New Mexico State University
Klipsch School of Electrical Engineering

EE589 - Digital Speech Processing
Fall 2005 - Exam #1

Name: ____________________

Prob. 1	/ 25 points
Prob. 2	/ 20 points
Prob. 3	/ 35 points
Prob. 4	/ 20 points
Total	/ 100 points

Prob. 1

The waveforms and spectrograms in Figs. 1.1 and 1.2 (next two pages), were made from recordings ($f_s = 16$ kHz, 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 ______________________________ .
Comment:

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

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

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

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

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

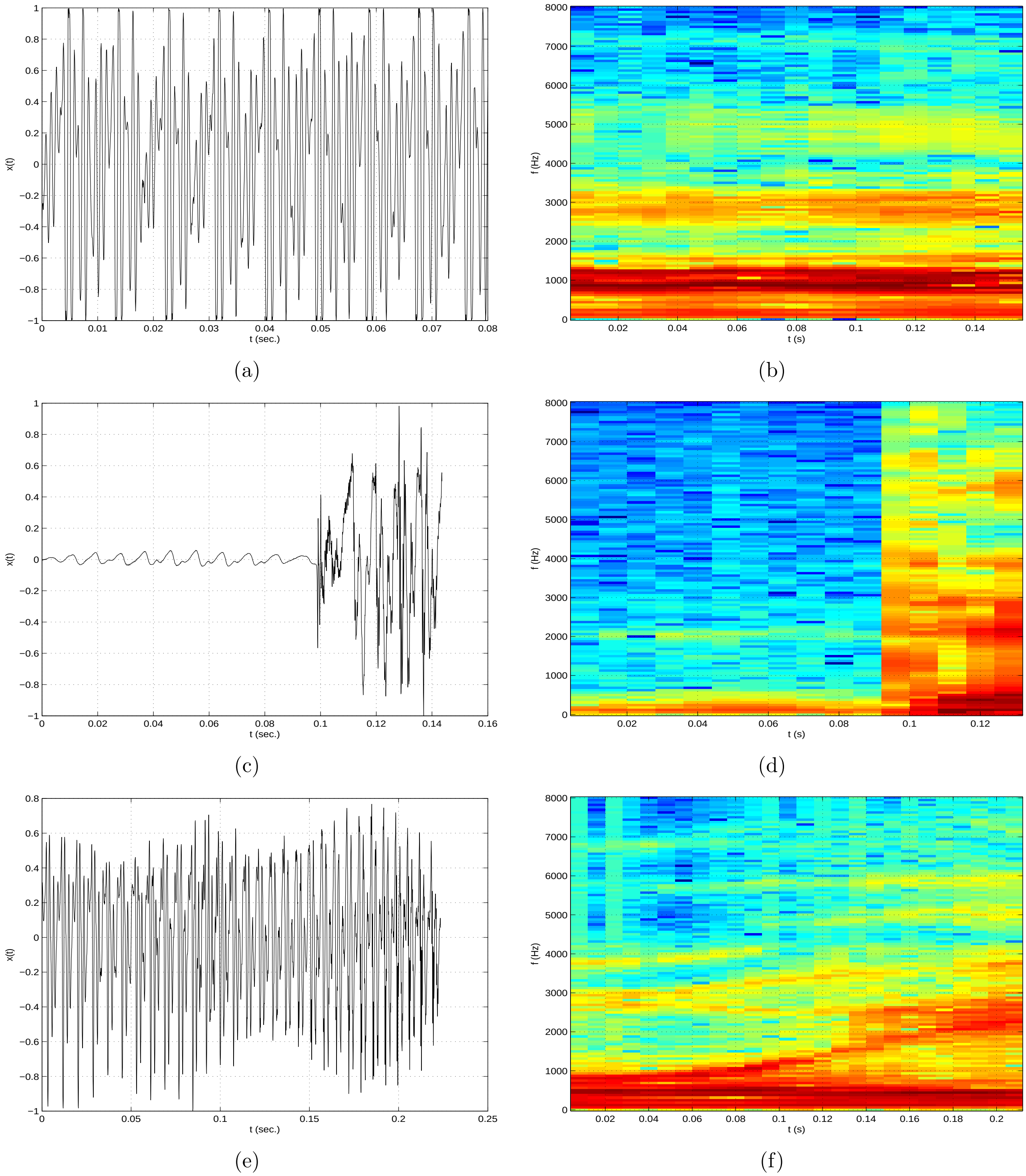

Figure 1.1: Plots of phoneme waveforms and periodograms for Prob. 1.

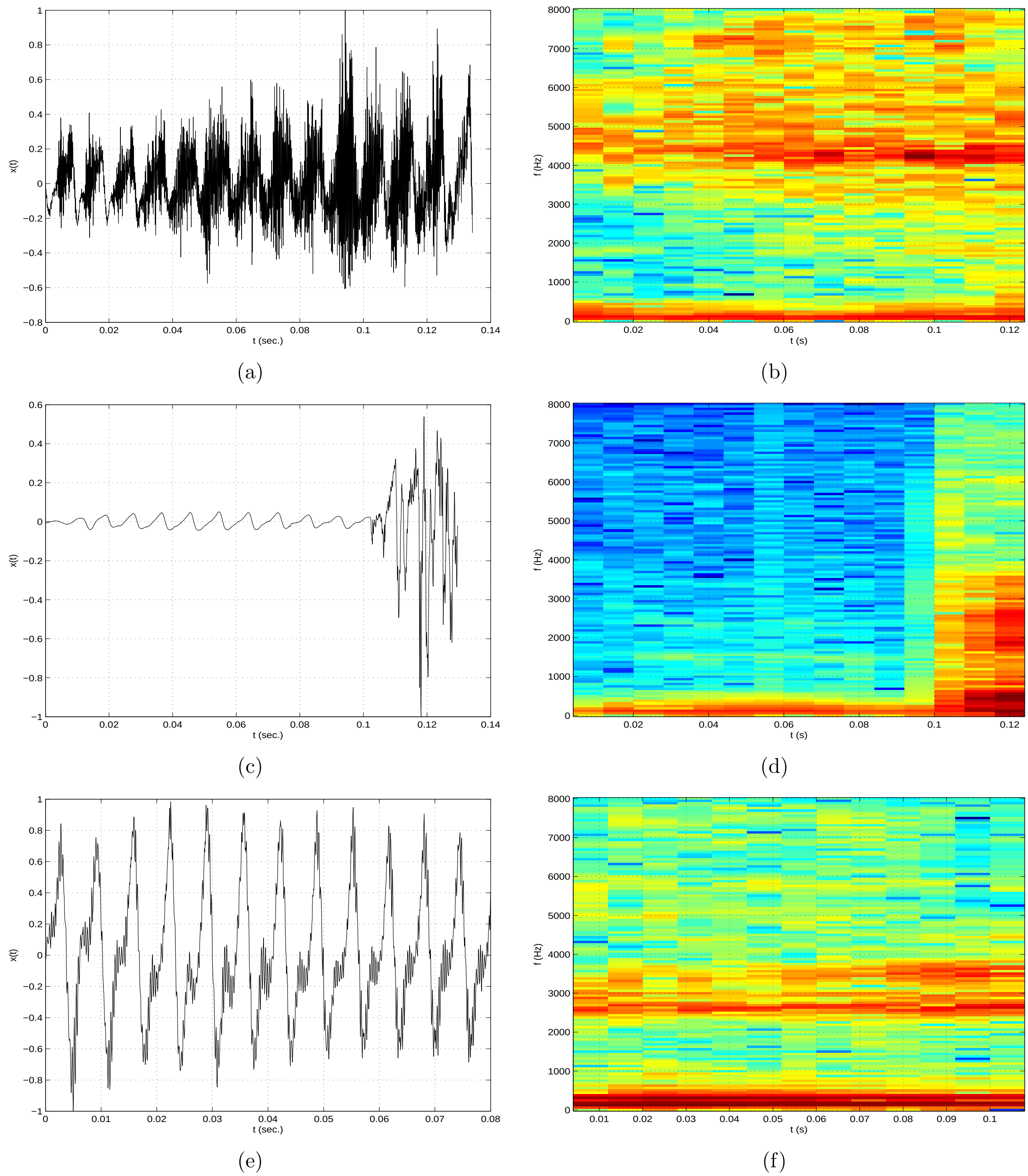

Figure 1.2: Plots of phoneme waveforms and periodograms for Prob. 1.

Prob. 2

[Text Prob. 4.14] Figure 4.33 represents the magnitude of the discrete-time Fourier transform of a steady-state vowel segment which has been extracted using a rectangular window. The envelope of the spectral magnitude, $|H(\omega)|$, is sketched with a dashed line. Note that three formants (resonances) are shown, and that only the main lobe of the window Fourier transform is depicted.

(a) Suppose the sampling rate is 6000 samples/s and was set to meet the Nyquist rate. What is the pitch period milliseconds? How long is the rectangular window in milliseconds?

(b) If $F_1 = 750$ Hz and the vocal tract is considered to be a single acoustic tube, what is the length of the vocal tract? Assume zero pressure drop at the lips, an ideal volume velocity source, and speed of sound $c = 350$ m/s.

(c) If the length of the vocal tract were shortened, how would this affect the spacing of the window main lobes that make up the discrete-time magnitude spectrum of the signal? Explain your answer.

(d) Suppose that $H(\omega)$ represents the frequency response between the lip pressure and glottal volume velocity. How would the spectral magnitude change if there were no radiation load at the lips?

Prob. 2 (cont.)

Prob. 3

[Text Prob. 3.2(a)] Suppose that the pitch period of the speaker is steady except for a small deviation ϵ that alters in sign every pitch period (Figure 3.31a). Please use Fig. **??** in place of Figure 3.31a since it is not a correct illustration.

(a) Show that the glottal pulse train (assume no shaping) can be expressed as

$$p[n] = \sum_{k=-\infty}^{\infty} \delta[n-(2k)P] + \sum_{k=-\infty}^{\infty} \delta[n+\epsilon-(2k+1)P]. \tag{3.1}$$

(b) Derive the following expression:

$$|P(\omega)|^2 = 2\left(1+\cos\left[(\epsilon - P)\,\omega\right]\right)\left[\sum_{k=-\infty}^{\infty} \frac{2\pi}{2P}\delta(\omega - k\frac{2\pi}{2P})\right]^2 \tag{3.2}$$

where $P(\omega)$ is the Fourier transform of $p[n]$.

(c) Plot $|P(\omega)|^2$ for $\epsilon = 0$, $0 < \epsilon \ll P$, and $\epsilon = P$.

(d) What is the effect of pitch jitter on the short-time speech magnitude spectrum, $|S_w(\omega)|$? Assume

$$\begin{aligned} s_w[n] &= s[n]w[n] \\ &= (p[n] * h[n])w[n] \end{aligned} \tag{3.3}$$

where $s[n]$, $w[n]$, $p[n]$, and $h[n]$ are the speech signal, windowing signal, irregular impulse train, and vocal tract impulse response respectively.

Prob. 3 (cont.)

Prob. 4

(Quatieri, p. 196) "The sequence $s[n]$ follows an all-pole model, with some poles inside and some poles outside the unit circle. Observe that if we flip all maximum-phase poles inside the unit circle to their conjugate reciprocal locations, then the autocorrelation (sic) remains intact."

Consider a real-valued stable sequence, $s[n]$ and its autocorrelation, $r_s[n] = s[n] * s[-n]$.

(a) Determine the z-transform of $r_s[n]$, $R_s(z)$ and its region of convergence.

(b) Based on (a), argue that if you flip maximum-phase poles inside the unit circle to their conjugate reciprocal locations, then the autocorrelation remains intact (unchanged).

Prob. 4 (cont.)