

ECE 537 Homework 6

Due: October 16, Urbana time

1. Consider a two-tube model of the vocal tract with $\ell_1 = \ell_2 = 8.85\text{cm}$, and assume that $c = 354\text{m/s}$ and $F_s = 4000$ samples/second, so that $\ell_1 F_s / c = \ell_2 F_s / c = 1$ sample. Suppose that the glottis has infinite impedance ($Z_g = \infty$) and the radiation impedance is zero ($Z_R = 0$). As a function of the reflection coefficient $r_1 = \frac{A_2 - A_1}{A_2 + A_1}$, find the system function $H(z) = U_L(z)/U_G(z)$, where $u_L(t) = u_2(\ell_2, t)$ is the volume velocity of air leaving the lips, and $z = e^{j\omega}$ with ω equal to the frequency measured in radians per sample. Note that there are no losses in this vocal tract model, so $H(z)$ should have four poles on the unit circle. Find the frequencies of both formants as functions of r_1 . Demonstrate that your result is sensible as $r_1 \rightarrow 1$ (a good model of /a/) and as $r_1 \rightarrow -1$ (a good model of /i/).

Solution: The specified impedances give $r_0 = r_2 = 1$. Taking the DTFT of the boundary conditions at the glottis, lips, and the boundary of the two tubes, we have

$$\begin{aligned} U_1^+(z) &= 2U_G(z) + U_1^-(z) \\ U_2^+(z) &= (1 + r_1)z^{-1}U_1^+(z) + r_1U_2^-(z) \\ zU_1^-(z) &= -r_1z^{-1}U_1^+(z) + (1 - r_1)U_2^-(z) \\ zU_2^-(z) &= -z^{-1}U_2^+(z) \end{aligned}$$

Setting $U_L(z) = z^{-1}U_2^+(z) - zU_2^-(z)$ and simplifying we find that

$$\frac{U_L(z)}{U_G(z)} = \frac{4(1 + r_1)z^{-2}}{1 + 2r_1z^{-2} + z^{-4}}$$

The roots of this transfer function are

$$\begin{aligned} z^2 &= -r_1 \pm j\sqrt{1 - r_1^2} \\ \angle z &= \pm \frac{1}{2} \left(\pi - \text{atan} \frac{\sqrt{1 - r_1^2}}{r_1} \right) \end{aligned}$$

As $r_1 \rightarrow 1$, $\angle z \rightarrow \pi \pm \epsilon$, and its square roots are both at frequencies of $\angle z \approx \frac{\pi \pm \epsilon}{2}$, which is 1000Hz if $F_s = 4000$. As $r_1 \rightarrow -1$, $\angle z \rightarrow 0 + 2\pi k$ for any integer k , and its square roots are at $\angle z = 0$ and $\angle z = \pi$, i.e., $F_1 = 0$ and $F_2 = 2000\text{Hz}$.

2. Suppose that $s[n]$ is well modeled by $s[n] = v[n] + a_1s[n-1] + a_2s[n-2]$ for an impulse train or white noise signal $v[n]$. Estimate the frequency and bandwidth of the first formant, in Hertz, as functions of a_1 , a_2 , and F_s . Assume that $a_1^2 + 4a_2 < 0$ and $|a_2| < 1$.

Solution: The transfer function is

$$\frac{S(z)}{V(z)} = \frac{1}{1 - a_1z^{-1} - a_2z^{-2}}$$

which has roots at

$$z = \frac{a_1}{2} \pm j\sqrt{-(a_1^2 + 4a_2)/4}$$

The formant frequency and bandwidth are

$$\begin{aligned}
 F_1 &= \frac{F_s}{2\pi} \angle z \\
 &= \frac{F_s}{2\pi} \operatorname{atan} \left(\frac{\sqrt{-(a_1^2 + 4a_2)}}{a_1} \right) \\
 B_1 &= \frac{F_s}{\pi} \ln |z| \\
 &= \frac{F_s}{\pi} \ln |a_2|
 \end{aligned}$$

3. Ignoring their perceptual weighting filter, Rose and Barnwell's self-excited LPC vocoder chose β_3 and T_0 to minimize $\mathcal{E}_{\text{Adaptive}}$, then chose β_2 and k to minimize $\mathcal{E}_{\text{Stochastic}}$, where $\hat{s}[n]$ is synthetic speech from previous frames, $c_k[n]$ is the k^{th} stochastic codevector, and

$$\begin{aligned}
 \mathcal{E}_{\text{Adaptive}} &= \sum_{n=0}^{N-1} (s[n] - \beta_3 \hat{s}[n - T_0])^2 \\
 \mathcal{E}_{\text{Stochastic}} &= \sum_{n=0}^{N-1} (s[n] - \beta_2 h[n] * c_k[n] - \beta_3 \hat{s}[n - T_0])^2
 \end{aligned}$$

Find functions $R_1[m]$, $P_1[m]$, $R_2[m]$, and $P_2[m]$ such that the minimizers of $\mathcal{E}_{\text{Adaptive}}$ and $\mathcal{E}_{\text{Stochastic}}$ can be written as:

$$\begin{aligned}
 T_0 &= \operatorname{argmax}_m \frac{R_1[m]}{P_1[m]}, \quad \beta_3 = \frac{R_1[T_0]}{P_1[T_0]} \\
 k &= \operatorname{argmax}_m \frac{R_2[m]}{P_2[m]}, \quad \beta_2 = \frac{R_2[k]}{P_2[k]}
 \end{aligned}$$

Solution:

$$\begin{aligned}
 R_1[m] &= \sum_{n=0}^{N-1} s[n] \hat{s}[n - m] \\
 P_1[m] &= \sum_{n=0}^{N-1} \hat{s}^2[n - m] \\
 R_2[m] &= \sum_{n=0}^{N-1} (s[n] - \beta_3 \hat{s}[n - T_0]) (h[n] * c_m[n]) \\
 R_2[m] &= \sum_{n=0}^{N-1} (h[n] * c_m[n])^2
 \end{aligned}$$