EECS 225d: Midterm, March 19, 1999

Name:

The total test time is estimated as 50 minutes, but you may stay afterwards and work on the test for as long as you would like (up to an additional hour) since the room is free that long.

The test is composed of two main parts. The first part consists of small problems with a point total of 26. The second part consists of 6 larger problems that are worth 8 points each: choose 3 to answer (maximum score for part II is 24).

For multiple choice (or true/false) questions, also write a brief (1-2 sentence) explanation of why the answer is correct; a correct answer without an explanation will only get half-credit. The parenthesized numbers after each question number give the relative point value of each question.

PART I: Small Problems

1. (2) An idealized (uniform cross-sectional area tube) vocal tract with a length of 17 cm has its first three resonances at 500 Hz, 1500 Hz, and 2500 Hz. An 8.5 cm vocal tract (for a small child) would have these resonances at

   a) 250, 750, and 1250

   b) 1000, 3000, and 5000

   c) 250, 1250, and 2250

   d) 750, 1750, and 2750

   $\Rightarrow$ **The answer is b), since resonant frequencies (modes) are inversely proportional to the length of a uniform acoustic tube.**

2. (2) The ceiling of a cubical room is lifted by a crane so that it is twice as high, and the walls are extended up (to meet the new ceiling position) by using perfectly reflective materials. Ignoring air absorption, frequency dependencies, nonlinearities, or spatial dependencies, what is the effect on the reverberation time (e.g., RT60) of this change?

   a) increase the reverberation time for short sounds and decrease it for long sounds

   b) double the reverberation time

   c) halve the the reverberation time

   d) increase the reverberation time by a factor of 4.

   $\Rightarrow$ **The answer is b), since the effect is to double the volume while keeping the total absorption the same, and** $RT60 = \frac{.049V}{S\bar{\alpha}}$**.**

3. (2) Two noise monitoring consultants are checking on noise pollution outside of this building using sound pressure level (SPL) meters. Consultant number 1 is using a meter that measures A-weighted SPL, while consultant number 2 is using a meter that measures unweighted SPL. If the noise is primarily below 100 Hz, which would be the higher level?

   a) The A-weighted SPL

   b) The unweighted SPL

   c) The meters would show the same value

   d) The meters would explode, leaving the city council no choice but to close down the university.

   $\Rightarrow$ **The answer is b), since the A-weighted SPL downgrades the significance of low frequency signals, e.g., below 100 Hz.**

4. (2) A multivariate Gaussian is fitted to the examples in each of two classes. These two Gaussians are then used to design a minimum probability of error classifier. Under which condition is the decision surface for this classifier guaranteed to be a hyperplane (a linear classifier)?

   a) The variance for each feature is the same

   b) The covariance matrix is non-diagonal but equal for each class

   c) The covariance matrix is diagonal but different for each class

   d) The data was collected in Cincinnati

   $\Rightarrow$ **The answer is b), even though the covariance matrix is non-diagonal, since the quadratic term of the discriminant function is the same for each class if the matrix is the same. Setting these discriminant functions to be equal, the quadratic terms will cancel.**

5. (2) The articulation index (AI) for nonsense syllables filtered from 500 Hz to 3 kHz is x, and for filtering from 3 kHz to 6 kHz is y. According to Fletcher's AI theory, the AI for syllables filtered from 100 Hz to 6 kHz is

   a) $x - y$

   b) $x + y$

   c) at least $x + y$

   d) at least $x \times y$

   ⇒ **The answer is c), since the AI is additive, and $x + y$ would then be the AI for the range from 500 Hz to 6 kHz.**

6. (2) Lower spectral resolution at high frequencies is a common property of

   a) Linear predictive analysis

   b) Critical band filter banks

   c) Cepstral analysis

   d) Swedish basketball players

   ⇒ **The answer is b). Critical bands have been measured to be about 100 Hz for frequencies below 800 Hz. For higher frequencies, bandwidths increase with center frequencies. At 10 kHz, the critical bandwidth is well over 1 kHz.**

7. (2) A Multi-Layer Perceptron (MLP) is composed of units that compute a sigmoidal nonlinearity $f(y) = \frac{1}{1+e^{-y}}$, where y is a weighted sum of inputs to the unit. Which of the following statements is true about this system?

   a) The chain rule of differentiation cannot be used directly to optimize the network weights, due to the nonlinearity.

   b) The outputs of the system can be anywhere in the range [-1,1].

   c) The input layer is often called "hidden" because it is not directly observable from the network outputs.

   d) When the number of weights in the MLP is large in comparison with the number of training patterns, a large number of training epochs can lead to overfitting of the training data.

   $\Rightarrow$ **a) is not true, since the sigmoid is differentiable. b) is not true, because the outputs of the sigmoid cannot be less than 0. c) is not true because a "hidden" layer refers to layers between the input and output. d) is true because a very detailed model can be overtrained, leading to (essentially) memorization of the peculiarities of the relatively small training sample. Thus the resulting parameters will "overfit" the training data and will not generalize well to an independent sample.**

8. (2) Consider a 10th order linear predictor, used for speech that is sampled with a 10 kHz sampling rate. Choose the condition under which the predictor would be expected to have the largest error signal:

   a) when the speech is low in energy

   b) for female speech

   c) at the start of each new period of voiced speech

   d) in the middle of each new period of voiced speech

   $\Rightarrow$ **The answer is c). The accuracy of linear prediction is a function of the signal autocorrelation, and is based on the hypothesis that the underlying structure of the system does not change over the measurement interval. The introduction of a glottal pulse does change the environment, and the resulting signal is less predictable from a weighted sum of the 10 previous points. Thus the error is larger.**

9. (2) Critical bandwidths based on perceptual experiments

   a) increase logarithmically over the complete range of hearing

   b) increase like a $\frac{1}{3}$ octave bank of filters

   c) remain constant over most of the frequency range

   d) remain constant for low frequencies and increase logarithmically thereafter

   $\Rightarrow$ **The answer is d), based on many measurements.**

10. (2) True or False: The perception of pitch is only possible when listening to a harmonic series.

    $\Rightarrow$ **False. Pitch is perceivable even if the signal is NOT perfectly harmonic.**

11. (3) A typical auditory fiber has a refractory period of about 3 ms. This means that no new firings can occur for 3 ms after a spike. Explain how the auditory system manages to perceive pitch as high as several kilohertz despite this constraint.

   $\Rightarrow$ **The human peripheral auditory system contains about 30000 neurons. Evidently, more centrally located neurons can process the information arriving from many neurons that are in synchrony to the stimulus.**

12. (3) Define 3 of the following 5 terms (1 sentence per definition):

a) neural phase-locking $\Rightarrow$ **Auditory neurons tend to spike at the peaks of an increasing periodic tone.**

b) critical band $\Rightarrow$ **The frequency band surrounding a tone such that changing the noise bandwidth <u>outside</u> the band does <u>not</u> affect the tone threshold of hearing.**

c) simultaneous masking $\Rightarrow$ **Masking of a tone by noise when noise and tone are simultaneously affected.**

d) cochlear apex $\Rightarrow$ **The "inside" of the "snail" (cochlea) - i.e., the last place on the spiral to be stimulated by an acoustic input. This is the end that is more sensitive to low frequency energy.**

e) forward masking $\Rightarrow$ **Masking of a tone by noise in which the noise (the masker) is applied <u>before</u> the signal to be detected (the target).**

PART II: Larger problems (choose 3 to do)

13. (8) Let $h(n) = \alpha^n + \beta^n$.

a) Find the z-transform $H(z)$ of $h(n)$

b) Under what conditions is the digital filter stable? Unstable?

c) If $\alpha = 0.8$ and $\beta = -0.8$, find the z-plane pole-zero plot of $H(z)$.

d) Sketch the response along the unit circle.

$\Rightarrow$

a) $H(z) = \frac{1}{1 - \alpha z^{-1}} + \frac{1}{1 - \beta z^{-1}} = \frac{2 - (\alpha+\beta) z^{-1}}{(1 - \alpha z^{-1})(1 - \beta z^{-1})} = \frac{2z - (\alpha+\beta)}{(z - \alpha)(z - \beta)}$

b) Filter is stable (poles inside the unit circle) when $|\alpha| < 1$ and $|\beta| < 1$.

c) Poles are at $z = \pm 0.8$, zeros are at $z = 0$.

d)

14. (8) Consider the finite sequence $x(n)$, where $x(0) = 1$, $x(1) = 1$, $x(2) = 1$, $x(3) = 2$, $x(4) = 1$, $x(5) = 1$, and $x(6) = 1$.

   a) Find the *linear* convolution of $x(n)$ with itself, using direct convolution.

   b) Describe an algorithm that should yield the same result, using the DFT.

   $\Rightarrow$ **a)**

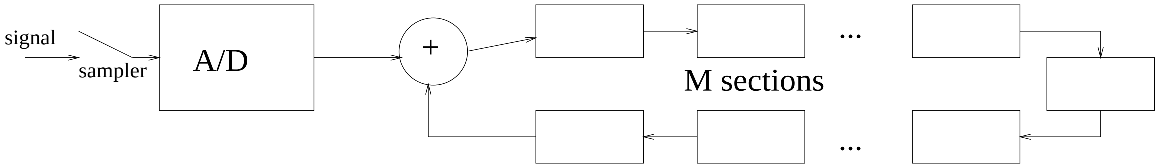
   b) The convolution can be implemented by DFT, by first augmenting with 6 zeros:

   i)

   ii) take DFT of this sequence

   iii) multiply the DFT by itself

   iv) the inverse DFT should be the desired linear convolution



15. (8) An acoustic tube is represented digitally by a spatial pattern.

    Describe the conditions required to prevent aliasing in time and in space.

    ⇒ **The Nyquist condition for sampling an analog signal is to have a sampling rate of at least twice the bandwidth. So, by knowing the spectrum of the input signal, the rate of sampling to obtain aliasing can be obtained.**

    Now, if each new incoming sample allows a shift from one section to the next, you have not created aliasing in space.

16. (8) Describe the EM procedure as it would be used to determine the parameters that maximize the likelihood of the data. Your discussion should not depend on any particular form for the probability estimators or distributions, but should otherwise describe the process as clearly as possible.

$\Rightarrow$ **Let k be the hidden variables, X the observed variables, and Θ the parameters to be optimized. Then**

(a) Choose some distributional form and initial parameters Θ_{old}.

(b) Compute the probabilities $P(k|X, \Theta_{old})$.

(c) Develop expressions of the expectation of $P(k, X|\Theta)$ over the distribution $P(k|X, \Theta_{old})$.

(d) Generate equations by setting equal to zero the partial derivatives of this expression, augmented with Lagrangian terms to represent constraints.

(e) Solve the resulting equations to get the new parameters

(f) Assess the resulting system, and return to (b) if the stopping criterion is not met.

17. (8) Describe Voice Onset Time (VOT). Why did experimental results on VOT measurements lead to the concept of "categorical perception?" Describe any physiological evidence that supports the notion that VOT is perceived in connected speech.

$\Rightarrow$ **Voice Onset Time is the time interval from the plosive burst to the onset of voicing of the following vowel. Psychoacoustic experiments (by Haskins Lab, using their Pattern Playback) yielded "categorical" results; when the VOT exceeded a threshold, the perception of the sound changed suddenly. Thus, although the VOT was varied <u>continuously</u>, the percept was <u>quantized</u> to one of the plosive sounds.**

Much later (1995), Delgutte measured PSTH for connected speech (e.g., "Joe took Father's green shoe bench out") and showed that the time between the onset of high CF neurons spiking and low CF neurons spiking was equal to the psychoacoustically measured VOT.

18. (8) Speech production, music production by string, and music production by horns (trumpets, trombones, etc.) have in common that an excitation signal drives an acoustic device. In each of these 3 cases, describe the excitation signal or signals and the corresponding acoustic device.

$\Rightarrow$

(a) **Speech Production:**

During vowels, the vocal tract is primarily open and the pressure gradients caused by vocal cord oscillations provides the excitation. During fricatives, a small section of the vocal tract is constricted (not completely closed) enough to cause turbulent (noisy) air flow. The noise produced at the constriction excites the entire vocal tract. During plosives, the constriction is complete, causing a pressure buildup, followed by a transient burst on release. Again, this excitation propagates through the entire vocal tract.

(b) **Strings:**

If a string is plucked, is vibrations are transmitted to the instruments' body via the bridge and the sound heard is based on the vibration of the body and the air inside the body. The string may be plucked, bowed, or struck; in each case the string vibrations serve as the excitation.

(c) **Horns:**

Excitation signals are produced by lip vibrations that send pressure waves into the horn. In the case of speech or strings, the excitation signals are usually modified independently of acoustic tube response, while in horns, the pressure waves produced by the horn feed back and significantly influence the lip vibration.