EEL 6586: HW#2

Assignment is due Monday, Feb 19, 2007 in class. Late homework loses $e^{\#\ of\ days\ late} - 1$ percentage points.

PART A: Short Answer (No more than a few sentences each)

A1 Explain why it is difficult to choose a proper window size in analyzing speech.

A2 Why are overlapping windows typically used in speech analysis?

A3 Assume that a speech signal was framed with a 25ms rectangular window. What is the main lobe width in Hz due to the rectangular window? Assume a 20KHz sampling rate.

A4 A common algorithm for pitch determination is to perform autocorrelation on the LPC residual (error) and look for peaks. Why does this algorithm work better than performing autocorrelation on the original speech signal and looking for peaks?

A5 Explain why humans have no problem determining the pitch of voices through the telephone, even though the cutoff frequency (about 300Hz)is larger than typical pitch frequencies.

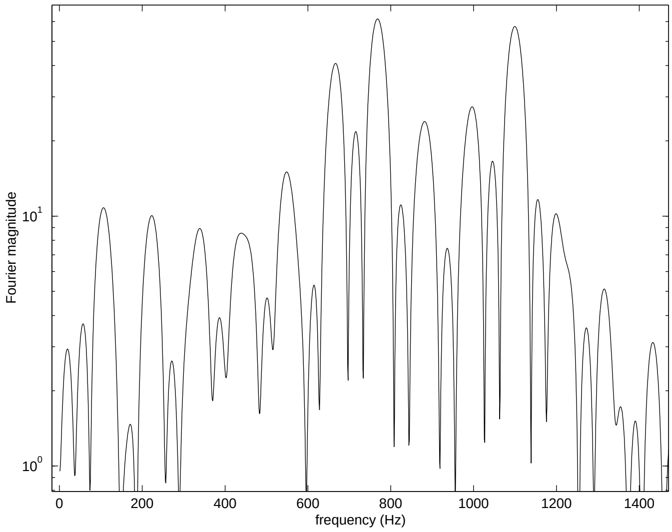
PART B: Textbook problems (Use Matlab only to optionally check your work)

B1 Compute $r[k]$, of the following discrete-time signal:

$$x[n] = \sin(\omega n)$$

Use the definition of autocorrelation in your response (i.e., don't use data samples).

B2 A train of impulses is fed through a simple all-pole model to produce a stationary voiced phoneme. The resulting sequence is windowed by a rectangular window and its Fourier magnitude is plotted as shown below. The sampling rate is 10KHz. Answer the following questions.



a) Estimate the frequency of the first two formants (F1 and F2) in Hz.

b) Estimate the pitch of the voice in Hz.

c) Estimate the size of the rectangular window (in ms) used in processing the speech. Explain your answer.

B3 An infinite train of impulses is created with the following relation

$$e(n) = \sum_k \delta(n + kP)$$

Assume that the sampling frequency is 10kHz.

1. Determine the value of P to create a pitch frequency of 100Hz.
2. The infinite train of impulses is fed through an all-pole model of

$$H(z) = 1/(1 + .9z^{-1} + .81z^{-2})$$

   What is the dominant formant frequency in the signal?
3. Is this formant frequency higher or lower than typical first formant frequencies for humans?
4. How will the formant frequency change if pre-emphasis is applied to the signal $(s(n) - 0.95s(n-1))$?

B4 Assume that an infinite impulse train

$$\sum_k \delta(n + kP)$$

is filtered by a vocal-tract model given by $H(z) = 1/(1+.9z^{-1}+.81z^{-2})$ to produce a speech signal $s(n)$.

1. Derive the difference equation for $s(n)$.
2. Compute the autocorrelation function $r(0)$ for the speech signal $s(n)$.
3. Compute the autocorrelation function $r(1)$ for the speech signal $s(n)$.
4. Compute the single LPC coefficient ($p = 1$) for this system.
5. How does this coefficient compare to the first coefficient when $p = 2$? Explain.

B5 Assume that white noise excitation $w(n)$ is filtered by an *all-pole* vocal-tract model $H(z) = 1/(1 + .25z^{-2})$ to produce a speech signal $s(n)$. $w(n)$ is defined:

$$E\{w(n)w(m)\} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

In this problem you will use LPC to derive an all-pole approximation to $H(z)$.

1. Derive the difference equation for $s(n)$.
2. Compute the autocorrelation function $r(0)$ for the speech signal $s(n)$.
3. Compute the autocorrelation function $r(1)$ and $r(2)$ for the speech signal $s(n)$.
4. Compute the first two LPC coefficients ($p = 2$).
5. (5 points) Derive $\hat{H}(z)$, the all-pole approximation to $H(z)$. Does your answer make sense?

PART C: Computer Analysis of Speech In this part you will write a program for automatic pitch analysis. You will run your program on three recorded sentences at http://www.cnel.ufl.edu/hybrid/courses/EEL6586/sentence.html for an adult male (sentence 1), adult female (sentence 2) and a child (sentence 3). Through the following steps, you will develop a pitch algorithm that processes the autocorrelation of the LPC residue for each window of speech:

C1 Break the sentence into overlapping windows. Describe how you choose the window type, length and overlap for this pitch estimation algorithm.

C2 Compute the LPC coefs for each window and inverse filter the signal in that window to get the residue. Show a typical example of the windowed signal, with plots of the time domain signal, its power spectrum, the smooth envelope from the LPC coefs and the residue (in the time domain).

C3 Run an autocorrelation on the residue signal and show an example plot using the residue from [C2].

C4 Write a procedure that automatically computes the pitch by finding the "first biggest peak" after the lag zero peak. What pitch is detected for your example window from [C3]?

C5 Put all of the pieces together and write an algorithm to compute pitch for each window Write a program that determines the pitch of each window of a sentence (if the pitch exists). Show a plot of F0 (in Hz) vs. time (in seconds). You may need to add an additional filtering step to smooth out the pitch values. Indicate unvoiced regions and silence with a pitch of zero. Hand in plots showing the results on the three sentences. Compute the average pitch of each sentence making sure to only consider the voiced regions.

As always, hand in all of your matlab code as the appendix of your homework. Discuss your algorithms in detail and comment on the accuracy of your algorithms.