<u>Homework 7 (ECE6255 Spring 2010) Grade=4/100</u>

1. (This is Problem 14.1 of Quatieri simplified to the 1-dim case) Consider a random variable *x*, with a probability density function (pdf), characterized as a Guassian mixture model, defined as

$$p(x) = \sum_{k=1}^{K} \omega_k \bullet \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp[-\tfrac{1}{2\sigma_k^2}(x-\mu_k)^2]$$

   Show that $\sum_{k=1}^{K} \omega_k = 1$, the necessary and sufficient condition for *p(x)* to be a valid pdf [Hint: find the definition of the Gaussian pdf from your textbook on random variables, we need the condition $\int_{-\infty}^{\infty} p(x) = 1$ for *p(x)* to be a valid pdf].

2. Work out Problem 14.3 (a) and (d) of Quatieri [Hint: Study Section 14.3.3; for Part (d), is the background model the same for all speakers?].

3. Let *x(n)*, the input to a linear shift-invariant system *H(z)*, be a stationary, zero mean white noise process with a variance σ_x^2. It is well-known that the autocorrelation function (defined in the following) of a random process is an important tool for signal analysis [Hint: look up the definition of the white noise in your textbook on random variables and autocorrelation functions]. Show that the autocorrelation function of the output, $\phi_y(m)$, can be expressed as a function of the autocorrelation function, $\phi_x(m)$, of the input

$$\phi_y(m) = E[y(n)y(n+m)] = \sigma_x^2 \bullet \sum_{k=-\infty}^{\infty} h(k)h(k+m),$$

   where *h(n)* is the impulse response of the system, and *E*(.) stands for mathematical expectation.

4. The following 4×5 matrix of numbers shows a 20-state network, with each *(i, j)* network element indicating the cost of visiting that state.

0	4	3	6	4
7	8	6	8	8
2	3	1	8	7
6	2	9	3	0

   Suppose it is desired to move from the upper left corner state, i.e., the source at state (1, 1), to the lower right corner state, the sink at state (4, 5), and an individual movement is allowed only one step rightward or downward, but not both (i.e., moving diagonally in the southeast direction). There is also an additional constraint that you cannot make more than three rightward or downward transitions on the same row or same column in the same path (this is similar to a durational or slope constraint in DTW). Find the minimum cost path, the states never visited by any sub-path due to the durational constraint, and the corresponding minimum path cost. Explain in steps how you arrive at the answers. You can break ties arbitrarily when choosing same cost sub-paths. Do you get the same answer if you work backwards? [Hint: Use the string edit example in Lecture 25 to figure out the solution. Make sure you follow the dynamic programming principle discussed in Lecture 25]

5. Write a MATLAB program to compute distance between two utterances using the LPC-derived cepstral vectors to represent speech frames, a Euclidean distance between cepstral vectors as a frame distance, and DTW (dynamic time warping) as a way to evaluate distance between two utterances of unequal lengths. You can use your programs in Problem 5 of HW#6 for LPC analysis. You can use the LPC-derived cepstrum formulation in Lecture 28 (also in Problem 3 of Quiz #2) to implement extraction of cepstral vectors. You can choose the dimension of the cepstral vector to be the same as the order of the LPC analysis, e.g., 16. Use the DP formulation in Lecture 25 to implement DTW. Use the slope constraint ($1/2 \le$ slope ≤ 2) to limit the red admissible path region, i.e., the length of unknown utterance, *U*, has to be at most twice or at least half the length of the reference utterance, *R*. Download the six speech utterances from the class website. Three of them are from male talkers and the other three are from female talkers, all speaking the same sentence, "*She had your dark suit in greasy wash water all year*". You are asked to:
   a) Compute the pair-wise 6x6 distance matrix, i.e. given utterance *k*, compute the overall average frame distance *D(k, l)*, using DTW, allowing horizontal, vertical and diagonal transitions. The average frame distance is obtained by dividing the total utterance distance with the total number of frames for utterance *k*; [Hint: *D(k, k)=0*. The frame distance between the *i*-th cepstral vector of the reference utterance and the *j*-th frame of the unknown utterance can be considered as the cost for a warping path to visit that state when performing DTW].
   b) Based on the results in Part (a), find the average speaker distance between female and male speakers for each speaker, and for each gender group. Comment on the effect of gender difference on comparing utterance distances.