EEEM034: Solutions Spring 2009/10, Semester 2

1.(a)(i) Any three factors from the following:

- the need to conserve energy and optimise effectiveness of speech output for a given amount of effort (physical and/or cognitive)
- articulators have stiffness, damping and inertia
- muscles have finite strength and power
- articulators move slowly and continuously (c.f., speed of sound)
- sparse articulatory targets leave some parts free to vary
- the organisation of speech as a sequence of actions causes blending, overlap and timing effects
- articulators are interconnected and have limited freedom to move
- articulatory-acoustic mapping involves a many-to-many relationship
- not all parts of an utterance are equally important for the communicative intent

8

(ii) Any two examples along these lines:

- as one phoneme transits to the next within an utterance (especially where a large movement is required)
- when some part of the articulatory apparatus is not fully constrained in the current phone yet has a specification that derives from neighbouring phones, a.k.a. the influence of phonetic context or context-sensitivity
- when part of an utterance is de-emphasised, as in fluent speech
- carry-over or anticipatory coarticulation

6

1/2

1(a)(iii) Any two of the following:

- slow continuous movement of articulators between targets
- target undershoot
- timing adjustments, as with feature spreading 6

(b)(i) Distinct variants of a phoneme that are pronounced differently yet perceived as belonging to the same phonetic category for a given language. 5

(ii) The vowel context moves the tongue forward (for [i]) or backward (for [a]) so that the tongue dorsum touches the palate at a different location 5

(iii) $F_1 = \frac{c}{4L}$

$F_1^{ki} = \frac{340}{4 \times 0.03} = 2833\ Hz$

$F_1^{ka} = \frac{340}{4 \times 0.057} = 1491\ Hz$

10

(c)(i) As the tongue and jaw moved continuously from the front-mid configuration to the front-close one, F_1 would decrease (from 700Hz to half) and F_2 would increase (from 1800Hz to 2500Hz). The acoustic transition of the formants would be smooth and continuous. 10

(ii) The change to the final phoneme has its strongest effect on the adjacent phoneme /t/, which has unspecified lip rounding, whereas the lip rounding is absent for /i/ and required for /u/. Thus, we would see wide lips for 80 versus lips narrowing into the rounded position during 8-2.

4

JK

1(c)(ii) (cont.)

Acceptable alternative answers include:

- the effect on /t/ of the change in vowel place from front to back
- the effect on the realisation of /i/ given the change in tongue body configuration in the final phone
- the prosodic changes, to syllable stress and duration for example, as a result of the word structure differences which could produce more intensity on /t/ in 8·2, increase its duration or that of the preceeding vowel, whereas 80 could ~~be~~ contain a flap rather than a stop

10

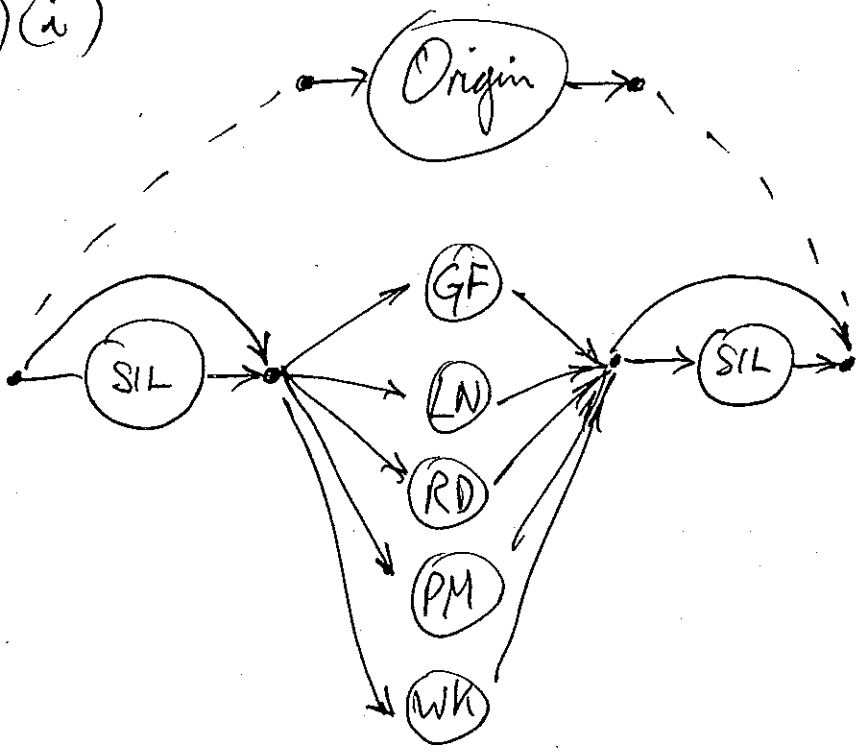
(iii) The acoustic effect of moving the lips from spread to rounded is to reduce formant frequencies as it lengthens the effective length of the vocal tract.
Also, see related points in (c)(i) and (c)(ii) above.

10

(d) Book work — see notes for details.
Marks (up to 15) given for explanations, examples, equations and working.

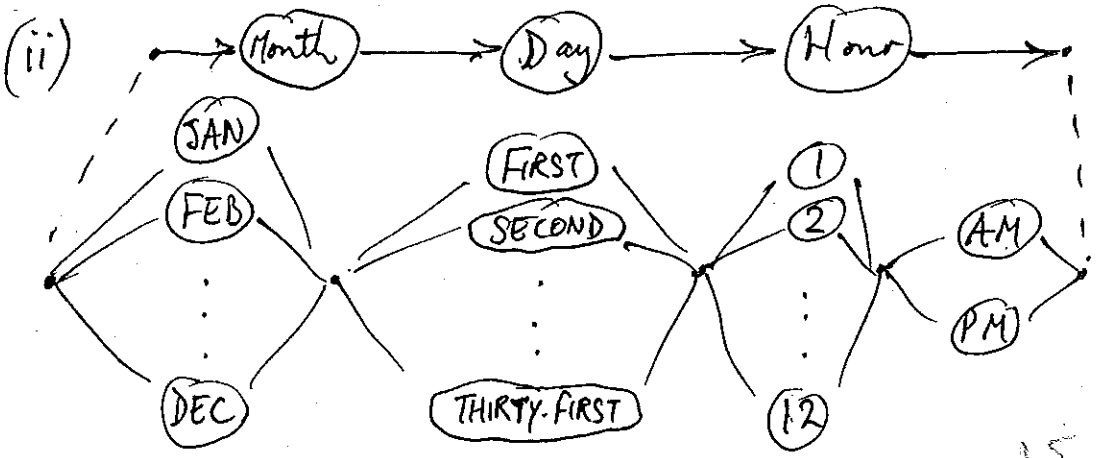
30

2.(a)(i)



10

(ii)



15

(b)(i) $\frac{1}{12}$

6

(ii) $P(\text{OutTime}) = P(\text{Month})P(\text{Day})P(\text{Hour})$

$= \frac{1}{12} \times \frac{1}{31} \times \frac{1}{12} \times \frac{1}{2} = \frac{1}{8928} = 1.12 \times 10^{-4}$

13

(iii)
- To set the probability of illegal dates to zero
- To set the probability to zero when no service is running
- To weight the values according to their prior probability
- To use N-gram conditional probabilities of next part given the previous ones

6

2(c) (i) First, recommend adjusting P(Month) prior probability according to time between booking and travel, because it has a strong effect on likelihoods, eliminates erroneous bookings and reduces computation. Using the current month (e.g., from the OS), calculate the difference as a number of months from the enumerated list and use this to access the values stored in a look-up table such that:

~~$P^*(\text{Month}) = P(\text{Month} \mid \text{Today})$~~

~~$= P^I(\text{Month})\, P(\text{Duration} \mid \text{Today})$~~

where $P^I(\text{Month})$ is the probability value for the current month irrespective of the current date, based on annual sales statistics, and P(Duration) from the booking durations.

$$P(\text{Month} \mid \text{Today}) = P(\text{Month}, \text{Duration})$$
$$= P^I(\text{Month})\, P(\text{Duration})$$

10

(ii) The seasonal probability $P^S(\text{Month})$ can be used to replace the month prior that was independent of duration:

$$P(\text{Month} \mid \text{Today}) = P^S(\text{Month})\, P(\text{Duration})$$

6

(iii) Assuming we're in June:

$P^S(\text{Month})$		$P(\text{Duration})$		$P(\text{M} \mid \text{Today})$
0.12	×	0.55	=	0.066
0.15	×	0.15	=	0.0225
0.06	×	0	=	0

JK 9

2(d)(i) With an N-gram, we make the approximation

$$P(w_1, w_2, \ldots w_M) = P(w_1)P(w_2|w_1)\ldots P(w_M|w_1, w_2, \ldots w_{M-1})$$

$$\approx P(w_1)\,P(w_2|w_1)\ldots P(w_n|w_{n-N+1}, \ldots, w_{n-1})\ldots$$

or alternatively $P(w_M|w_{M-N+1}, \ldots, w_{M-1})$

$$\approx \prod_{n=1}^{M} P(w_n|w_{n-N+1}, \ldots, w_{n-1})$$

So, here we can use bigram ($N=2$) probabilities of $P(\text{Destination}|\text{Origin})$ to obtain the joint probability of the sequence or combination ~~$P(\text{Destination, Or}$~~

$$P(\text{Origin, Destination}) = P(\text{Origin})\,P(\text{Destination}|\text{Origin})$$

13

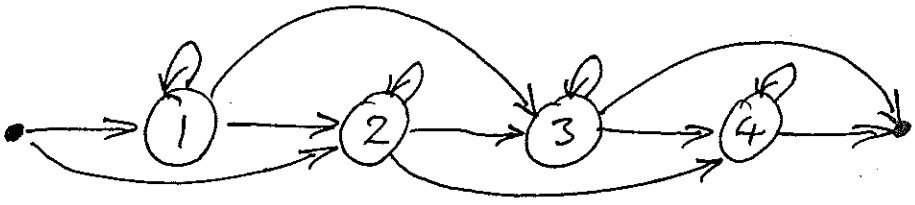
(ii)

N	$P(GF)$ $P(\text{Origin})$		$P(LN/GF)$ $P(\text{Destination}/\text{Origin})$		$P(GF,LN)$ $P(\text{Origin, Destination})$
2	0·2	×	0·7	=	0·14
1	0·2	×	0·4	=	0·08
0	0·2	×	0·2	=	0·04

12

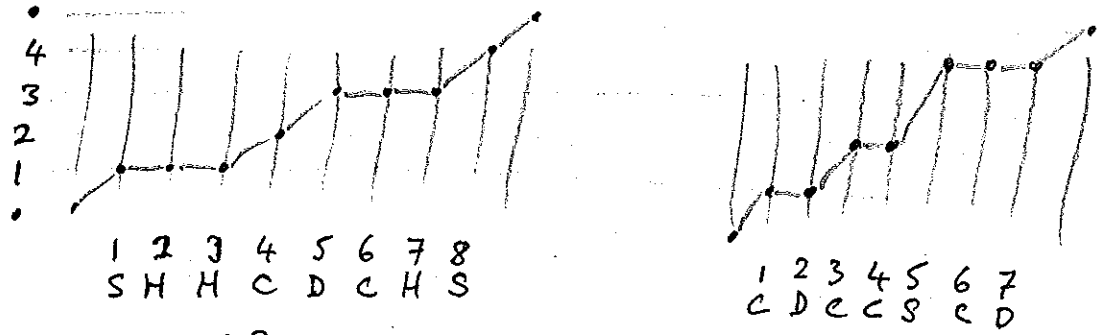
3(a) To determine the optimal parameter settings for the models (in a ML sense), which act as templates of the recorded speech patterns. 5

(b)



15

(c)(i)



8

(ii) $\pi^{Vit} = \left[\frac{2}{2} \quad \frac{0}{2} \quad \frac{0}{2} \quad \frac{0}{2}\right] = [1 \; 0 \; 0 \; 0]$

5

(iii) $a_{1j} = \left[\frac{3}{5} \quad \frac{2}{5} \quad \frac{0}{5} \quad \frac{0}{5}\right] = [.60 \; .40 \; 0 \; 0]$

$a_{2j} = \left[\frac{0}{3} \quad \frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}\right] = [0 \; .33 \; .33 \; .33]$

$a_{3j} = \left[\frac{0}{3} \quad \frac{0}{3} \quad \frac{2}{3} \quad \frac{1}{3}\right] = [0 \; 0 \; .67 \; .33]$

$a_{4j} = \left[\frac{0}{4} \quad \frac{0}{4} \quad \frac{0}{4} \quad \frac{2}{4}\right] = [0 \; 0 \; 0 \; .50]$

$$\Rightarrow A^{vit} = \begin{bmatrix} .60 & .40 & 0 & 0 \\ 0 & .33 & .33 & .33 \\ 0 & 0 & .67 & .33 \\ 0 & 0 & 0 & .50 \end{bmatrix}$$

16

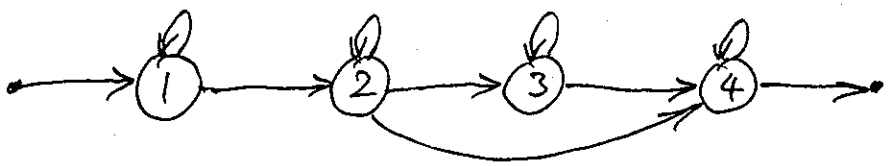
JK

3(c)(iv)

$$B^{Vit} = \begin{bmatrix} S & H & C & D \\ 1/5 & 2/5 & 1/5 & 1/5 \\ 0/3 & 0/3 & 3/3 & 0/3 \\ 0/3 & 1/3 & 1/3 & 1/3 \\ 2/4 & 0/4 & 1/4 & 1/4 \end{bmatrix} = \begin{bmatrix} .20 & .40 & .20 & .20 \\ 0 & 0 & 1 & 0 \\ 0 & .33 & .33 & .33 \\ .50 & 0 & .25 & .25 \end{bmatrix}$$

16

(d)(i)



9

(ii) The topology is much simpler as the transitions that did not occur in the two training examples have been deleted. There remains only one skip (from 2 to 4) and otherwise every state must be occupied at least once in a strict left-right manner. If no additional training utterances are included, the initial state and state transitions are severely limited which could lead to a poor match with alternative test utterances.

6

JK

3(e)(i) Viterbi training only considers the best path. This makes a hard assignment of observations to states at each time frame. B-W uses the full forward and backward probs to give soft allocation via the occupation likelihood $\gamma_t(i)$ and transition likelihood $\xi_t(i,j)$. 10

(ii) By including other potential paths through the trellis, some weighting is given to alternative transitions and observation assignments. Thus, the training by BW would result in fewer zero probs, making a less rigid constraint on the topology and the ability of the model to generalise. 10

JLK

Module: EEM.~~asp~~ ssr ~~ADVANCED SIGNAL PROCESSING~~ Speaker & Speech Recognition
Year: 2009/2010
Examiner: J Kittler
Special Requirements: None
Exam shared with: COMM027 Biometrics
SOLUTION No.: 4

(a) *Bayes minimum error rule:* This is a general decision rule which assumes that the costs of classification errors are zero/one costs, i.e. zero for a correct decision and one for any incorrect decision. The rule assigns pattern vector $\mathbf{x}$ to class ω_j if

$$P(\omega_j)p(\mathbf{x}|\omega_j) = \max_{i=1}^{m} P(\omega_i)p(\mathbf{x}|\omega_i)$$

or if

$$P(\omega_j|\mathbf{x}) = \max_{i=1}^{m} P(\omega_i|\mathbf{x})$$

where $P(\omega_i)$ and $P(\omega_i|\mathbf{x})$ are the i^{th} class a priori and aposteriori probabilities respectively an $p(\mathbf{x}|\omega_i)$ is the i^{th} class probability density function.

[15%]

(b) Given a pattern, the a priori class probability specifies the probability of occurrence of a particular class before any observation is made on the pattern. By making an observation, i.e. taking measurements on the pattern we gain information which is reflected in a change of the class probability. The class probability after the experiment is referred as the aposteriori class probability.

[10%]

(c) i. The integral of the density function must be equal to one. Hence

$$\int_0^2 (-ax + 1)dx = [-a\frac{x^2}{2} + x]_0^2 = 1$$

from which we get

$$a = 0.5$$

[10%]

ii.

PG

[10%]

iii. When $x = 2.5$ the aposteriori probability of the identity claim being false will be

$$P(\omega_2|x) = \frac{p(x|\omega_2)P(\omega_2)}{p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2)} = 1$$

since $p(2.5|\omega_1) = 0$.

[10%]

iv. The minimum error decision threshold is given by

$$p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2)$$

Substituting we get

$$(-0.5x + 1) \times 0.8 = (0.5x - 0.5) \times 0.2$$

which gives

$$x = 1.8$$

[15%]

v. With the above optimal threshold the false acceptance rate of imposters is given by

$$P(\omega_2) \int_1^{1.8} (0.5x - 0.5)dx = 0.2[0.25x^2 - 0.5x]_1^{1.8} = 0.032.$$

[15%]

vi. To reduce the false acceptance rate to 1% the threshold would have to be set to value b for which

$$0.2[0.25x^2 - 0.5x]_1^b = 0.01$$

Solving the quadratic equation we find an appropriate root

$$b = 1.447$$

[15%]