasalized Vowels
Vowel Nasalization

Nasalized Vowel

Nasal Consonant
Nasalized Vowel

\[ P_R(\omega) = R(\omega)(U_L(\omega)+U_N(\omega)) \]

\[ U_N(\omega) = \text{Volume Velocity from Nostrils} \]

\[ P_R(\omega) = R(\omega)(T_L(\omega)+T_N(\omega))U_G(\omega) \]

\[ = R(\omega)T(\omega)U_G(\omega) \]

\[ T(\omega) = T_L(\omega) + T_N(\omega) \]
Nasalized Vowel

\[ T(s) = T_L(s) + T_N(s) \]

\[ = \frac{1}{(1-s/s_{Ln})(1-s/s_{Ln}^*)} + \frac{1}{(1-s/s_{Nn})(1-s/s_{Nn}^*)} \]

\[ = \frac{2(1-s/s_{Zn})(1-s/s_{Zn}^*)}{(1-s/s_{Ln})(1-s/s_{Ln}^*)(1-s/s_{Nn})(1-s/s_{Nn}^*)} \]

\[ 1/s_{Zn} = \frac{1}{2}(1/s_{Ln} + 1/s_{Nn}) \]

\[ s_{Zn} = n^{th} \text{ spectral zero} \]

\[ T(s) = 0 \text{ if } s = s_{Zn} \]
The "Pole-Zero Pair"

$$20\log_{10} T(\omega) =$$

$$20\log_{10} \left( \frac{1}{1-s/s_{Ln}}(1-s/s_{Ln}^*) \right)$$

$$+ 20\log_{10} \left( \frac{(1-s/s_{Zn})(1-s/s_{Zn}^*)}{(1-s/s_{Nn})(1-s/s_{Nn}^*)} \right)$$

= original vowel log spectrum

+ log spectrum of a pole-zero pair
Additive Terms in the Log Spectrum

\[ \log \left( \frac{1 - be^{-j\omega}}{(1 - \lambda e^{-j\omega})(1 - \eta e^{-j\omega})} \right) = \log \left( \frac{1}{1 - \lambda e^{-j\omega}} \right) + \log \left( \frac{1 - be^{-j\omega}}{1 - \eta e^{-j\omega}} \right) \]
Transfer Function of a Nasalized Vowel

![Graphs showing transfer functions and their sum](image-url)
Pole-Zero Pairs in the Spectrogram

spin
Summary

- **Perturbation Theory:**
  - Squeeze near a velocity peak: formant goes down
  - Squeeze near a pressure peak: formant goes up

- **Formant Transitions**
  - Labial closure: loci near 250, 1000, 2000 Hz
  - Alveolar closure: loci near 250, 1700, 3000 Hz
  - Velar closure: F2 and F3 come together ("velar pinch")

- **Vocal Tract Transfer Function**
  - \[ T(s) = \prod s_n s_n^* / (s-s_n) (s-s_n^*) \]
  - \[ T(\omega=2\pi F_n) = Q_n = F_n / B_n \]
  - 3dB bandwidth = \( B_n \) Hertz
  - \( T(0) = 1 \)

- **Nasal Vowels:**
  - Sum of two transfer functions gives a spectral zero between the oral and nasal poles
  - Pole-zero pair is a local perturbation of the spectrum