UNIVERSITY OF SURREY©

Faculty of Engineering and Physical Sciences
Department of Electronic Engineering

Undergraduate and Postgraduate Programmes in Electronic Engineering

Module EEEM034; 15 credits

Speaker & Speech Recognition

Level HEM Examination

Time allowed: Two hours

Semester 2, 2010

Read these instructions!

Answer **three questions** out of four.
If you do more, your best three will count.

All questions carry equal credit.

Where appropriate the mark carried by an individual part of a question is indicated in square brackets [ ].

Additional materials: None

Answer three questions

1. (a) Coarticulation is a well-known phenomenon in speech production.

   i. What properties and requirements of the human speech production system cause coarticulation? Identify at least three distinct factors.

   ii. When and under what circumstances does coarticulation occur? Give examples of at least two instances.

   iii. What are the main effects of coarticulation that can be observed? Describe at least two effects observed in speech. **[20%]**

   (b) In fricatives and stop consonants, the frequency of the first main formant is associated with the resonances of the front cavity, i.e., the part inside the mouth from the constriction to the lips. Two allophones of the velar plosive phoneme /k/ occur in different vowel contexts: $[\mathrm{k_i}]$ as in "keep", and $[\mathrm{k_a}]$ as in "card".

   i. What are allophones?

   ii. Why do they occur here?

   iii. Calculate the frequency of the first acoustic resonance, given that the front cavity lengths are 3.0 cm and 5.7 cm respectively, and that this part of the oral cavity may be treated as a uniform tube, closed at one end.

   [Hint: assume speed of sound $c = 340\,\mathrm{ms}^{-1}$.] **[20%]**

   (c) A system for recognizing spoken numbers 0–99 uses a pronunciation dictionary to combine phone models into word models. Consider the phone transcriptions of "eight-two" /ɛitu/ and "eighty" /ɛiti/ together with the table of articulatory features in Table 1.

   i. Describe what coarticulation effects would occur during the initial diphthong /ɛi/ and how this would appear in a spectrogram.

   ii. Identify another form of coarticulation that would produce differences between the two phonetic realisations, other than the differences during the final phone. Explain your reasoning to justify your answer.

   iii. For your answer to part (ii) above, describe the expected articulatory change and any corresponding changes in the acoustic signal. **[30%]**

   (d) Discuss **two** of the following methods for re-estimating model parameters using limited amounts of training data. Marks will be given for clear explanations, illustrative examples, relevant equations where all symbols have been defined, and derivations.

   i. Parameter tying based on articulatory features

   ii. Good-Turing discounting

   iii. Witten-Bell discounting

   iv. Maximum a posteriori (MAP) adaptation

   v. Maximum-likelihood linear regression (MLLR) adaptation **[30%]**

Question 1 continues on next page

TABLE FOR QUESTION 1 CONTINUED FROM PREVIOUS PAGE

Phoneme	Voicing	Manner	Place	Lip rounding
/ɛ/	Voiced	Vowel	Front, open-mid	Unrounded
/i/	Voiced	Vowel	Front, close	Unrounded
/t/	Unvoiced	Stop	Alveolar obstruent	–
/u/	Voiced	Vowel	Back, close	Rounded

Table 1: Articulatory features specified for each of the listed phonemes, or unspecified (–).

2. (a) A low-cost coach operator, called Guildford Tube, would like you to design the voice user interface (VUI) for its telephone ticket sales line. Your task concerns the language model (LM) used in the dialogue system.

To create a booking, the system asks the customer a series of questions to gather the information: \$Booking = { \$Origin, \$Destination, \$OutTime, \$Return, \$Fare }, as set out in the grammar below, where [·] denotes an option, | alternatives and SIL silence.

```
$Origin = [SIL] $Place [SIL]
$Destination = [SIL] $Place [SIL]
$OutTime = $Month $Day $Hour
$Return = ( SINGLE | ONE-WAY ) | ( RETURN $RtnTime )
$RtnTime = $Month $Day $Hour
$Fare = CHILD | ADULT | CONCESSION

$Place = GUILDFORD | LONDON | READING | PORTSMOUTH | WOKING
$Month = JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC
$Day = FIRST | SECOND | THIRD | ... | THIRTIETH | THIRTY-FIRST
$Hour = ( 1 | 2 | ... | 12 ) ( AM | PM )
```

i. Draw the grammar diagram for \$Origin including all the details given above.
ii. Draw the grammar diagram for \$OutTime including all the above details.

[25%]

(b) Initially, the default settings in the LM are for each alternative to have equal probability at any branch, e.g., $P(\text{AM}) = P(\text{PM}) = 1/2$ within the \$Hour lattice.

i. Hence, what would be the default probability value given for $P(\$\text{Month} = \text{JUN})$?
ii. Assuming independence of \$Month, \$Day and \$Hour within \$OutTime, what would be the default joint probability for $P(\$\text{OutTime} = \text{FEB THIRTY-FIRST 1 AM})$?
iii. Suggest one simple improvement that could be made to reduce the number of booking errors.

[25%]

(c) Counts from records of the journeys made show significant seasonal variation, as in Table 2. Similarly, statistics of the difference between purchase date and travel date show a steep decline truncated after three months, which is the maximum period allowed for advance bookings (Table 3).

i. Based on these data, what modification would you recommend making first to the LM probabilities for \$Month? Justify your recommendation and briefly explain how you would implement this (in one or two sentences).
ii. Assuming the seasonal variation and advance booking duration are independent, write an expression to calculate the revised LM probability combining these two factors.
iii. From today's date, calculate the values for $P(\$\text{Month} = \text{JUN})$, $P(\$\text{Month} = \text{AUG})$ and $P(\$\text{Month} = \text{OCT})$.

[25%]

QUESTION 2 CONTINUES ON NEXT PAGE

QUESTION 2 CONTINUED FROM PREVIOUS PAGE

(d) The concept of an N-gram LM is to used for updating the probabilities relating to \$Place, based on the popularity of each route as in Table 4.

i. Once the \$Origin has been recognized, explain how the concept could be used to provide a conditional probability for recognizing the \$Destination, such as $P(\text{\$Destination} = \text{LONDON}|\text{\$Origin} = \text{GUILDFORD})$

ii. Accordingly, use the bigram, unigram and zero-gram probabilities, $N = \{0, 1, 2\}$, to estimate the prior probability for this journey, $P(\text{\$Origin} = \text{GUILDFORD}, \text{\$Destination} = \text{LONDON})$.

[25%]

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
3	3	6	9	9	12	13	15	12	6	3	9

Table 2: Counts of monthly bookings, shown as percentages.

0	1	2	3	4+
55	25	15	5	0

Table 3: Counts of time between booking and travel, shown as percentages.

	To					
From	GF	LN	RD	PM	WK	Total
GF	0	140	10	15	35	200
LN	140	0	80	50	130	400
RD	10	80	0	5	5	100
PM	15	50	5	0	30	100
WK	35	130	5	30	0	200
Total	200	400	100	100	200	1000

Table 4: Proportion of journeys per thousand travelled between Guildford (GF), London (LN), Reading (RD), Portsmouth (PM) and Woking (WK).

3. (a) During development of an HMM-based recognition system, what is the purpose of training? **[5%]**

(b) Given the initial-state probabilities π and the state-transition probabilities A below for a prototype HMM (where ^{T} denotes the vector transpose), draw the model topology:

- show emitting states as circles containing the state number,
- denote start and end null nodes as smaller filled circles,
- add arrows to indicate each permissible transition.

$$\pi^T = \begin{bmatrix} 0.75 \\ 0.25 \\ 0 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.6 & 0.3 & 0.1 & 0 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0 & 0 & 0.7 & 0.2 \\ 0 & 0 & 0 & 0.8 \end{bmatrix}, \quad B = \begin{array}{c} \begin{matrix} \spadesuit & \heartsuit & \clubsuit & \diamondsuit \end{matrix} \\ \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.0 & 0.0 & 0.8 & 0.2 \\ 0.0 & 0.1 & 0.3 & 0.6 \\ 0.5 & 0.0 & 0.2 & 0.3 \end{bmatrix} \end{array}$$

[15%]

(c) The best state sequences $X = \{1, 1, 1, 2, 3, 3, 3, 4\}$ and $X = \{1, 1, 2, 2, 4, 4, 4\}$ were found by the Viterbi algorithm using this HMM prototype for the observation sequences $\mathcal{O} = \{\spadesuit, \heartsuit, \heartsuit, \clubsuit, \diamondsuit, \clubsuit, \heartsuit, \spadesuit\}$ and $\mathcal{O} = \{\clubsuit, \diamondsuit, \clubsuit, \clubsuit, \spadesuit, \clubsuit, \diamondsuit\}$, respectively, including a final transition to the (non-emitting) exit node.

i. Sketch the trellis diagram for each of these two training examples, showing the path of the best state sequence.

ii. According to Viterbi training, calculate the new initial-state probabilities π^{Vit}.

iii. Similarly, calculate the new state-transition probabilities A^{Vit}.

iv. Calculate the new output probabilities B^{Vit}. **[45%]**

(d) i. From your values of π^{Vit} and A^{Vit}, sketch the topology of the newly re-estimated model.

ii. Comment on any changes from the original prototype HMM. **[15%]**

(e) i. What is the main difference between Baum-Welch training and Viterbi training?

ii. With reference to your A^{Vit} matrix, re-estimated by Viterbi training in part (c), describe qualitatively how you would expect the values re-estimated by Baum-Welch training to differ. **[20%]**

4. (a) State the Bayes minimum error decision rule.

[15%]

(b) Explain the terms *a priori class probability* and *a posteriori class probability* in statistical decision making.

[10%]

(c) Consider a biometric personal identity verification system which takes a biometric measurement on a client and evaluates its consistency with the claimed client identity. The outcome of the consistency test is expressed in terms of a one-dimensional score x. It is known that the client score distribution is given as

$$p(x|\omega_1) = \begin{cases} -ax + 1 & 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

whereas the class of impostors has a distribution

$$p(x|\omega_2) = \begin{cases} ax - 0.5 & 1 \leq x \leq 3 \\ 0 & \text{elsewhere} \end{cases}$$

Suppose that the a priori probability of a claim being fraudulent is 0.2.

i. Find the value of parameter a so that $p(x|\omega_1)$ and $p(x|\omega_2)$ satisfy the axiomatic properties of probability density functions.

[10%]

ii. Sketch the class probability density functions.

[10%]

iii. What will be the aposteriori probability of an identity claim being false when the biometric score $x = 2.5$?

[10%]

iv. Find the minimum error decision threshold.

[15%]

v. What will be the false acceptance rate of imposters with the decision threshold set to the value found in (c)iv?

[15%]

vi. How should you set the threshold so as to reduce the false acceptance rate to 1%?

[15%]

Examiners: Dr. P.J.B. Jackson, Prof. J. Kittler
External examiners: Prof. L. Cuthbert, Prof. P. Rees