ECE 537 Fundamentals of Speech Processing
Problem Set 3

UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

Assigned: Monday, 9/5/2022; Due: Friday, 9/16/2022
Reading: Osamu Fujimura, "Analysis of Nasal Consonants," 1962

All parts of this problem set will use the Laplace transform analysis of the one-dimensional wave equation, as did Fujimura's paper. Since your undergraduate courses on waves and DSP might not have covered that analysis, it is reviewed here; for more extensive coverage, see any textbook on acoustics or electromagnetics. The two-sided Laplace transform of a signal $p(t)$ is given by

$$P(s) = \int_{-\infty}^{\infty} p(t) e^{-st} dt \tag{1}$$

Eq. (1) is probably a bit different from anything you've seen before (e.g., ECE 210 teaches the one-sided Laplace transform instead of this two-sided Laplace transform), and that's because the integral in Eq. (1) fails to converge for a lot of interesting signals. For example, it fails to converge for $p(t) = \cos(\omega t)$. However, it converges successfully for any finite-duration, finite-amplitude $p(t)$, so it works for real-world signals. For such signals, its inverse is just the inverse continuous-time Fourier transform:

$$p(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(j\omega) e^{j\omega t} d\omega \tag{2}$$

The Fujimura article assumes acoustic waves to be a relationship between air pressure $p(x,t)$ and volume velocity $u(x,t)$, where x is position and t is time. Volume velocity is defined to be the air particle velocity multiplied by the cross-sectional area of the wave front, so it has units of $\left[\frac{m}{s}\right] \times \left[m^2\right] = \left[\frac{m^3}{s}\right]$. Volume velocity is more useful than air particle velocity if you're studying wave propagation in a tube with varying cross-sectional area, like the vocal tract, because volume velocity has an intuitive built-in normalization for variations in the cross-sectional area: if you have a constant $0.001m^3/s$ of air entering the wide end of a tube, you'd better have a constant $0.001m^3/s$ coming out the small end, regardless of what the two areas are!

Like any other wave equation (electromagnetic, sound, water waves, waves on a violin string, or waves on a Slinky), the one-dimensional acoustic wave equation is solved by the addition of a rightward-traveling wave, $r(t - x/c)$, and a leftward-traveling wave, $l(t + x/c)$, where c is the speed of the wave. For the acoustic wave, the ratio of pressure to volume velocity is $\frac{\rho c}{A(x)}$, where ρ is the density of air and $A(x)$ is the cross-sectional area of the vocal tract at position x, thus:

$$p(x,t) = r\left(t - \frac{x}{c}\right) + l\left(t + \frac{x}{c}\right) \tag{3}$$

$$u(x,t) = \frac{A(x)}{\rho c}\left(r\left(t - \frac{x}{c}\right) - l\left(t + \frac{x}{c}\right)\right), \tag{4}$$

If you take the Laplace transform of Eqs. (3) and (4), you get

$$P(x,s) = R(s) e^{-sx/c} + L(s) e^{sx/c} \tag{5}$$

$$U(x,s) = \frac{A(x)}{\rho c}\left(R(s) e^{-sx/c} - L(s) e^{sx/c}\right) \tag{6}$$

The Fujimura article is justly famous for summarizing all of the complicated details of the mouth, nose, and pharynx by three **susceptance** functions, $B_m(s)$, $B_n(s)$, and $B_p(s)$. These are defined to be

$$B_m(s) = \frac{U_m(0,s)}{P_m(0,s)} = \frac{A_m}{\rho c}\left(\frac{R_m(s) - L_m(s)}{R_m(s) + L_m(s)}\right) \tag{7}$$

$$B_n(s) = \frac{U_n(0,s)}{P_n(0,s)} = \frac{A_n}{\rho c}\left(\frac{R_n(s) - L_n(s)}{R_n(s) + L_n(s)}\right) \tag{8}$$

$$B_p(s) = \frac{U_p(0,s)}{P_p(0,s)} = \frac{A_p}{\rho c}\left(\frac{R_p(s) - L_p(s)}{R_p(s) + L_p(s)}\right), \tag{9}$$

where $U_m(x,s)$, $P_m(x,s)$, A_m, $R_m(s)$, and $L_m(s)$ are the volume velocity, pressure, area, rightward wave, and leftward wave in the mouth cavity, respectively; likewise n for nose and p for pharynx. Eqs. (7-9) assume that the velum (the juncture between mouth, nose, and pharynx) occurs at position $x = 0$, and that the position x can be imagined to increase as one moves away from the velum in any direction. In particular, if $U_m(x,s)$, $U_n(x,s)$, and $U_p(x,s)$ all describe air velocity *away* from the velum, then resonance occurs if, for any air pressure shared by all three cavities ($P_m(0,s) = P_n(0,s) = P_p(0,s)$), the excess volume velocity coming out of two cavities is completely compensated by volume velocity going *into* the third cavity ($U_m(0,s) + U_n(0,s) + U_p(0,s) = 0$).

Many Laplace-domain functions can be most concisely written in terms of hyperbolic functions. The basic hyperbolic functions are:

$$\sinh(x) = \frac{1}{2}\left(e^x - e^{-x}\right) \tag{10}$$

$$\cosh(x) = \frac{1}{2}\left(e^x + e^{-x}\right) \tag{11}$$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} \tag{12}$$

$$\coth(x) = \frac{\cosh(x)}{\sinh(x)} \tag{13}$$

The hyperbolic functions are related to trigonometric functions by:

$$\sinh(jx) = j\sin(x) \tag{14}$$

$$\cosh(jx) = \cos(x) \tag{15}$$

1. (1 point) When viewed from the velum, the nasal cavity is a tube that's open at the other end. Since the cross-sectional area of the room is much larger than the cross-sectional area of your nose, the air pressure in the room is basically zero, regardless of how much air you blow out from your nose. In other words,

   $$P_n(d_n, s) = 0, \tag{16}$$

   where d_n is the length of the nasal tube. Notice that Eq. (16) can be interpreted as a constraint on the relationship between $L_n(s)$ and $R_n(s)$. Impose that constraint, solve for the resulting value of $B_n(s)$, and express your answer in terms of the hyperbolic functions in Eqs. (10-13).

2. (1 point) When viewed from the velum, the pharynx is a tube that's closed at the other end. Since the cross-sectional area of the glottis is much smaller than the cross-sectional area of your pharynx, there's basically no way for a standing wave in your pharynx to push air back through the glottis. In other words,

   $$U_p(d_p, s) = 0, \tag{17}$$

   where d_p is the length of the pharynx. Notice that Eq. (17) can be interpreted as a constraint on the relationship between $L_p(s)$ and $R_p(s)$. Impose that constraint, solve for the resulting value of $B_p(s)$, and express your answer in terms of the hyperbolic functions in Eqs. (10-13).

3. The Fujimura article summarizes the nose and pharynx with an “internal” susceptance, $B_i(s) = B_p(s) + B_n(s)$. The form of the internal susceptance is generally pretty complicated, so Fujimura measured it empirically. A not-quite-correct but useful approximation can be obtained, however, if we assume that $A_n = A_p$ and $d_n = d_p$.

   (a) (1 point) Under the assumption that $A_n = A_p$ and $d_n = d_p$, find $B_i(s)$. Hint: if you've done problems 1 and 2 correctly, this problem should be made trivially easy by the following trig identity:

   $$\tanh(x) + \coth(x) = 2\coth(2x) \tag{18}$$

   If Eq. (18) doesn't make this problem easy, then go back and check your answers to problems 1 and 2. If Eq. (18) makes this problem easy, then use the definitions in Eqs. (10-13) to verify the truth of Eq. (18).

   (b) (1 point) Evaluate your answer to part (a) at the frequency $s = j2\pi f$. Assuming that the total pharynx + nose has a length of about $d_n + d_p = 24.0$cm, and that the speed of sound in air at body temperature is $354\frac{m}{s}$, find the resonant frequencies for the nasal consonant /ŋ/. Compare your result to the measurements given for KS on page 1869 of the article.

4. Suppose that the mouth, for an /n/, is about 5cm long, from the velum to the tongue tip closure.

   (a) (1 point) What is the frequency of the corresponding antiformant?

   (b) (1 point) By setting $B_i(s) + B_m(s) = 0$, it is possible to calculate the formant frequencies of the /n/. Suppose that $B_i(s)$ is given by the value you calculated in problem 3. The exact values of the formants of /n/ could be calculated using the method shown in Figure 2, if we knew the values of A_m and A_n. Since we don't know the values of A_m and A_n, use the fact that each formant is somewhere between one of the singularities of $B_i(s)$ and one of the zeros of $B_i(s)$, as shown in Figure 2, to specify lower and upper bounds on the possible frequencies of the first five formants of /n/.