# Introduction to Digital Speech Processing, Midterm Exam

Nov. 28, 2018, 10:00-12:00

- OPEN Lecture Power Point (Printed Version) and Personal Notes
- You have to use CHINESE sentences to answer all the questions, but you can use English terminologies
- Total points: 100

---

1. Take a look at the block diagram of a speech recognition system in Figure 1.

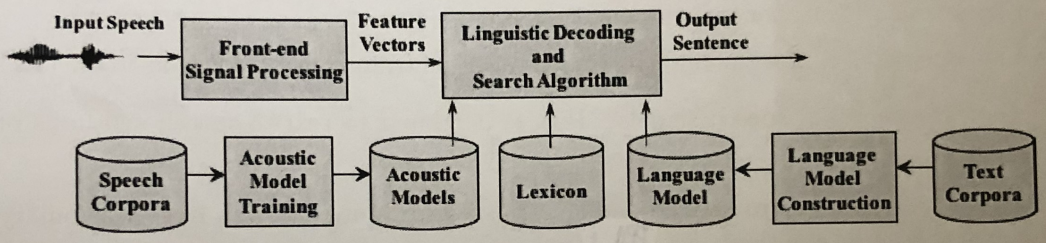


Figure 1: A Speech Recognition System

(a) In the block of front-end processing, why do we use the filter-bank? (4%) 大抄 P.44 講義 P.70,

(b) Explain the rules of the acoustic models, lexicon, and language model in Figure 1? (12%)

(c) Why do we need smoothing in the language model? (2%) 講義 Ch.6

(d) Which part includes the HMM-GMM? (2%) Acoustic models

2. Given a HMM $\lambda = (A, B, \pi)$ with $N$ states, 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) What is $\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)$? Show your results. (4%) $P(\bar{O}|\lambda)$

(b) What is $\frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}$? Show your results. (4%) $P(q_t=\bar{i}|\bar{O},\lambda)$.

(c) What is $\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)$? Show your results. (4%) $P(\bar{O}, q_t=\bar{i}, q_{t+1}=\bar{j}|\lambda)$

(d) Formulate and describe the Viterbi algorithm to find the best state sequence $\bar{q}^* = q_1^* q_2^* \cdots q_t^* \cdots q_T^*$ giving the highest probability $\text{Prob}[\bar{O}, \bar{q}^*|\lambda]$. Explain how it works and why backtracking is necessary. (4%) 大抄 P.

大抄 P43. Explain what is a tree lexicon and why it is useful in speech recognition. (8%)

大抄 P36. 4. (a) Given a discrete-valued random variable $X$ with probability distribution

$$\{p_i = \text{Prob}(X = x_i), i = 1, 2, 3, \ldots, M\}, \quad \sum_{i=1}^{M} p_i = 1$$

Explain the meaning of $H(X) = -\sum_{i=1}^{M} p_i[\log(p_i)]$. (4%)

(b) Explain why and how $H(X)$ above can be used to select the criterion to split a node into two in developing a decision tree. (4%)

大抄 P43 5. (a) What is the perplexity of a language source? (4%)

(b) What is the perplexity of a language model with respect to a corpus? (4%)

(c) How are they related to a "virtual vocabulary"? (4%)

大抄 P43 6. Please answer the following questions.

(a) Explain what a triphone is and why it is useful. (4%)

(b) Explain why and how the unseen triphones can be trained using decision trees. (4%)

7. What is the prosody of speech signals? How is it related to text-to-speech synthesis of speech? (6%)

見下 8. Explain why and how **beam search** and **two-pass search** are useful in large vocabulary continuous speech recognition. (8%)

大抄 P1. 9. Please briefly describe **LBG algorithm** and **K-means algorithm** respectively. Which one of the above two algorithms usually performs better? (Explain your answer with **descriptions**, not just formula only.) (8%)

10. Homework problems (You can choose either HW2-1 or HW2-2 to answer)

**HW2-1**

(a) We added the sp and sil model in HW2-1. How can they be used in digital recognition? (2%) short pause, silence, 加入效果變好

(b) Write down two methods to improve the baseline of the digital recognizer and explain the reason. (4%) 增加 state_num, GMM

**HW2-2**

(a) Why do we use Right-Context-Dependent Initial/Final to label? (2%)

(b) What characteristics can we use to help distinguish the Initials and Finals? (4%)

8. a. beam search: 一般來說，有兩種方式，①設定寬度L，只走前L名
②設定條件，$P < D_{max} - Th$ 會被刪除

b. two pass search: 先用較簡單的model做一次粗估找出N條解，
再用較精準的model做出最佳解

7. Prosody：韻律，以較來說都是抑揚頓挫，
要follow corpus-based Text-to-Speech Synthesis (大抄 P1)