EEE 6586 – SPEECH SIGNAL PROCESSING
Computer Assignment #3 (Due Date: See Course Homepage)

I. SHORT-TERM SPEECH PROCESSING

Background Reading: *Theory and Applications of Digital Speech Processing*, L. R. Rabiner and R. W. Schafer, 2011, Chapters 6 and 10 (Sections 10.3-10.5).

Overview: The purpose of this exercise is to extract temporal and/or spectral characteristics or features such as energy, zero crossing rate, autocorrelation, average magnitude difference (AMDF) and discrete Fourier transform that is useful in various speech applications: speech segment classification, endpoint detection, coding and recognition.

Problem 1

For the vowel speech utterance (i.911 data)

a) Compute the short term autocorrelation given by equations 6.29 and 6.35 in text (you can perform correlation in MATLAB using *xcorr* function) for a Hamming window of length 50 ms (any 500 samples in the mid part of the data) for lags m = 0, 1, 2, …, 100.

b) Compute the N = 500 point magnitude spectrum (use MATLAB *fft* function) of the waveform based on a Hamming window and short-term DFT.

c) Repeat steps (a) and (b) after center clipping the waveform according to equation 10.20 and Figures 10.26 and 10.27 in text with clipping level set to 30% of the maximum signal amplitude.

d) Comment on the changes in both the autocorrelation and the spectrum and what these changes indicate about the effects of the clipping operation on the waveform

e) Estimate the pitch using the two autocorrelation results and which would provide a better performance in an automated procedure

Problem 2

For the same vowel data (i.911)

a) Compute the short-term average magnitude difference (AMDF) given by equation 6.43 in text using the same window type, length, and data frame in Problem 1(a) and observe the similarities or differences in their results.

b) Estimate the pitch and make conclusion as to the effect of center clipping here.

Problem 3

For the words two.911 and six.911

a) Compute and sketch the short-term normalized energy (use equation 6.10 in text). Use boxcar window of length N = 300 (30 ms) with no overlap.

b) Can you use the energy plots of the words to distinguish between different sounds (voiced/unvoiced/silence or background noise)?

c) Based on this information obtained in part (b), can you identify the word endpoints? Which of these two words is more difficult to accomplish this? Compare your results with visual inspection.