

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

Problem Set 4

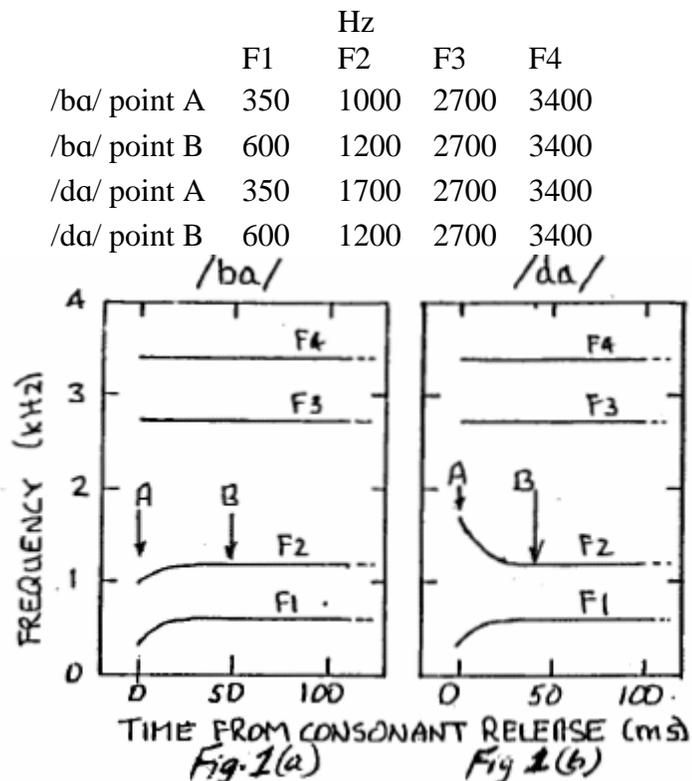
Assigned: 03/02/04

Due: 03/09/04

Read Chapter 7; Chapter 8, Sections 8.1 and 8.2; and Chapter 9.

**Problem 1**

At the release of a labial stop consonant before the vowel /a/ (such as /ba/), assume that the formants change in the manner shown in Fig. 1a. For the syllable /da/, the first formant follows a similar path but the movement of the second formant is different, as in Fig. 1b. We assume for purposes of this problem that the frequencies of the higher formants (F3, F4, etc.) remain fixed. The frequencies of the first four formants at points A and B for each syllable are tabulated below.



Determine the **change** (in dB) in the **amplitude** of the fourth-formant peak in the spectrum as we move from point A near the beginning of the transition to point B in the steady-state vowel for each

of these idealized syllables. What is the difference in the amplitude of the F4 peak at point A in the /da/ syllable relative to point A in the /ba/ syllable? Assume that the spectrum and amplitude of the glottal source remains the same throughout both syllables. (The factors that influence the relative amplitudes of formant prominences are discussed in Section 3.2 of the text.)

To get an approximate answer (all that is necessary here), recall that as the frequency of a low-frequency formant like F1 or F2 changes, higher-frequency components of the spectrum change in amplitude in proportion to the square of the ratio of the new formant frequency to the old formant frequency. Thus, as the first-formant frequency changes from F11 to F12, the change in spectrum amplitude at frequencies much higher than F1 is approximately  $40 \log \frac{F12}{F11}$  dB.

### Problem 2

For fricative consonants there is a turbulence noise source just downstream from a constriction in the vocal tract. This noise source can be represented as a sound-pressure source in series with the vocal tract. To simplify the problem we will assume that the vocal-tract downstream from the constriction has a uniform cross-sectional area and the sound-pressure source  $p_s$  is located distance  $\ell_b$  from the constriction, as illustrated in Fig. 2.1. The noise source is assumed to be a consequence of a jet of air from the narrow constriction impinging on an obstacle in the vicinity of the point marked  $p_s$ .

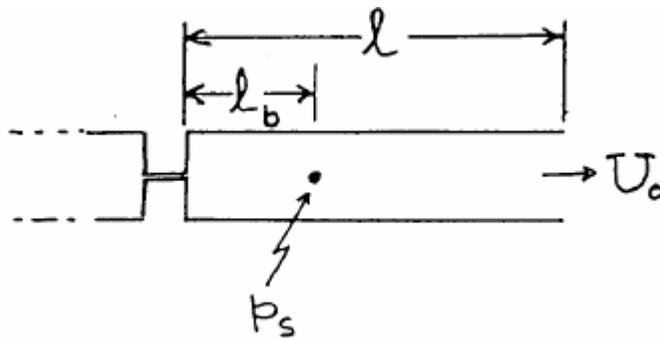


Fig. 2.1

We assume the constriction to have infinite impedance. The length of the tube anterior to the constriction is  $\ell$ , and its cross-sectional area is  $A$ . An equivalent circuit for this arrangement is as shown in Fig. 2.2.

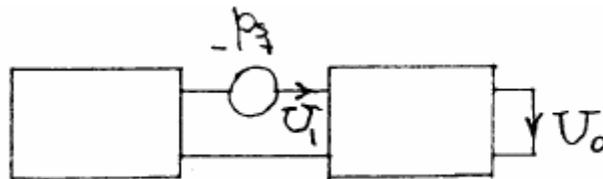


Fig. 2.2

(a) Assume the tube is lossless. Show that the transfer function  $\frac{U_o}{P_s}$  is

$$\frac{U_o}{P_s} = j \frac{A \sin k\ell_b}{\rho c \cos k\ell} \quad (1)$$

(Hint: It may be useful to write

$$\frac{U_o}{P_s} = \frac{U_o}{U_1} \cdot \frac{U_1}{P_s} \quad (2)$$

and write expressions for each of these terms). Interpret the poles and zeros in this expression.

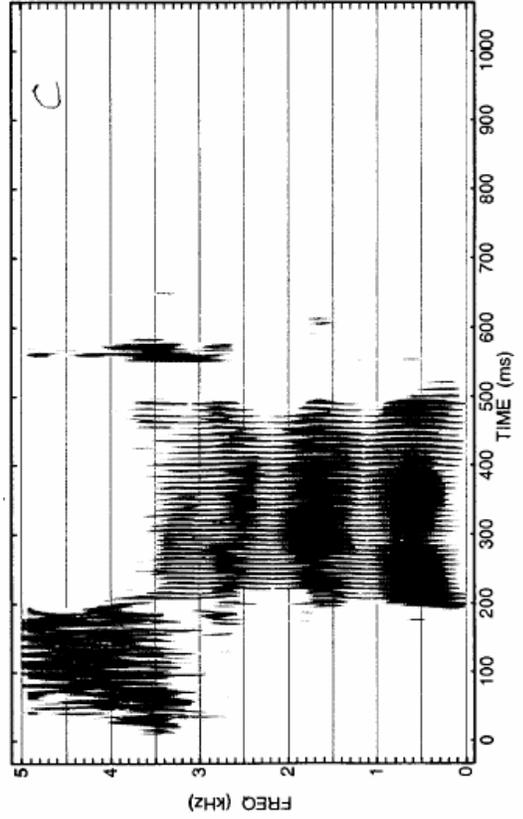
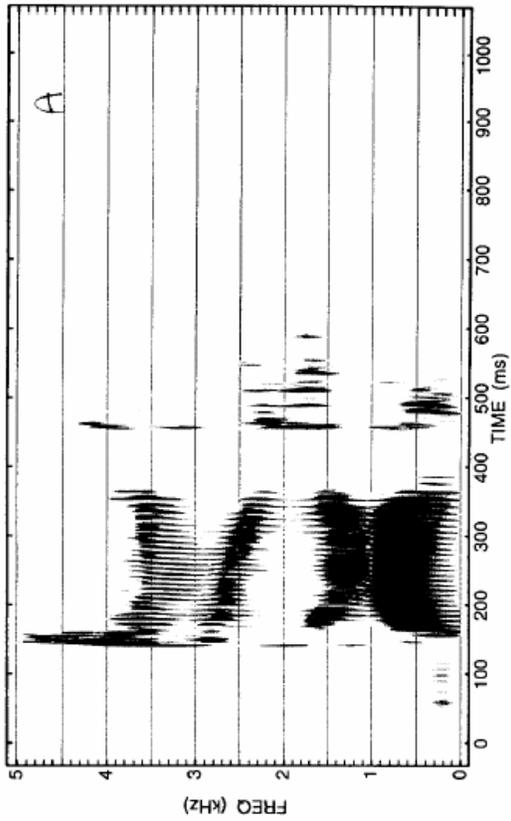
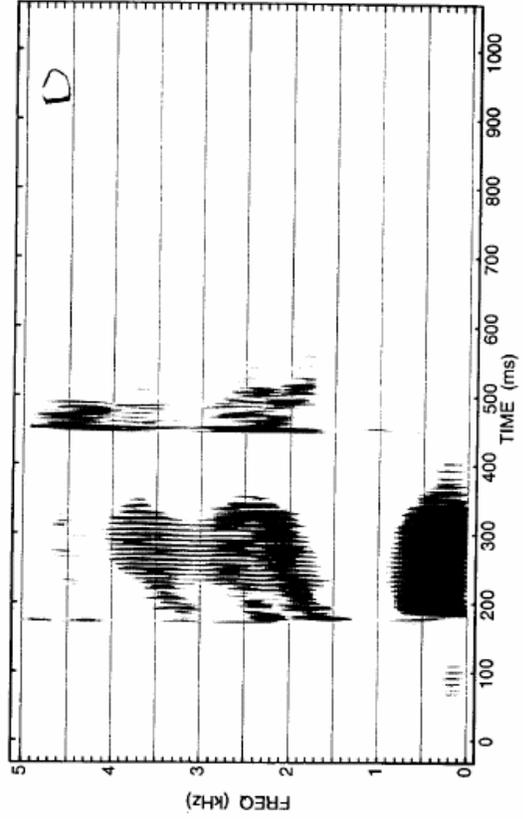
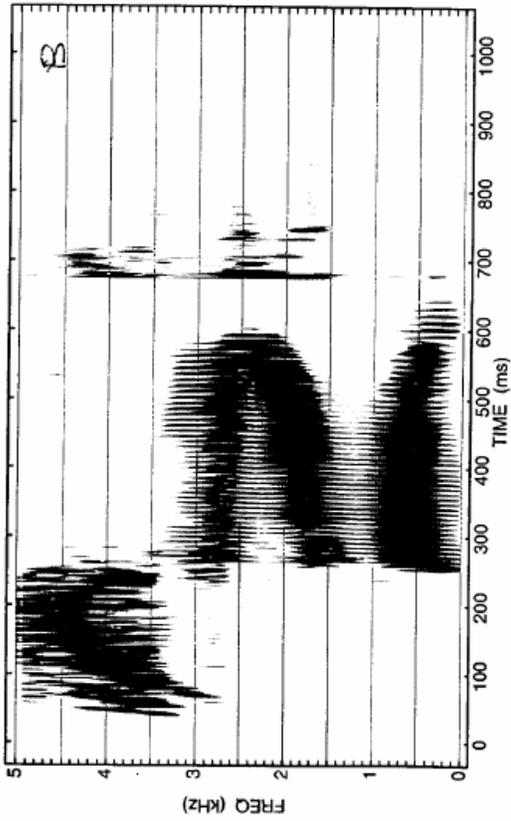
(b) Assume now that the length  $\ell$  of the front cavity is 3 cm, and that the distance  $\ell_b$  from the constriction to the source is 2 cm. What are the frequencies of the lowest-frequency pole and the lowest-frequency zero of the transfer function  $\frac{U_o}{P_s}$ ?

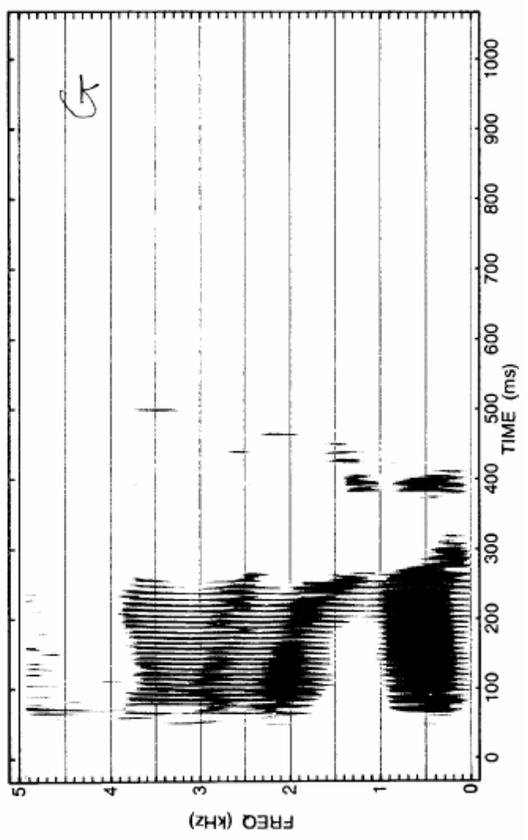
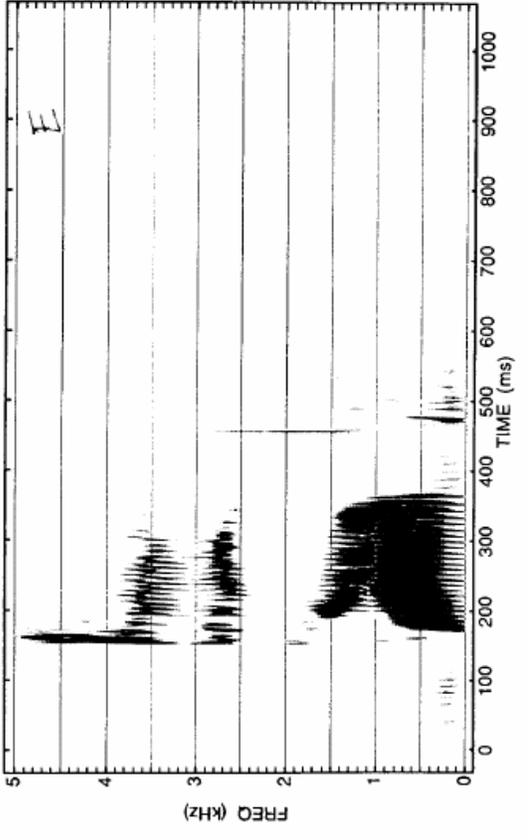
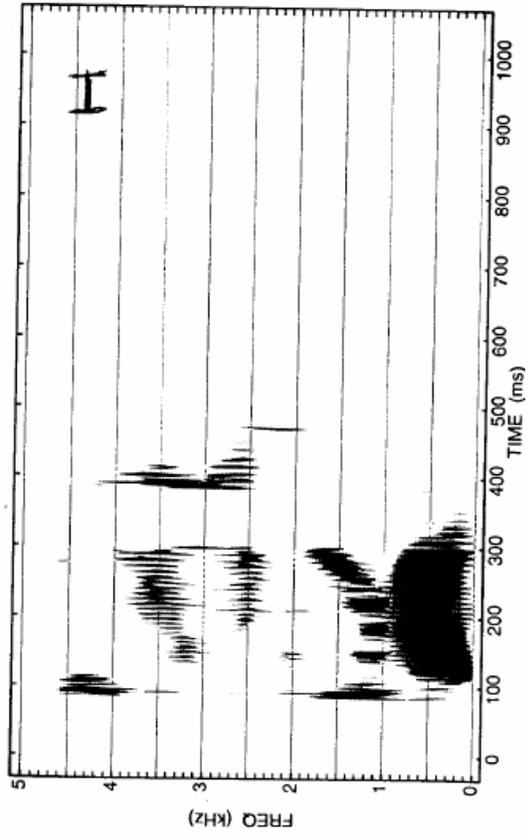
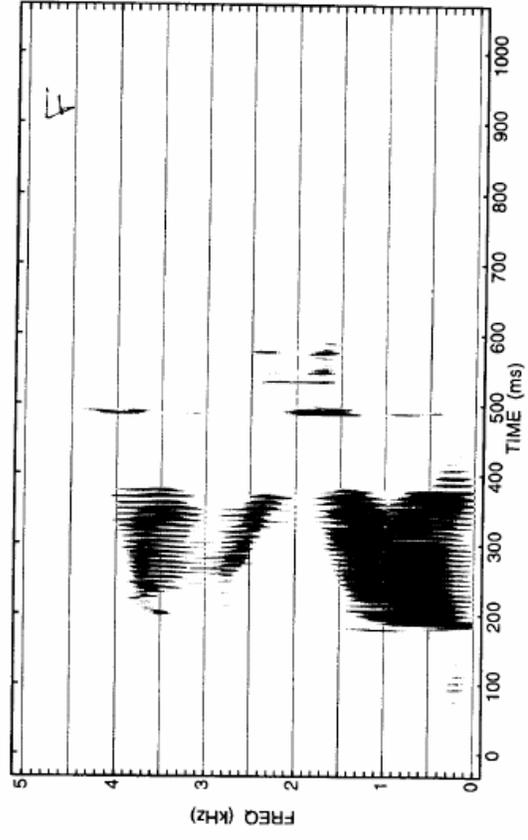
### Problem 3

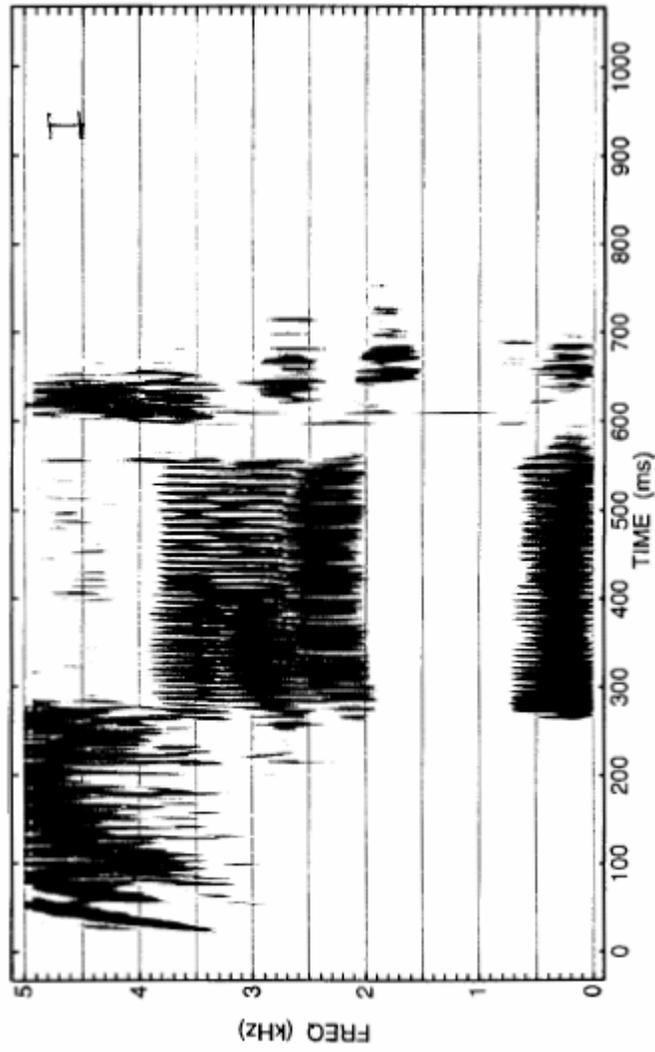
Attached are spectrograms of nine words. The words are taken from the following, longer, list of words.

bad	dug
beg	ebb
bed	egg
big	good
bode	sad
bug	
daze	sag
dog	said
dub	seed

Identify the words on the spectrograms.







**Problem 4**

Attached are 8 spectrograms of short utterances. Some are spoken by a female speaker and some by a male speaker. The sentences from which these utterances are drawn are:

- A. I saw Timmy.
- B. I see Timmy.
- C. I saw Tommy.
- D. I see Tommy.
- E. I saw Tammy.
- F. I see Tammy.

Some utterances may be repeated (and maybe some are not represented).

- (a) For each spectrogram, identify which utterance was produced.
- (b) Identify which utterances were spoken by the female speaker, and which by the male speaker.
- (c) Measure the average fundamental frequency over the words **see** and **saw** for the female and for the male speaker. Note that each glottal pulse is represented as a vertical line on the spectrogram.
- (d) Measure F2 and F3 in the middle of the vowels **see** and **saw** for each repetition by each speaker. Tabulate your results giving average F2 for female **see**, for female **saw**, for male **see**, and for male **saw**, and also average F3 for each. Calculate the percent difference between male and female values for each of the four pairs. Note that if there was uniform scaling of vocal-tract size between female and male speakers, and if the vowels were produced with the same vocal-tract shape for males and females, these percentages should all be the same. Do you have any hypotheses to explain the results? For example, what if male speakers have a relatively longer pharynx than female speakers? (See Table 1.1 of Chapter 1.)

