New Mexico State University
Klipsch School of Electrical Engineering

EE590 - Digital Speech Processing
Fall 2003 - Exam #1

Name: ____________________

Prob. 1	/ 20 points
Prob. 2	/ 20 points
Prob. 3	/ 20 points
Prob. 4	/ 20 points
Prob. 5	/ 20 points
Total	/ 100 points

Prob. 1

The waveform plots and spectrograms in Figs. 1.1 and 1.2 (next two pages), were made from recordings ($f_s = 11025\text{Hz}$, 16 bit resolution) of a single, spoken phoneme. Using these figures, determine the type of phoneme which was spoken:

- Vowel
- Nasal
- Voiced fricative
- Unvoiced fricative
- Voiced plosive
- Unvoiced plosive
- Dipthong

Note some phoneme types may occur more than once, some phoneme types may not occur. Providing comments regarding your decision may lead to partial credit.

(a) The phoneme in Fig. 1.1(a) and (b) is a ______________________________ .

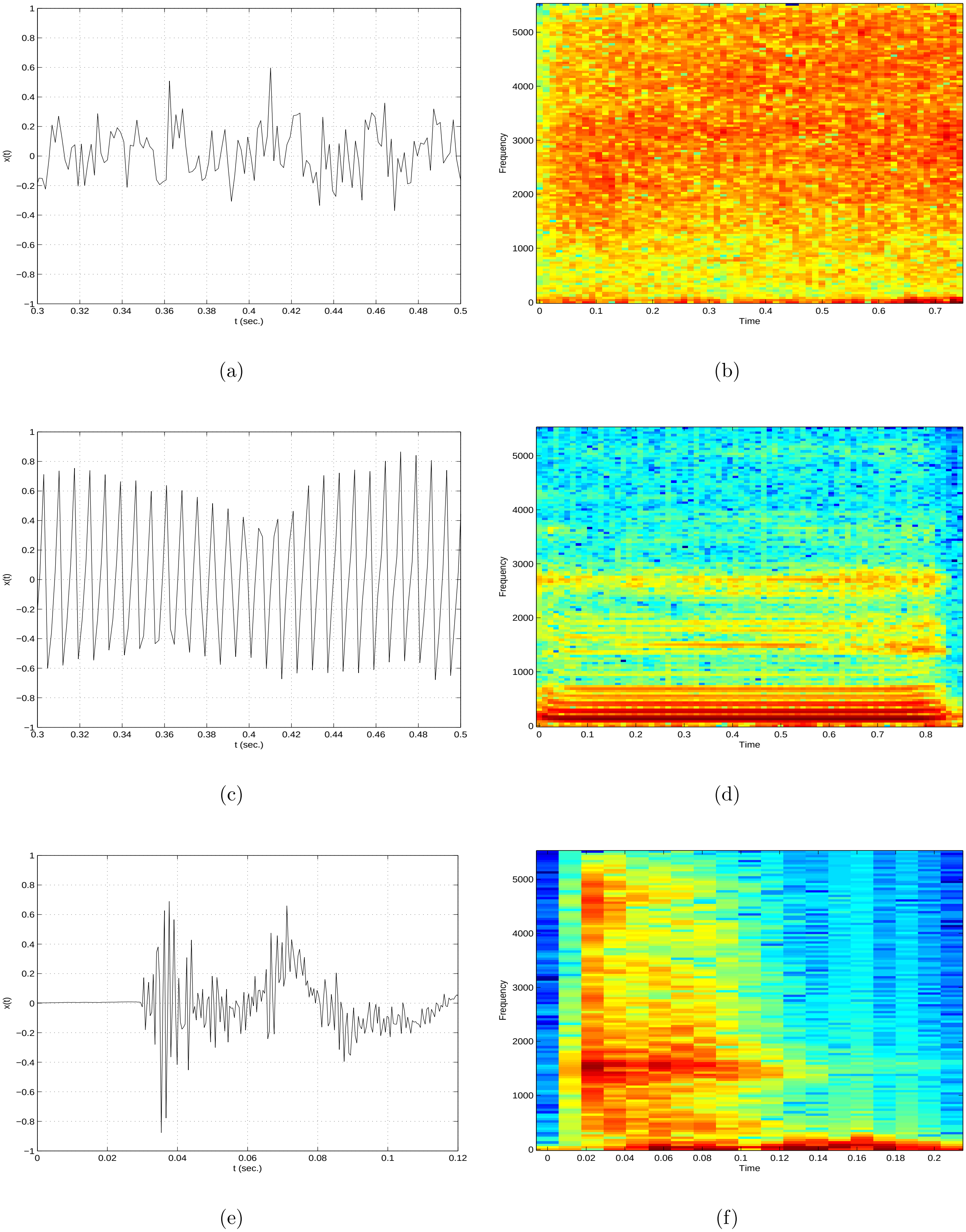
(b) The phoneme in Fig. 1.1(c) and (d) is a ______________________________ .

(c) The phoneme in Fig. 1.1(e) and (f) is a ______________________________ .

(d) The phoneme in Fig. 1.2(a) and (b) is a ______________________________ .

(e) The phoneme in Fig. 1.2(c) and (d) is a ______________________________ .

(f) The phoneme in Fig. 1.2(e) and (f) is a ______________________________ .

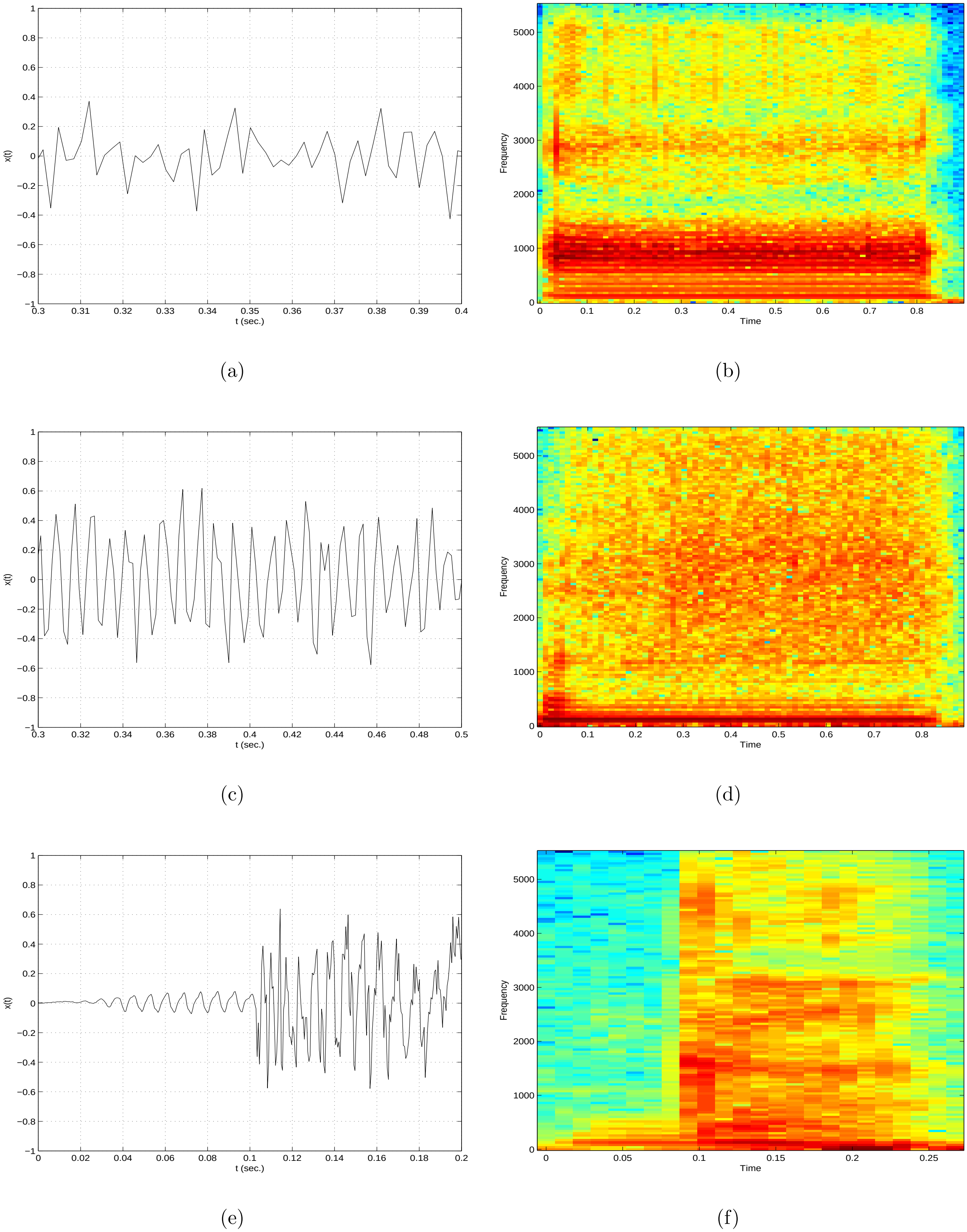


(a) (b)

(c) (d)

(e) (f)

Figure 1.1: Plots of phoneme waveforms and periodograms for Prob. 1.



(a) (b)

(c) (d)

(e) (f)

Figure 1.2: Plots of phoneme waveforms and periodograms for Prob. 1.

Prob. 2

Fig. 2.1 (Text Figure 4.20, p. 150) is an overview of the complete discrete-time speech production model with components labeled (a)–(j). Fig. 2.2 (next page) shows spectra (periodograms) from two signals emitted by the model. The signals are sampled at a rate of 8000 samples/second and are windowed with different window lengths prior to Fourier transform.

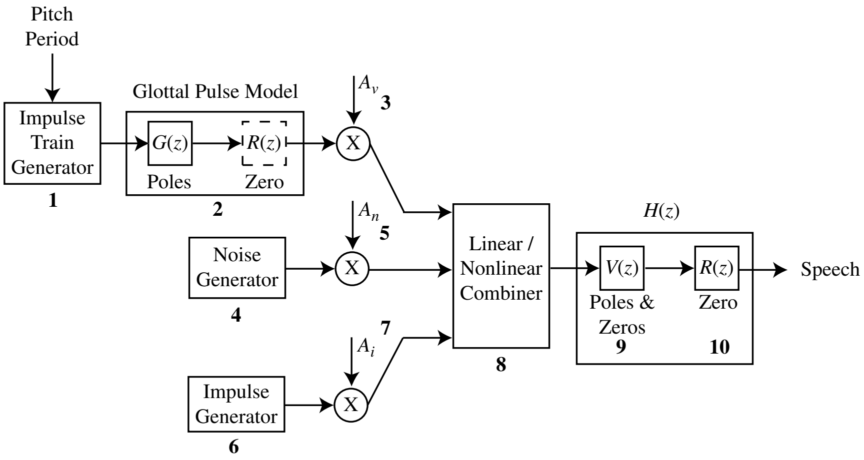


Figure 2.1: Discrete-time system model for speech production for Prob. 2.

(a) Which part of the model in Fig. 2.1 determines the formant frequencies?
Circle your answer: **1 2 3 4 5 6 7 8 9 10**

(b) Which part of the model in Fig. 2.1 determines the spacing of the prominent local peaks in Spectrum A? Circle your answer: **1 2 3 4 5 6 7 8 9 10**

(c) Which spectrum(s) in Fig. 2.2 correspond to voiced speech? Circle your answer(s): A B C D E

(d) Which spectrum(s) in Fig. 2.2 correspond to unvoiced speech? Circle your answer(s): A B C D E

(e) Which spectrum(s) in Fig. 2.2 were computed with the longest window?
Circle your answer(s): A B C D E

(f) Which spectrum(s) in Fig. 2.2 were computed with the shortest window?
Circle your answer(s): A B C D E

(g) Estimate the pitch or fundamental frequency (in Hz) in Spectrum A in Fig. 2.2.

(h) Estimate the first three formant frequencies (in Hz) in Spectrum B in Fig. 2.2.

$f_1 =$, $f_2 =$, and $f_3 =$

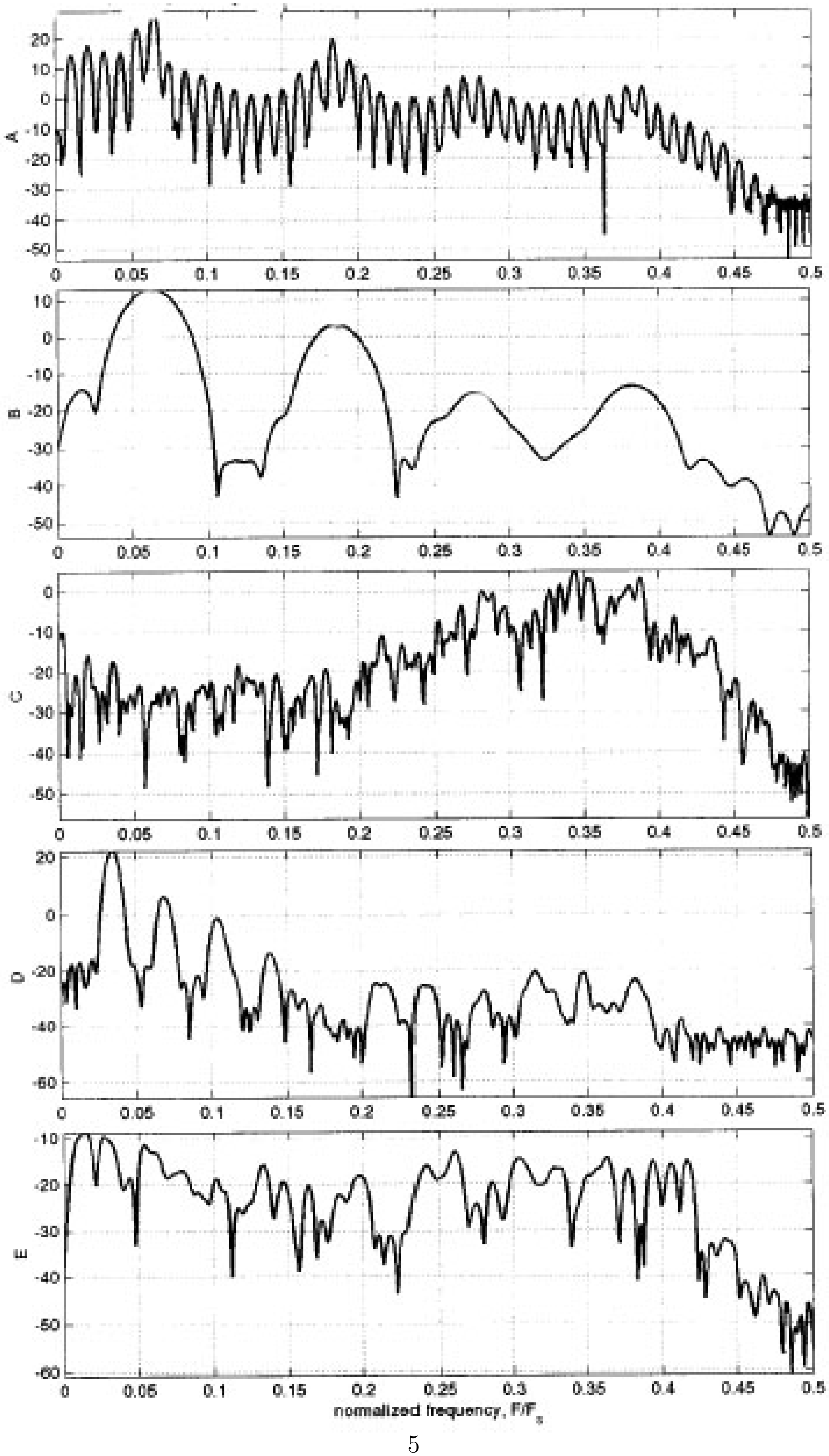


Figure 2.2: Periodograms for Prob. 2.

Prob. 3

[Text Prob. 4.9(a)]. Consider the signal flow graph of the lossless, two-tube model, depicted in Fig. 3.1 [Text Figure 4.18(a), p. 146]. Derive the transfer function for the two-tube model,

$$V_a(s) = \frac{be^{-s2\tau}}{1 + a_1 e^{-s2\tau} + a_2 e^{-s4\tau}} \tag{3.1}$$

where $b = (1 + r_g)(1 + r_L)(1 + r_1)/2$, $a_1 = r_1 r_g + r_1 r_L$, and $a_2 = r_L r_g$ corresponding to Fig. 3.1.

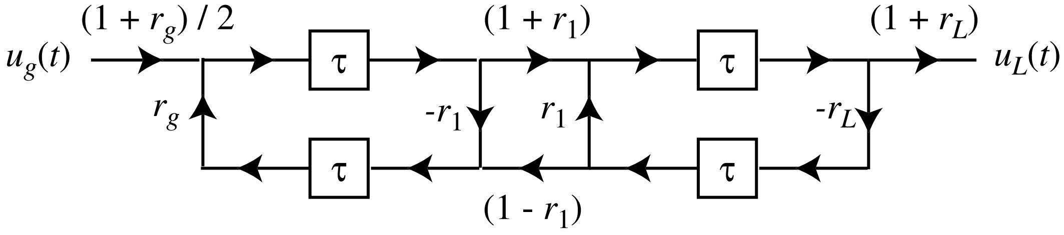


Figure 3.1: Signal flow graph for lossless, two-tube model in Prob. 3.

Prob. 4

Consider the signal flow graph of the lossless, two-tube model, depicted in Fig. 3.1 [Text Figure 4.18(a), p. 146]. The reflection coefficients are given as $r_g = r_L = 1$. The reflection coefficient r_1 is expressed in terms of the areas of the two tubes as

$$r_1 = \frac{A_2 - A_1}{A_2 + A_1}. \tag{4.1}$$

Assume the velocity of sound is $c = 344\text{m/sec}$ and that the dimensions of the two tubes are identical; $l_1 = l_2 = 8.5\text{cm}$ and $A_1 = A_2 = 5\text{cm}^2$. Finally, in Prob. 3, we showed that the system function relating $U_L(s)$ to $U_G(s)$ is

$$V_a(s) = \frac{be^{-s2\tau}}{1 + a_1 e^{-s2\tau} + a_2 e^{-s4\tau}} \tag{4.2}$$

where $b = (1 + r_g)(1 + r_L)(1 + r_1)/2$, $a_1 = r_1 r_g + r_1 r_L$, and $a_2 = r_L r_g$.

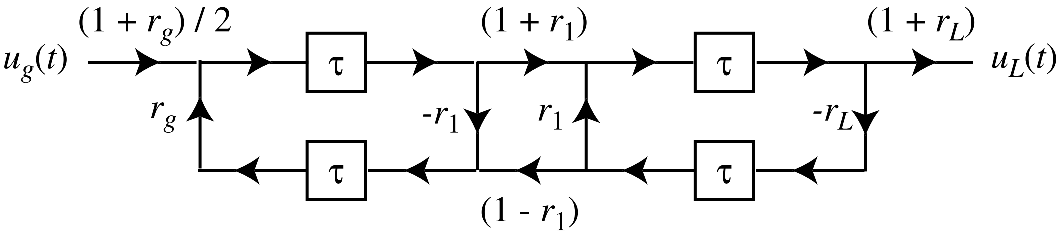


Figure 4.1: Signal flow graph for lossless, two-tube model in Prob. 4.

(a) Why is the system lossless?

(b) Determine the time delay τ and the reflection coefficient r_1.

(c) Can the two connected tubes described above, be considered a single uniform tube? Why or why not?

(d) Draw the simplest signal flow graph of the corresponding discrete-time system that is obtained by assuming a bandlimited input which is sampled at a rate $f_s = 1/T = 1/(2\tau)$. By simplest we mean the one with the fewest elements (delays, multipliers, summing nodes, etc.).

(e) Determine the system function $V(z)$ for the system in part (d). Also determine its poles and plot them in the z-plane.

Prob. 5

Let $e[n] = s[n] - \tilde{s}[n]$ denote the prediction error where $s[n]$ is the speech signal and $\tilde{s}[n]$ is the prediction. The prediction is formed with a pth order autoregressive model.

$$\tilde{s}[n] = \sum_{k=1}^{p} \alpha_k s[n-k]. \tag{5.1}$$

Let $A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$ denote the prediction error filter. To keep things simple, let the mean-squared prediction error be given by

$$E = \sum_{n} e^2[n]. \tag{5.2}$$

In this problem we will show that the zeros of $A(z)$ [poles of $H(z) = A/A(z)$] are inside the unit circle thus proving the minimum-phase nature of the model computed using the autocorrelation method.

(a) [3 points] Use Parseval's theorem to express the mean-squared prediction error, E in terms of $A(\omega)$ and $S(\omega)$.

(b) [17 points] Use your result in (a) together with the fact that $E > 0$ to prove that $A(z)$ has all its zeros inside the unit circle. Hint: assume that z_i is a zero of $A(z)$ and is outside the unit circle. Then $A(z) = (1 - z_i z^{-1})A'(z)$ where $A'(z)$ is the minimum phase portion. Then argue that $A(z)$ cannot be optimal, i.e. it does not minimize the mean-squared error. Thus z_i must be inside the unit circle.