ECE 797: Speech and Audio Processing

Assignment #1: End-point detection of discrete utterances

The Due date: Report and MATLAB code to be submitted (preferably via email) before class on Thursday, February 5.

The report: Should include plots and descriptions of results and answers to specific questions.

The code: The MATLAB M-files generating the results in your report should be submitted.

The rules: Each student should work on their own code and report. If you get stuck, you can ask for advice from another student. If you get really stuck, you can ask me.

The Assignment:

The problem of locating the endpoints in time (i.e., start and finish) of a discrete utterance is an important problem in many speech processing applications. One simple scheme for endpoint detection makes use of measures of (i) the *short-term energy* and (ii) the *short-term zero-crossing rate* of a signal as a function of time. The values of these two metrics as a function of time are combined to make a decision about the location of the endpoints of a discrete utterance.

1. Write a MATLAB function that implements this endpoint detection scheme and test it out on the speech waveform file supplied on course web site: `oilyrag.wav`. Try to adjust the parameters of your algorithm so that it can detect the endpoints of discrete utterances in the test sentence, where a discrete utterance might be considered:

    a. a phoneme,

    b. a syllable,

    c. a word, or

    d. a sentence.

2. Describe how each of the two metrics (i.e., the short-term energy and the short-term zero-crossing rate) contributes to the success or failure of the detection algorithm for the different utterances in this sentence (at the different utterance levels a–b above). Note that it may help to refer to the sound source and/or phonemic categories described in Chapter 3 of Quatieri.

3. How does the detection algorithm perform in the presence of white, Gaussian background noise generated by the MATLAB function `randn()` at a few different SNRs? For what source/phonemic categories does the background noise cause the most problems? Explain why.