**SPEECH COMMUNICATION
6.541J-24.968J-HST710J
Spring 2004**

Problem Set 3   *Assigned: 02/19/04*
                                                                        Due: 02/26/04

Read Chapter 6.

Problem 1

In this problem we examine the acoustic and perceptual consequences of speaking loudly compared with speaking in a normal voice. We focus here on just one word – the word **cat** – produced in a normal and a loud voice. Spectrograms of the two utterances are shown in Figures 1.1a and 1.1b, and spectra sampled near the middle of each vowel are given in Figures 1.2a and 1.2b. These are DFT spectra obtained with a 25.6 ms window, with a time window that is displayed superimposed on the waveform below each spectrum. The reference for amplitude scales (in dB) for the spectra are arbitrary, except that the reference is the same for the two spectra, i.e., the spectra (and the waveforms below them) correctly represent the relative amplitudes. In making the spectrograms in Figure 1.1, however, the overall amplitudes are normalized to have the same maximum amplitude.

a) Estimate the difference, in dB, between the overall sound pressure levels for the two vowels, as represented by the spectra in Figure 1.2?

b) Apart from differences in overall amplitude, there are several differences in the way the two utterances are produced. These differences may be related to the shaping of the vocal tract for the vowel and consonants (e.g., tongue position, jaw position, etc.,), the adjustment of the laryngeal configuration and vocal-fold stiffness, and the control of the respiratory system. Using the acoustic data from the spectrograms, the spectra, and the waveforms in Figures 1.1 and 1.2 as evidence, discuss the differences in the way these two utterances are produced. Be as quantitative as you can in making inferences about physiological parameters from the acoustics.
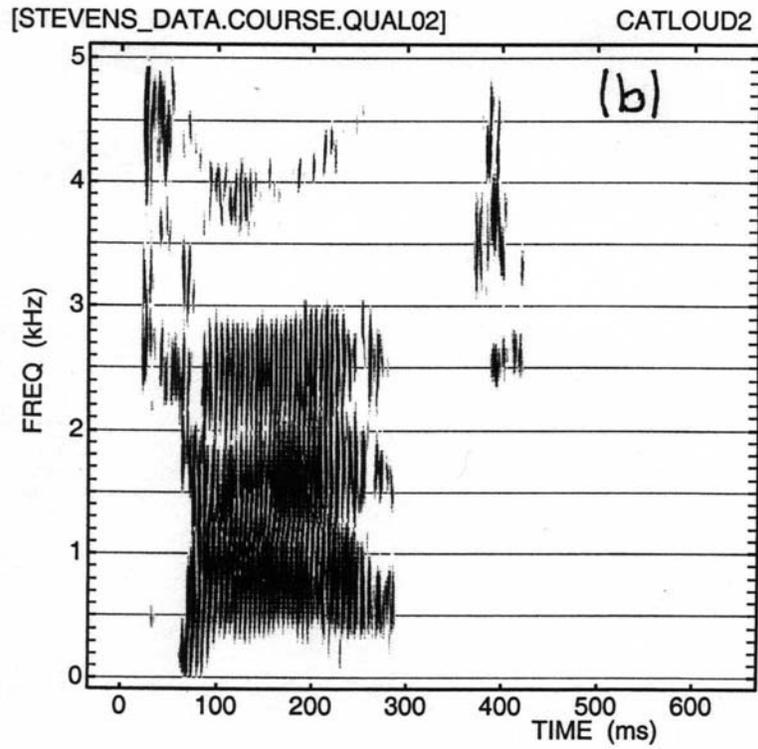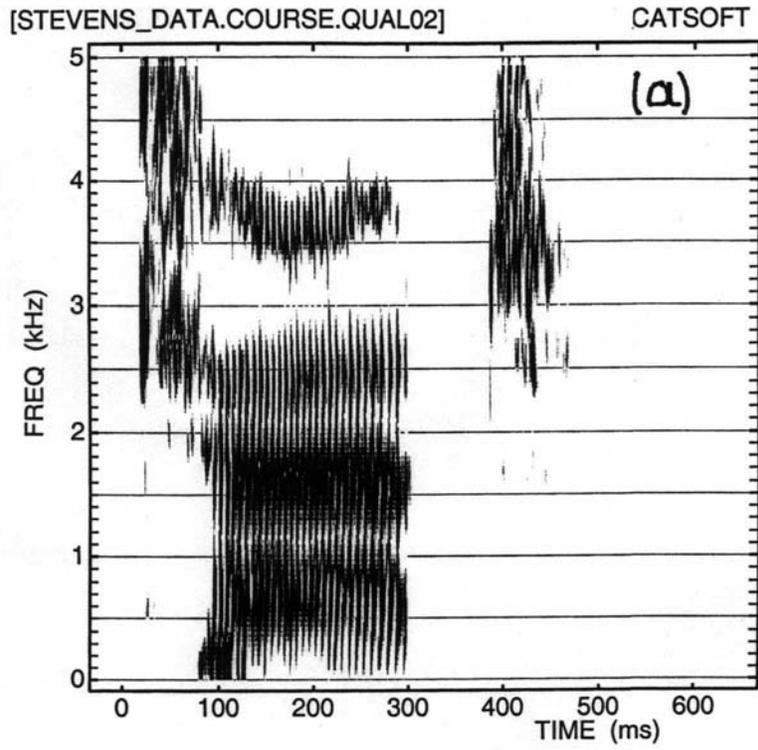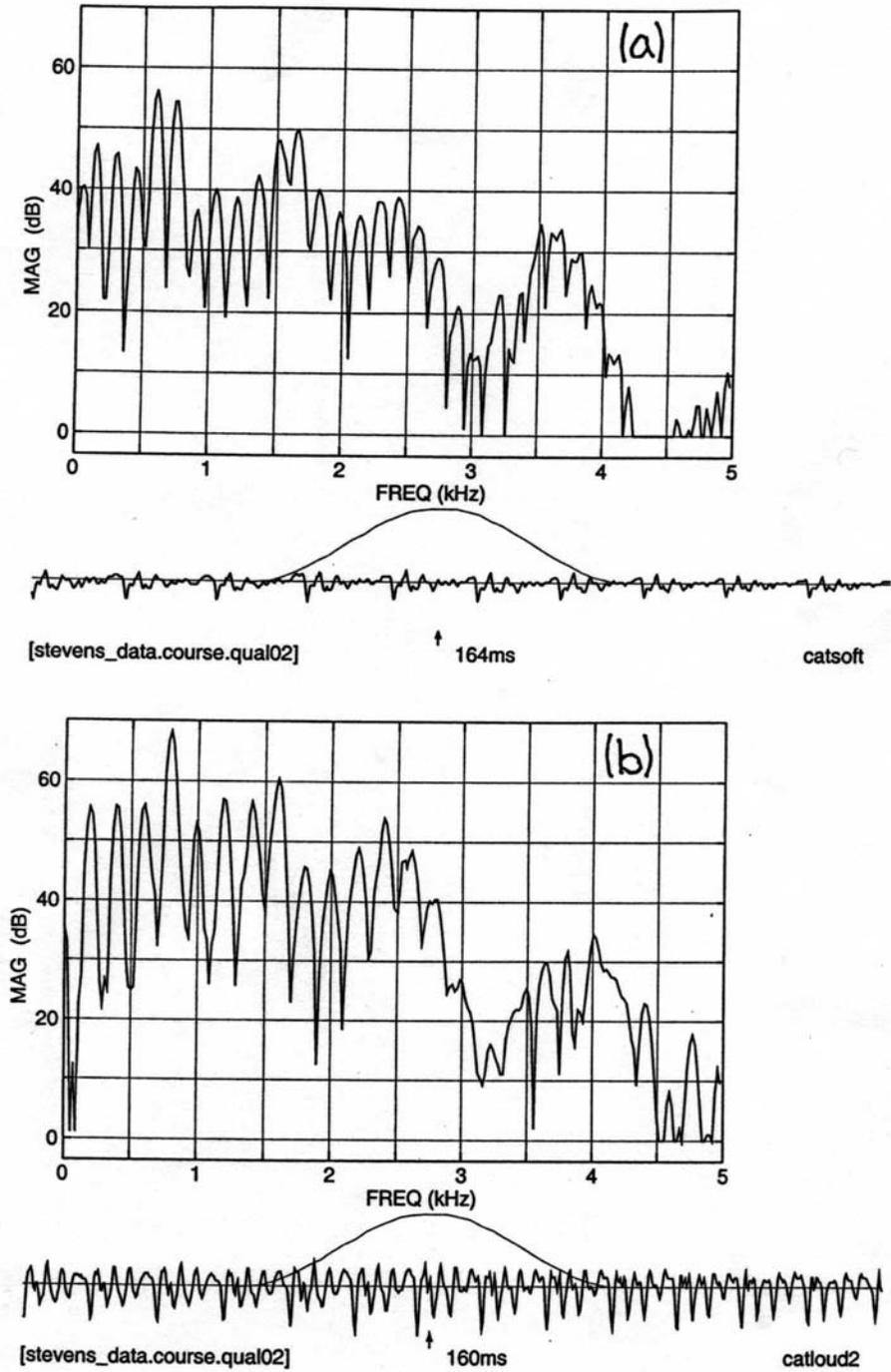
[STEVENS_DATA.COURSE.QUAL02]   CATSOFT

(a)

[STEVENS_DATA.COURSE.QUAL02]   CATLOUD2

(b)

Figure 1.1

[stevens_data.course.qual02]   164ms   catsoft

[stevens_data.course.qual02]   160ms   catloud2

Figure 1.2

Problem 2

An approximate model of the vocal tract configuration during the time interval of the lip closure for a labial stop consonant is shown in Fig. 2.1. The tube behind the closure is uniform, with area A and length ℓ.

Assume the following values for the parameters:

$$A = 4 \text{ cm}^2 \qquad \ell_c = 0.8 \text{ cm}$$
$$\ell = 15 \text{ cm} \qquad \rho = 0.00114 \text{ gm/cm}^3$$
$$c = 35400 \text{ cm/s}$$

When the consonant is released, the cross-sectional area at the lips increases rapidly as indicated in Fig. 2.2. We shall assume that the time course of the change in cross-sectional area $A_c(t)$ of the lip opening is as given in Fig. 2.3. That is, the area of the lip opening increases linearly with time at a rate of 40 cm^2/sec up to an area of 4 cm^2, which is the cross-sectional area of the tube that models the vocal tract behind the lips.

(a) Calculate the frequencies of the first three formants F1, F2, and F3 (1) during the time the lips are closed, and (2) after the lip opening has reached an area of 4 cm^2.

The frequency F1 of the first formant should be corrected for the effect of the walls. Thus if F1' is the first formant frequency on the assumption of hard walls for the tube, then the corrected value of F1 is given by

$$F1 = \sqrt{(F1')^2 + 180^2} \qquad\qquad (1)$$

as we have seen in Chapter 3 of the book (page 159).

(b) Calculate and plot the frequencies of the first three formants as a function of time during the 150-ms interval following the release.

Probably the simplest way of calculating the formant frequencies when the area of the lip opening is to solve the following equation for the frequency $f = \dfrac{\omega}{2\pi}$.

$$-j\frac{\rho c}{A}\cot\frac{\omega \ell}{c} + j\omega\frac{\rho \ell_c}{A_c} = 0, \qquad\qquad (2)$$

i.e., the sum of the impedances looking both ways from a point immediately behind the lip opening is zero. You might want to use MATLAB to solve the equation. (The equation can also be solved graphically.)

The procedure is to calculate the difference

$$\frac{2\pi f \rho \ell_c}{A_c} - \frac{\rho c}{A}\cot\frac{2\pi f\ell}{c} \tag{3}$$

and to find values of f that minimize the absolute value of this difference in particular frequency ranges.

Your calculations should show that F1, F2, and F3 do not rise synchronously, but that the formants increase at different rates.
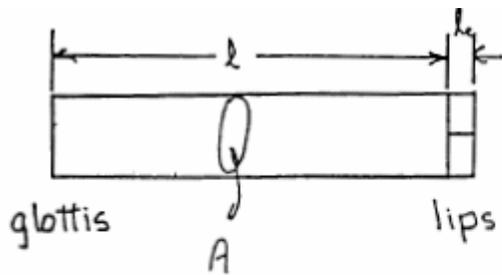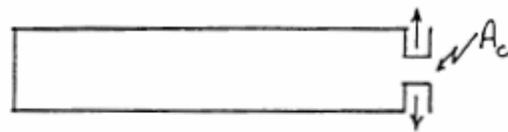

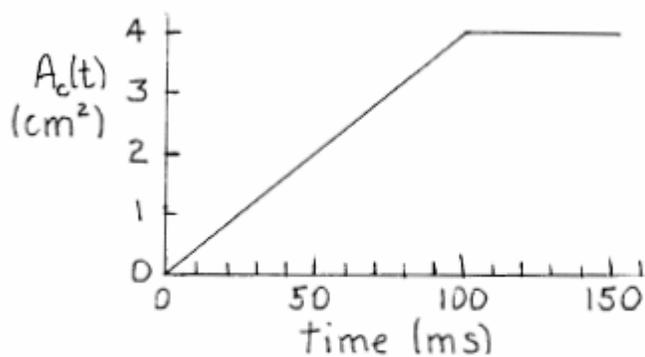Fig. 2.1   Fig. 2.2

Fig. 2.3

Problem 3

The spectrum U_s of the volume-velocity waveform at the glottis for a particular speaker is given in Fig. 3.1. This volume velocity forms the glottal excitation for a vowel that is produced with a uniform vocal tract such that the formant frequencies are 500, 1500, 2500, ... Hz. The cross-sectional area of the vocal tract is 4 cm^2. The bandwidths of the formants are, respectively, 100, 150, 200, ... Hz. Sound is radiated from the mouth opening, and the sound pressure is measured with a microphone at a distance of 20 cm in front of the mouth.

(a) Calculate the amplitude of the acoustic volume velocity at the mouth output for the harmonic at 1000 Hz.

The density of air is $\rho = 0.00114$ gm/cm^3

The velocity of sound is $c = 35400$ cm/s

(b) Calculate the sound pressure (at the microphone) of the harmonic at 1000 Hz. Your answer should be in dynes/cm^2.

(c) Calculate the sound pressure (at the microphone) of the harmonic that is at the frequency of the 2nd formant.

Note that when the formants are equally spaced, as they are in this problem, the peak-to-valley ratio for the transfer function (U_o/U_s, where U_o = output volume velocity and U_s = source volume velocity) is given by the formula $\frac{2S}{\pi B}$, where S = spacing of formants, B = bandwidth. (You might wish to derive this formula, but that is not required for this problem.)

(d) The first formant frequency is now shifted down to 350 Hz, but the other formant frequencies remain at their original values. Repeat part (b) for this new frequency of F1. (An approximate solution is adequate, noting that the spectrum at higher frequencies rides on the "skirt" of the component of the transfer function contributed by F1.)

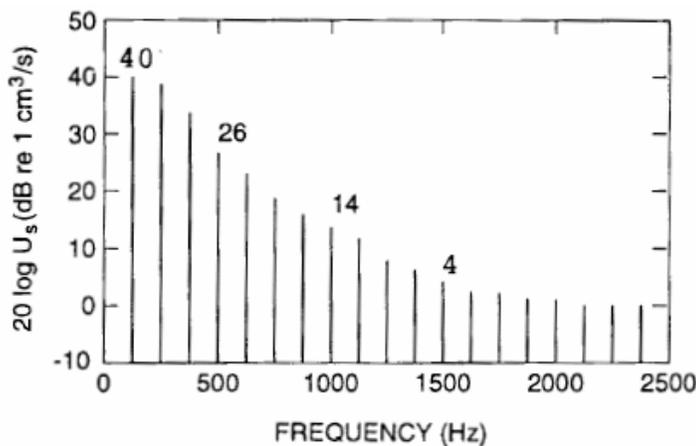(e) Repeat part (c) for the new frequency of F1 in part (d).

Figure 3.1