EECS 225d: Midterm 2, April 25, 1997

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 (say, up to an additional hour) since the room is free then.

For multiple choice questions, also write a brief (1-2 sentence) explanation of why the answer is correct. The parenthesized numbers after each question number give the relative point value of each question.

1. (2) In cepstral analysis of speech, the vocal tract characteristics would most clearly be seen in

   a) the zeroth coefficient

   b) low-time coefficients

   c) high-time coefficients

   d) equally in all coefficients

2. (2) LPC spectral analysis

   a) yields a pole-only representation

   b) can be more affected by pitch for children than for adults

   c) tends to model spectral peaks better than valleys

   d) all of the above

3. (2) A local slope constraint for DTW could be

   a) The start of an input word cannot be matched to the end of a reference template

   b) Subtract the local distance from the best global distance

   c) Cross-country skis are not permitted for the higher slopes

   d) Predecessor global distances used to compute the current global distance are constrained to the previous or current input or reference frames.

4. (2) A difference between a Hidden Markov Model (HMM) and a Markov Model (MM) is

   a) HMMs have better hiding places.

   b) MMs don't have observations.

   c) MM observations are deterministic, given a state transition

   d) HMM states have stochastic transitions

5. (2) Mel cepstra, PLP cepstra, and LPC cepstra are all sometimes used for the front end of ASR systems. A major characteristic of all three analysis methods (in contrast to power spectra estimated as the squared magnitude of the DFT of the windowed data) is

   a) the effects of fundamental frequency are reduced

   b) coarser resolution at high frequencies than at low frequencies

   c) easier access to the excitation characteristics

   d) all three were invented by Swedish basketball players

6. (2) The frequency response in the peripheral auditory system is tonotopic (meaning center frequency changes with place) because:

   a) the frequency response of the different hair cells varies.

   b) basilar membrane stiffness decreases with distance from the base.

   c) synchrony of auditory fibers decreases with frequency.

   d) none of the above; auditory filter bandwidths are a function of cortical neurons.

7. (2) 2400bps was chosen as a standard for secure speech (in the 1950's) because

   a) 2400bps resulted in very good synthetic speech.

   b) it matched the rates required by a teletype when sending real-time transcribed speech

   c) speech was marginal but telephone line modems could not accommodate higher bit rates.

   d) 2400bps could easily be sent down ordinary telephone lines without the need for modems.

8. (4) Explain the primary differences between the multipulse, CELP, and VSELP vocoders.

9. (5) The forward recursion to compute model likelihoods can be expressed as

$$\alpha_n(\ell) \;=\; \sum_{k=1}^{L} \alpha_{n-1}(k) P(q_\ell^n, x_n | q_k^{n-1}, X_1^{n-1}, M) \tag{1}$$

where X_1^{n-1} means the sequence of acoustic vectors $x_1, x_2, ...x_{n-1}$, the notation q_ℓ^n means the state at time n with category ℓ, and where there are L different state categories.

As it is commonly implemented, however, it is often expressed as

$$\alpha_n(\ell) \;=\; \sum_{k=1}^{L} \alpha_{n-1}(k) P(q_\ell^n | q_k^{n-1}, M) P(x_n | q_\ell^n, M) \tag{2}$$

Show the steps necessary to go from the first formulation to the second. For each step say whether any assumptions are required for the equality, and if so, say what they are.

10. (3) Define the following terms (1 sentence per definition):

    a) phoneme

    b) EM

    c) analysis by synthesis

11. (4) The phrase "carrier nature of speech" was proposed by Dudley as a way of explaining how a vocoder could represent speech using fewer bits (or less bandwidth). Explain how the two of the three major algorithms (channel vocoders, LPC vocoders and cepstral vocoders) implement this concept and, as a result, represent the speech signal with fewer bits (or less bandwidth) than a standard telephone channel or a standard PCM system.

12. (2) Describe at least 3 important steps in a text-to-speech synthesizer

13. (4) Draw a block diagram of an isolated word speech recognition system based on discrete probability distribution Hidden Markov Models. Explain the function of each of the constituent parts.

14. (4) Describe at least two distinct ways in which vector quantization (VQ) is used in different types of vocoders. Briefly compare this to their use in discrete density HMM-based speech recognizers. (I.e. are there similar reasons that VQ is effective in both cases? Is it useful for similar reasons?)