Automatic Speech Recognition Midterm

1. Pick the ASR task with the *higher* accuracy, assuming unmentioned task parameters are the same, from the following pairs.

   (a) clean speech vs. noisy speech

   (b) speaker-independent vs. speaker-dependent

   (c) large vocabulary vs. small vocabulary

   (d) read speech vs. spontaneous speech

   (e) isolated speech vs. continuous speech

2. (a) Show that if $y[n]$ is the convolution of two sequences $x[n]$ and $h[n]$, then the Fourier transforms satisfy $Y(e^{jw}) = X(e^{jw})\ H(e^{jw})$.

   (b) Let $W(z) = 3X(z) + 2z^{-2}Y(z)$. What is the relation between $w[n], x[n]$ and $y[n]$?

$$w[n] = 3x[n] + 2y[n-2].$$

3. What is the inverse z-transform of

$$\hat{H}(z) = \log(1 - az^{-1})?$$

$$\hat{h}[n] = -\frac{a^n}{n},\ n = 1, 2, \ldots.$$

4. Show that the number of complex multiplications in a radix-2 N-point FFT is $\frac{N}{2}\log N$.

5. Let X have a geometric distribution,

$$Pr(X = n) = p(1-p)^{n-1},\quad n = 1, 2, \ldots$$

   (a) Compute the expectation value of X.

$$EX = \sum_{n=1}^{\infty} n\ Pr(X = n) = \frac{1}{p}.$$

   (b) Given a set of N independent samples $\{x_1, \ldots, x_N\}$ of X, express the maximum-likelihood estimator of p in terms of the samples.

$$\log P(D|p) = \sum_{n=1}^{N} \log p + \log(1-p)^{x_i - 1} \Rightarrow p^* = \frac{N}{\sum_i x_i}$$

6. Suppose X_1 and X_2 are Gaussian random variables generators with distributions $X_1 \sim \mathcal{N}(-1, 1)$ and $X_2 \sim \mathcal{N}(1, 9)$. Let X_1 or X_2 be randomly chosen and a sample of $x = -2$ is generated. What is the probability that this sample comes from X_1,

   (a) if X_1 and X_2 are equally likely to be chosen? $\frac{3}{4}$

   (b) if X_1 is two times more likely than X_2 to be chosen? $\frac{6}{7}$

7. In a pattern classification problem, a pattern is classified to be one of K classes. The prior class probability is $p(k)$ and the class conditional probability is $q_k(x)$, $k = 1, \ldots, K$. Given a pattern x_0,

   (a) what is the probability of error if it is decided to be of class 1?

   $$1 - P(k = 1|x_0).$$

   (b) what decision minimizes the probability of error?

   $$k^* = \arg\min_k (1 - P(k|x_0)) = \arg\max_k P(k|x_0).$$

8. Show that the function $y(x, t) = f(x - ct) + g(x + ct)$ satisfies the wave equation.

   $$\text{Note } \frac{\partial y}{\partial x} = \frac{df}{du}\frac{\partial u}{\partial x} + \frac{dg}{dv}\frac{\partial v}{\partial x},$$

   where $u = x - ct, v = x + ct$. Continuing, one can show that

   $$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}.$$

9. Explain the following terms.

   (a) spectrum

   (b) basilar membrane

   (c) critical band

   (d) equal-loudness curve

   (e) tonotopical

10. (a) What is the SPL for a tone of 400 Hz to be just audible? 10 dB.

    (b) For a normal human subject, what is the rank of the loudness for tones of 100, 2000 and 7000 Hz at SPL of 40 dB?
    $2000 > 7000 > 100$ Hz.