E6820 Speech & Audio Processing & Recognition

Midterm Exam

Date: Monday 2002-03-11

Dan Ellis <dpwe@ee.columbia.edu>

This test consists of 4 questions. All questions have equal weight. You must show all your workings to get credit for an answer. You may use a calculator although it shouldn’t be needed.

This is an open-book, take-home test, meaning that you can do it at a time of your choosing, and you can refer to any notes or books. However, you must spend no longer than 2 hours working on the test, and you should complete the test at a single sitting.

You must return your answers to me by 5pm on Tuesday 2002-03-12. You can drop off your answers in my department mailbox (1312 Mudd) or fax them to the EE department at (212) 932-9421. Your test must be entirely your own work, and you should not discuss it with anyone until after the cut-off time.

If you have any questions, you can call me on (212) 854-8928 or send me email at dpwe@ee.columbia.edu.

1. A sheet of regular paper is rolled into a tube about 1 inch in diameter and 11 inches long, and held in free space so both ends are open.

(a) If a pressure impulse is injected at one end at time \( t = 0 \), sketch the pressure \( p(t) \) at the center of the tube as a function of time.

(b) If you blow over the end of the tube, you may hear a weak pitch. What is the fundamental frequency you expect for this pitch and why?

(c) How would you expect the sound to change if the far end of the tube was blocked off e.g. by placing a hand against it?

2. The following schematic diagram shows a test signal constructed from a set of sinusoids. The four test tones \( A \) through \( D \) are each constructed from 5 simultaneous sine tones of 50 ms duration; tones \( A \) through \( C \) have the sinusoids uniformly spaced in frequency between 880 and 1000 Hz, whereas the sinusoids in tone \( D \) lie uniformly between 820 and 1060 Hz. Each sinusoid in each tone has a level of 50 dB SPL. Sinusoid \( M \) is a masker at 950 Hz of 600 ms duration and 80 dB SPL.

(a) What is the total power of (in dB) of each of the tones \( A \) to \( D \)?
(b) A subject listens to this sound, and is asked to judge the relative loudness of the four tones $A$ through $D$. What result would you expect and why?

(c) If the five sinusoids making up tone $D$ are presented individually, would you expect them to have the same perceived loudness?

(d) Briefly, how might you conduct an experiment to verify your prediction for part (c)?

2. The spectrogram below is of an utterance from the TIDIGITS corpus of spoken digit strings.

(a) What can you say about the STFT window length used to construct this spectrogram?

(b) Describe the major features visible in the spectrogram, giving time, frequency, and what they represent.

(c) Can you tell what word(s) are being spoken?

4. In the figure, the points marked “x” belong to one class, and the points marked “o” belong to a different class.

(a) The bold X and O show the locations of the means of the two classes. Using a simple nearest-class-mean rule, what would be the classification of the points labeled A, B and C?

(b) The dotted lines show the unit standard deviation radius of Gaussians fit to each class. What are the classes of A, B and C under a MAP classifier using these models, if the priors are set to make the peak likelihood of each class equal?

(c) What might be a better approach to building a classifier for this data?