Prob. 1

The waveform for Name.wav is shown in Fig. 1. Individual phonemes for Fig. 1 are labeled on posted solution. Assume the 40 or so American English phonemes can each be represented by a unique 6 bit codeword. Then for the 18 phonemes in NAME.WAV which span 1.4s we have an information rate of $6 \times 18 \div 1.4 = 77$ bps.

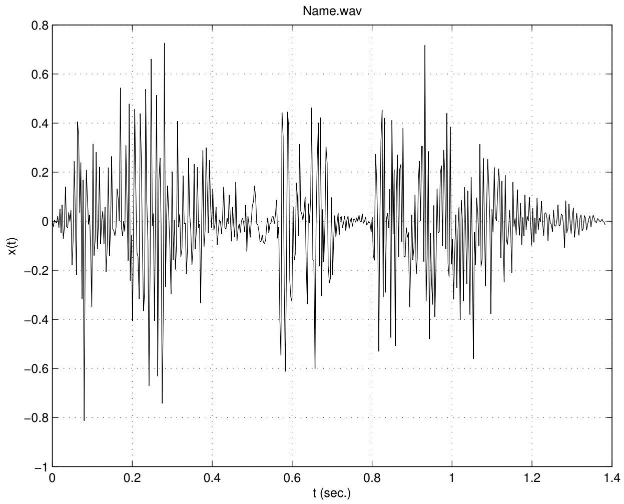


Figure 1: "My name is Phillip De Leon."

Prob. 2

The waveform and periodogram of SPEECH.WAV can be found in Fig. 2. Generally speaking, Prof. De Leon's long-term speech spectrum is low pass with a falloff of approximately 10dB per octave. Your mileage will vary.

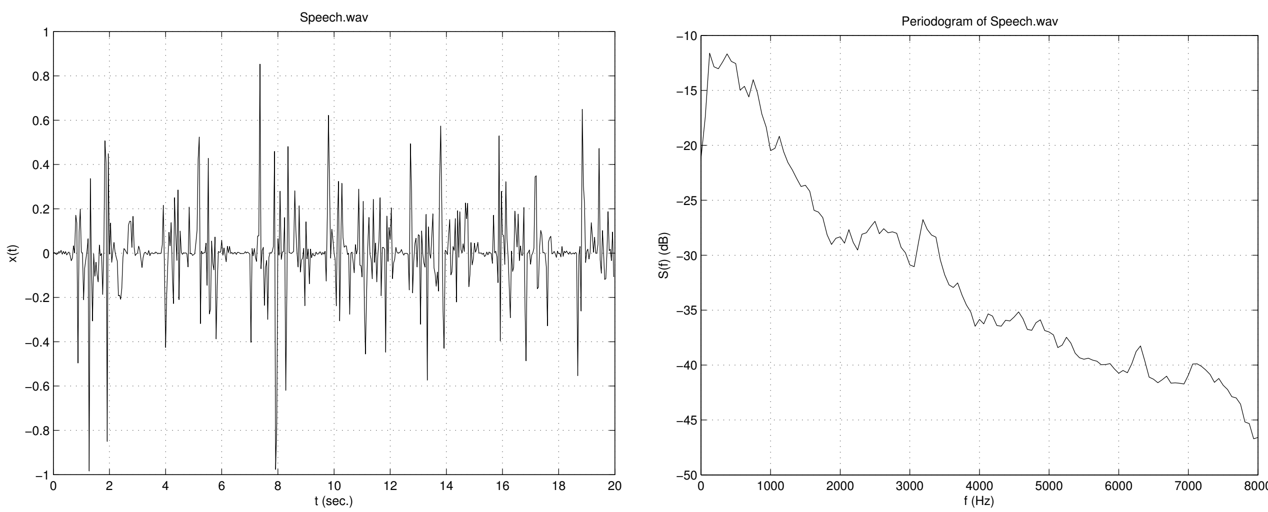
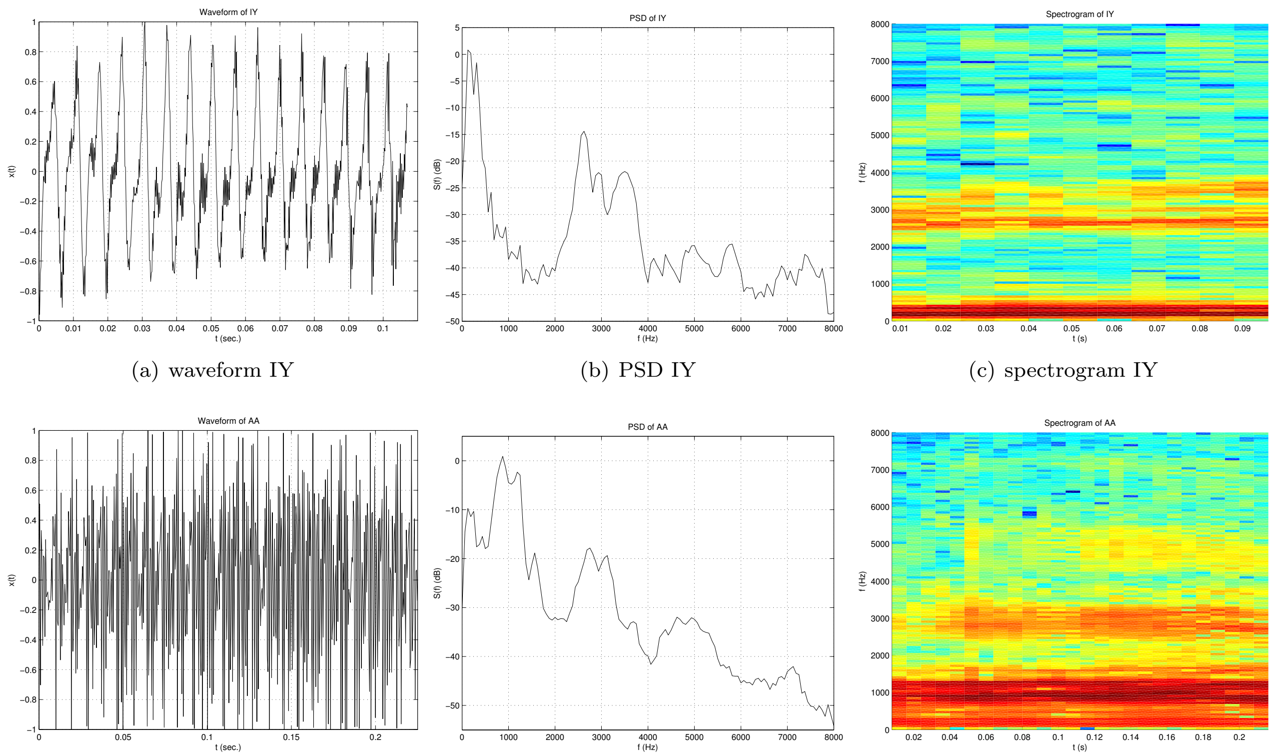


Figure 2: (a) Waveform and (b) periodogram of the first 20s of Speech.wav.

Prob. 3

Waveforms, power spectral densities, and spectrograms for vowel phonemes IY in "beat" and AA in "Bob" are shown in Fig. 3. The fundmental for IY is 155 Hz (14 cycles in 0.09 s) and for AA is 200 Hz(?). Peaks in the PSD show the formant frequencies which are the same as the red bands in the spetrogram.



(a) waveform IY (b) PSD IY (c) spectrogram IY

(d) waveform AA (e) PSD AA (f) spectrogram AA

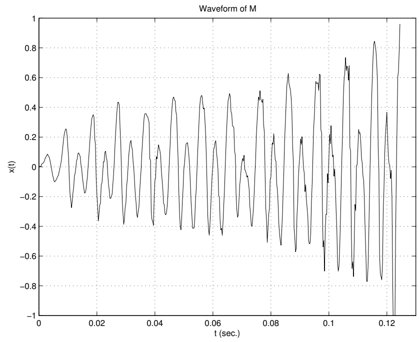
Figure 3: Vowels (front and mid). Phoneme IY (a) waveform, (b) PSD, (c) spectrogram. Phoneme AA (d) waveform, (e) PSD, (f) spectrogram

Prob. 4

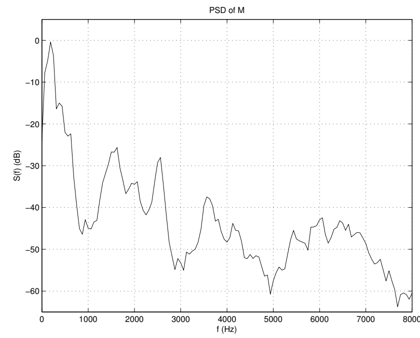
Waveforms, power spectral densities, and spectrograms for nasal phonemes M in "met" and N in "net" are given in Fig. 4. In the example words, both nasals occur during the first 0.1s. As described in the text p. 84, we see a dominant low resonance below 250 Hz for both nasals and little high frequency energy. Also as noted in the text, at the release of the constriction (approximately 0.1s) of the nasal, there is an abrupt change in the spectrogram (not shown).

Prob. 5

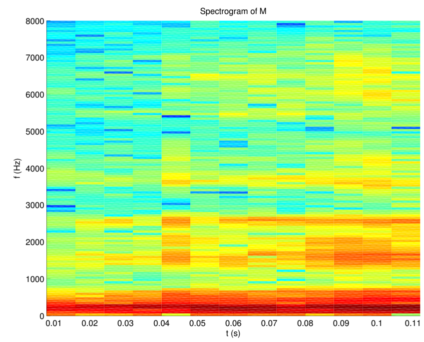
Waveform, power spectral density, and spectrogram for phoneme unvoiced fricative F in "fat" are given in Fig. 5. In the example words, the unvoiced fricative occur during the first 0.2s (0.175s for "F"). As described in the text p. 85, we a "noisy" spectrum. For the "F," there is little front cavity, so the spectrum is almost flat.



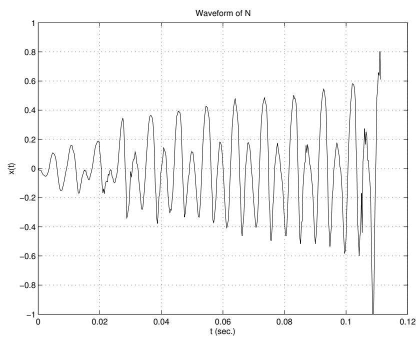
(a)



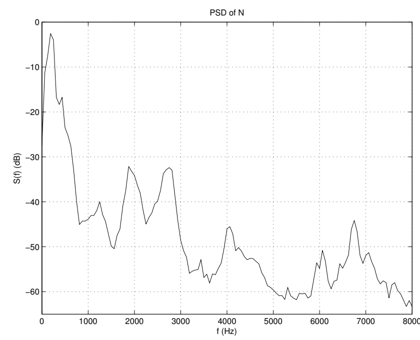
(b)



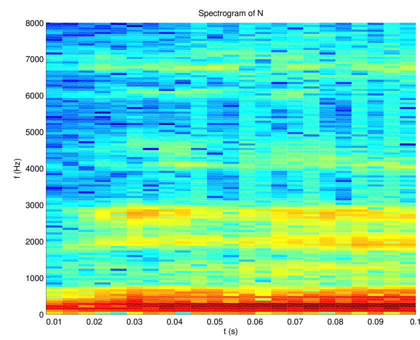
(c)



(d)

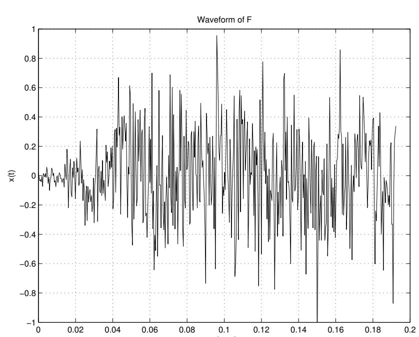


(e)

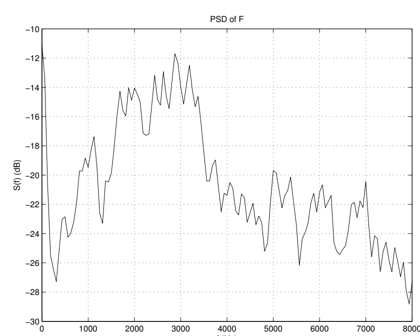


(f)

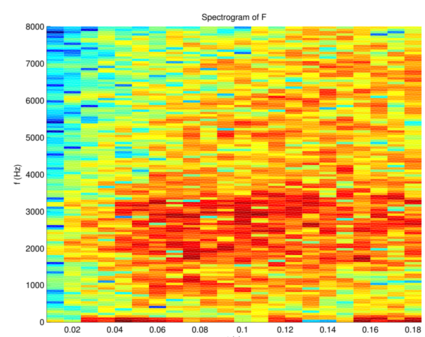
Figure 4: Nasals. Phoneme M (a) waveform, (b) PSD, (c) spectrogram. Phoneme N (d) waveform, (e) PSD, (f) spectrogram



(a)



(b)

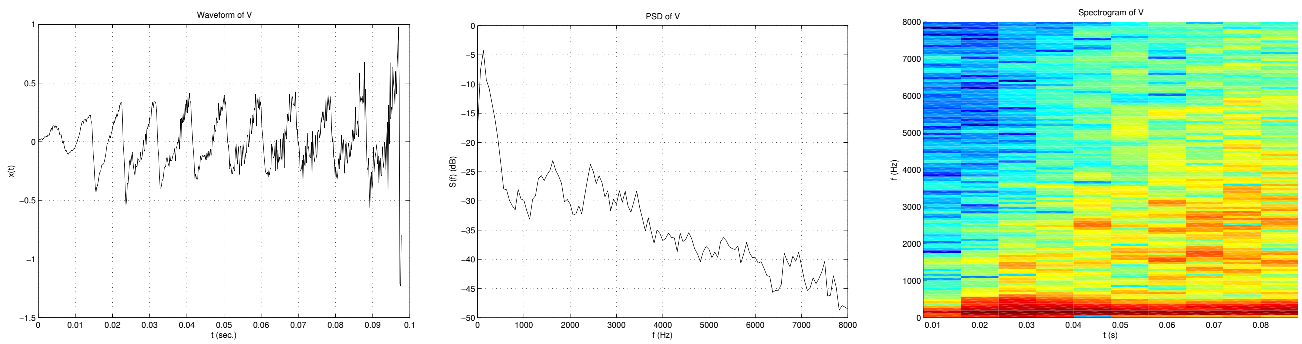


(c)

Figure 5: Unvoiced friactive. Phoneme F (a) waveform, (b) PSD, (c) spectrogram.

Prob. 6

Waveform, power spectral density, and spectrogram for voiced fricative phoneme V in "vat" are given in Fig. 6. For "vat", the voiced fricative occurs during the first 0.075s. As described in the text p. 85, we see a "noisy" spectrum *and* harmonics. The harmonics are clearly visible in Fig. 6.

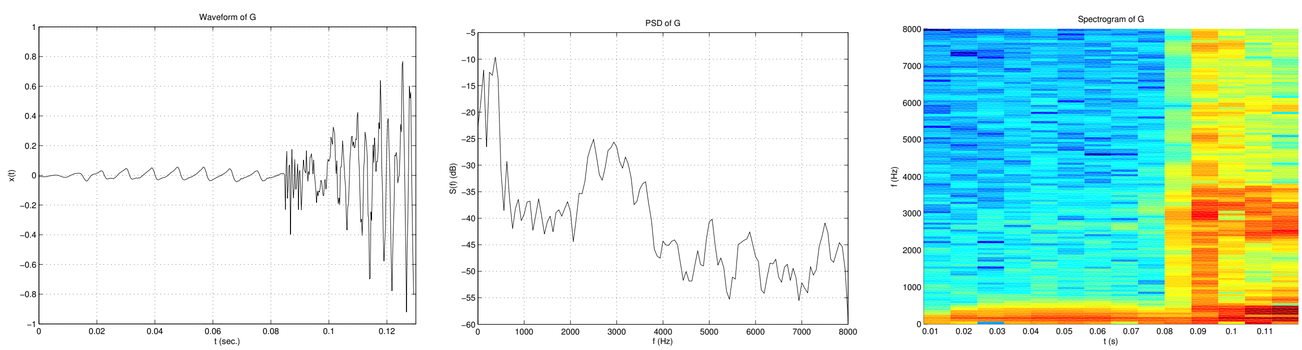


(a) (b) (c)

Figure 6: Voiced friactive. Phoneme V (a) waveform, (b) PSD, (c) spectrogram.

Prob. 7

Waveform, power spectral density, and spectrogram for voiced stop phoneme G in "get" are given in Fig. 7. In the example word, the voiced stop (plosive) occurs at 0.075s. As described in the text p. 90, we clearly see a low frequency "voice bar" prior to the burst.

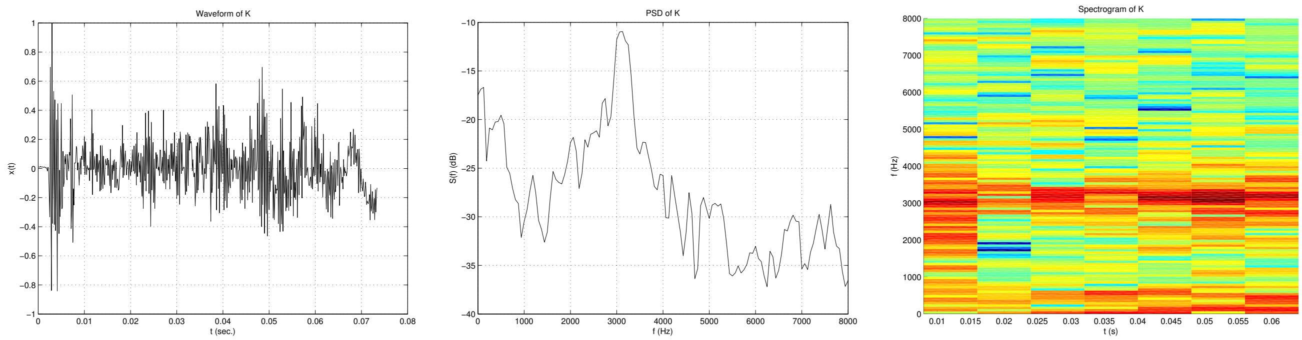


(a) (b) (c)

Figure 7: Voiced stop (plosive). Phoneme G (a) waveform, (b) PSD, (c) spectrogram.

Prob. 8

Waveform, power spectral density, and spectrogram for unvoiced stop phoneme K in "kit" are given in Fig. 8. In the example word, the unvoiced stop (plosive) occurs during the first 0.1s. As described in the text p. 90, we should see a gap of silence, followed by an abrupt burst, and then aspiration noise.

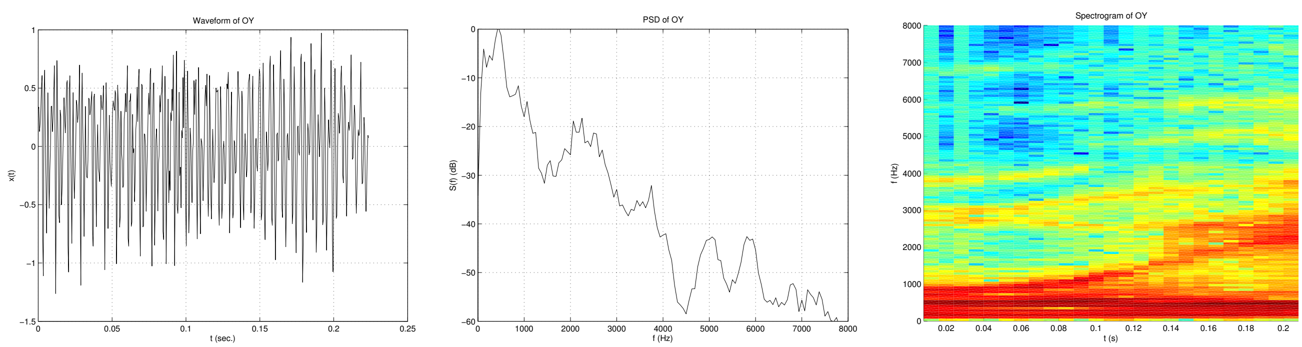


(a) (b) (c)

Figure 8: Unvoiced stop (plosive). Phoneme K (a) waveform, (b) PSD, (c) spectrogram.

Prob. 9

Waveform, power spectral density, and spectrogram for dipthong phoneme OY in "boy" are given in Fig. 9. As described in the text p. 93, we see in the spectrogram a rapid movement of the formants as the vocal tract changes configuration.



(a) (b) (c)

Figure 9: Dipthong honeme OY (a) waveform, (b) PSD, (c) spectrogram

Prob. 10

Waveforms, power spectral densities, and spectrograms for semivowel phonemes Y (glide) in "you" and L (liquid) in "let" are given in Fig. 10. As described in the text p. 93, we see in the spectrogram for the glide Y, a rapid movement of the formants as the vocal tract changes configuration similar to dipthong but faster.

Prob. 11

Formant frequency estimates for vowels are given in Table 1. A plot of f_1 vs. f_2 (first two formant frequencies of the vowels in Prob. 5) is given in Fig. 11 with the "triangle" drawn in, i.e. vowel triangle.

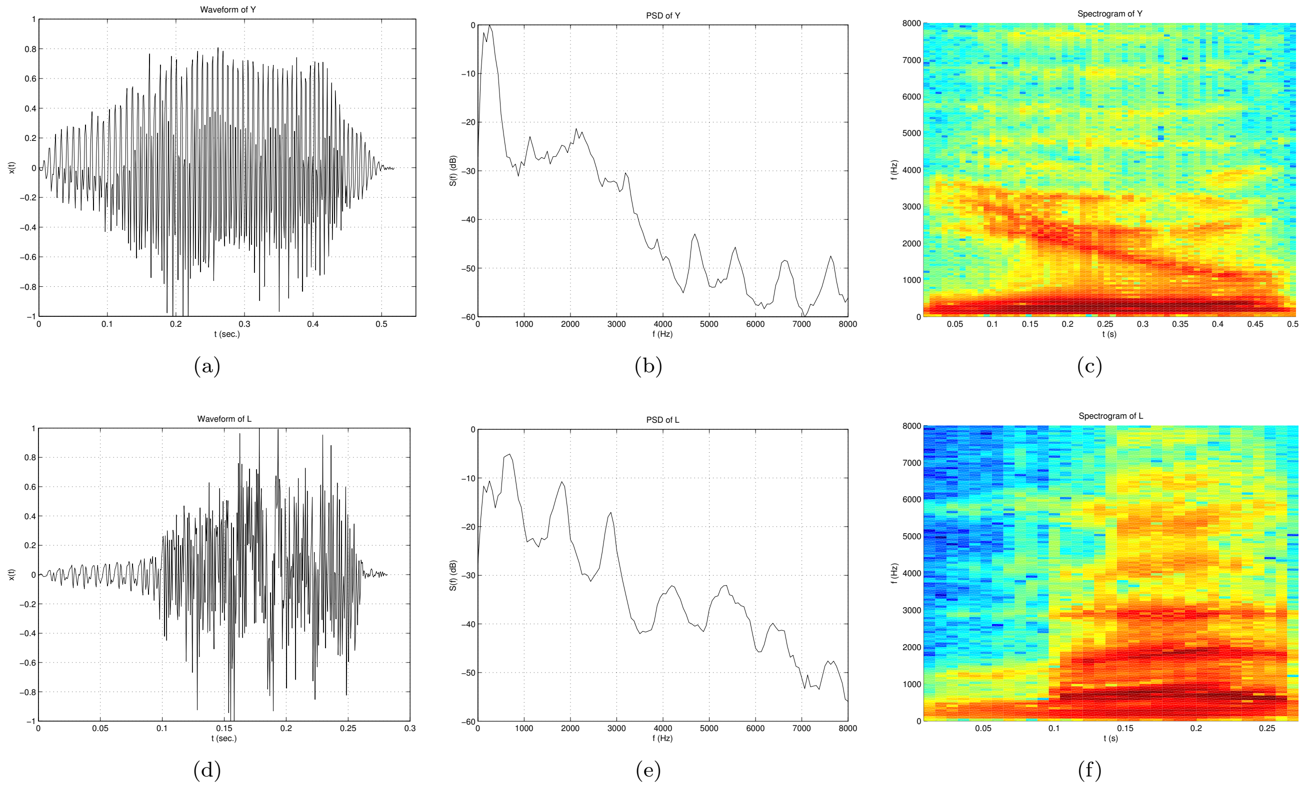


Figure 10: Semivowels. GLide phoneme Y (a) waveform, (b) PSD, (c) spectrogram. Liquid phoneme L (d) waveform, (e) PSD, (f) spectrogram

Phoneme	Example	f_1	f_2
i (IY)	beat	250	2750
I (IH)	bit	500	2000
ϵ (EH)	bet	500	2000
ae (AE)	bat	900	1750
a (AA)	Bob	900	1250
3 (ER)	bird	500	1250
ˆ (AH)	but	700	1250
$\supset$ (AO)	bought	700	1000
u (UW)	boot	250	1250
U (UH)	book	350	1100

Table 1: Formant frequency estimates.

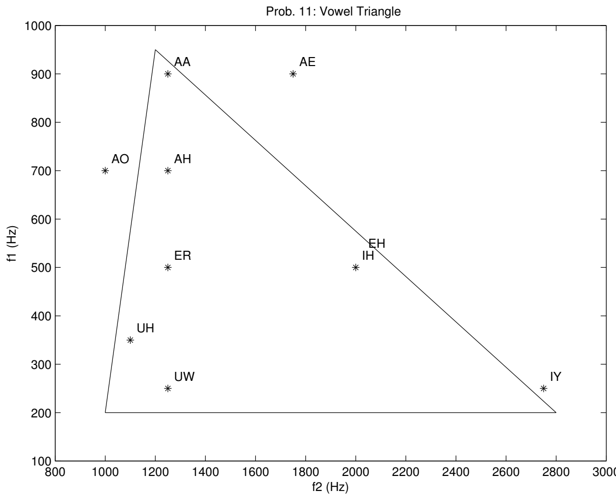


Figure 11: Vowel triangle.

Prob. 12

Error is $1.3573e-12$ for direct inversion (Hamming window with 50% overlap) of STFT for NAME.WAV.

Prob. 13

Error is $5.9046e-13$ for direct inversion (Hamming window with 50% overlap) of STFT for NAME.WAV.

Error is $7.2357e-13$ for direct inversion (Hamming window with 25% overlap) of STFT for NAME.WAV.

Error is $5.1500e-13$ for direct inversion (Hamming window with 75% overlap) of STFT for NAME.WAV.