

Digital Speech Processing, Midterm

Dec 2, 2005, 10:00-12:00

- OPEN EVERYTHING
 - 除專有名詞可用英文以外，所有文字說明一律以中文為限，未用中文者不計分
 - Total points: 120, Time allocation: 1point/min
-

1. (10) Explain the concept of "Corpus-based Text-to-Speech Synthesis", how it works and why it is good.
2. (20) Given a HMM $\lambda = (A, B, \pi)$, an observation sequence $\bar{O} = o_1 o_2 \dots o_t \dots o_T$ and a state sequence $\bar{q} = q_1 q_2 \dots q_t \dots q_T$, define

$$\alpha_t(i) = \text{Prob}[o_1 o_2 \dots o_t, q_t = i | \lambda]$$
$$\beta_t(i) = \text{Prob}[o_{t+1} o_{t+2} \dots o_T | q_t = i, \lambda]$$

- (a) (10) Let $\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)}$, Explain what $\gamma_t(i)$ is, where N is total number of states.
 - (b) (10) Formulate and describe the Viterbi algorithm to find the best state sequence $\bar{q}^* = q_1^* q_2^* \dots q_t^* \dots q_T^*$ giving the highest probability $\text{Prob}(\bar{q}^*, \bar{O} | \lambda)$. Explain how it works.
3. (10) What is LBG algorithm and why is it better than K-means algorithm?
 4. (10) Explain why and how the unseen triphones can be trained using decision trees.
 5. (10) Explain the meaning of the perplexity of a language model with respect to a testing corpus.
 6. (10) Explain the principles and procedures of estimating the probabilities for unseen events in Katz smoothing.
 7. (10)
 - (a) (5) What are the voiced/unvoiced speech signals and their time-domain waveform characteristics?
 - (b) (5) What is pitch in speech signals and how is it related to the tones in Mandarin Chinese?
 8. (10) Write down what you know about the techniques for speech end point detection.

9. (10) Explain how the tree lexicon can be used in the search algorithm for large vocabulary continuous speech recognition and how it is helpful.
10. (20) Write down anything you learned about the following subjects which were **NOT** mentioned in the class. Don't write anything mentioned in the class.
 - (a) Conversational interfaces
 - (b) Search problem/algorithms for large vocabulary continuous speech recognition