1(a) Marks awarded for relevant diagrams, equations and examples; as in the following lecture slides:

- C·4, G·15
- C·5, G·15
- G·18
- C·5, G·16, E·12, ~~E·16/17~~ E·15-17, G·23, H·8/11, etc.

[3 marks per point, [20]]

(b)

	/m/	/æ/	/s/
(i)	nasal, labial	vowel, open-front	fricative, alveolar
(ii)	Larynx	Larynx	turbulent jet
(iii)	mouth closed at lips velum lowered	mouth open velum closed (optimal)	tongue tip near palate velum closed
~~(iii)~~	nasal cavity	oral cavity	front cavity
(iv)	murmur	clear voicing	high frequency noise

[3 pts/point, upto 60]

(c) For MFCCs, as in slides H·14/15. ~~H·12~~

For PLPs, " " - H·8

- Pre-emphasis
- Filter bank
- Amplitude compression
- Spectral smoothing

} as in H·12

[20]

2. (a) (i) for decoding/recognition by finding the best alignment with a model and scoring in terms of max. cumulative likelihood, $\delta_t(i)$ [5]

(ii) It is a DP method which computes the $\delta_t(i)$ values from those at the previous frame $\delta_{t-1}(i)$ for $i \in \{1..N\}$. [5]

(b) $$p(O, X|\lambda) = p(O|X,\lambda)\, P(X|\lambda)$$
based on the model's Markov assumptions [10]

(c) (i) $$X^* = \arg\max_X p(o_1 .. o_T, x_1 .. x_T | \lambda)$$

$$\delta_{t+1}(j) = \max_{x_1^{t+1}} \left[p(O_1^{t+1}, x_1^t, x_{t+1} = j) \right]$$

$$= \max_{x_1^{t+1}} \left[p(O_1^t, x_1^{t+1}, x_t = i)\, p(O_{t+1}, x_{t+1} = j | x_t = i) \right]$$
$$\left[p(x_{t+1} = j | x_t = i)\, p(o_{t+1} | x_{t+1} = j) \right]$$

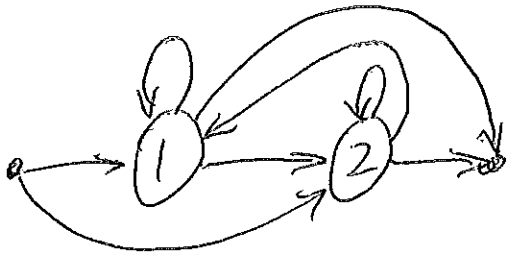
$$\delta_{t+1}(j) = \max_i \left[\max_{x_1^t} \left[p(O_1^t, x_1^{t+1}, x_t = i) \right] P(x_{t+1} = j | x_t = i)\, p(o_{t+1} | x_{t+1} = j) \right]$$

$$= \max_i \left[\delta_t(i)\, a_{ij} \right] b_j(o_{t+1})$$

For the HMM, $p(O_{t+1}, x_{t+1} = j | O_1^t, x_1^{t+1}, x_t = i) \approx p(O_{t+1}, x_{t+1} = j | x_t = i)$
$= p(x_{t+1} = j | x_t = i)\, p(O_{t+1} | x_{t+1} = j)$

2(d)

(i)



[10]

(ii) Ergodic (i.e., fully connected) [5]

(e) (i)

$t=1$: $\delta_1(1) = \pi_1 b_1(o_1) = 0.8 \times \cancel{0.2288}\ 0.223 = 0.1784$

$\delta_1(2) = \pi_2 b_2(o_1) = 0.2 \times \cancel{0.0397}\ 0.040 = 0.0080$

$t=2$: $\delta_2(1) = \max\left[(\delta_1(1)a_{11}), (\delta_1(2)a_{21})\right] b_1(o_2)$

$= \max\left[(0.1784 \times 0.5), (0.0080 \times 0.1)\right] \times 0.042$

≈ 0.00375

$\delta_2(2) = \max\left[(\delta_1(1)a_{12}), (\delta_1(2)a_{22})\right] b_2(o_2)$

$= \max\left[(0.1784 \times 0.4), (0.0080 \times 0.7)\right] \times 0.279$

≈ 0.01991 [25]

(ii) $P(O, X^* | \lambda) = \max\left[(\delta_2(1)\eta_1), (\delta_2(2)\eta_2)\right]$

$= \max\left[(0.00375 \times 0.1), (0.01991 \times 0.2)\right]$

$= 0.0003982 \approx 0.0004$ [10]

(iii) $X^* = \{1, 2\}$ [5]

2(e)(iv)

o_1 is close to μ_1 and o_2 to μ_2

The state transition probabilities of the underlying Markov model only provide a weak preference for staying in a state, compared to the transitions between states.

It is no surprise, therefore, that this short sequence includes the transition, and follows the states set by the ~~o~~ output pdfs, $X^* = \{1, 2\}$.

Marks for relevant notes [5]

3(a) To optimize the model parameters, acting as a template for an utterance. [10]

(b) (i) $$\gamma_t(i) = \frac{P(O_1^t, x_t=i)\,P(O_{t+1}^T \mid x_t=i)}{P(O_1^T)} = \frac{\alpha_t(i)\,\beta_t(i)}{P(O)}$$ [6]

(ii) $$\xi_t(i,j) = \frac{P(O_1^{t-1}, x_{t-1}=i)\,P(x_t=j \mid x_{t-1}=i)\,P(O_t \mid x_t=j)\,P(O_{t+1}^T \mid x_t=j)}{P(O)}$$

$$= \frac{\alpha_{t-1}(i)\,a_{ij}\,b_j(O_t)\,\beta_t(j)}{P(O)}$$ [9]

(c)

$t=1$:

$\alpha_1(1) = \pi_1 b_1(O_1) = 0.8 \times 0.1 = 0.08$

$\alpha_1(2) = \pi_2 b_2(O_1) = 0.2 \times 0.8 = 0.16$

$\alpha_1(3) = \pi_3 b_3(O_1) = 0 \times 0 = 0$

$t=2$:

$\alpha_1(1) = 0.08 \times 0.7 \times 0.0 = 0$

$\alpha_2(2) = ((0.08 \times 0.2) + (0.16 \times 0.6)) \times 0.1 = 0.0112$

$\alpha_2(3) = ((0.08 \times 0.1) + (0.16 \times 0.3)) \times 0.3 = 0.0168$

$P(O|\lambda) = (0.0112 \times 0.1 + 0.0168 \times 0.1) = 0.0028$ [10]

3 (d)

$t=1$: $\gamma_1(1) = \frac{\alpha_1(1)\beta_1(1)}{P(O)} = \frac{1}{0.0028}(0.08 \times 0.005) \approx 0.143$

$\gamma_1(2) = 0.16 \times 0.015 \approx 0.857$

$\gamma_1(3) = 0 \times 0.027 = 0$

$t=2$: $\gamma_2(1) = 0 \times 0$

$\gamma_2(2) = 0.0112 \times 0.1 = 0.4$

$\gamma_2(3) = 0.0168 \times 0.1 = 0.6$ [25]

(e)(i) $\hat{B}(\clubsuit) \approx \begin{pmatrix} \frac{0.143}{0.143} \\ \frac{0.857}{0.857+0.4} \\ \frac{0}{0.6} \end{pmatrix} \approx \begin{pmatrix} 1 \\ 0.682 \\ 0 \end{pmatrix}$ [15]

(ii) There is a large increase for state 1, whose only possible occupation arises at $t=1$.
The value for state 2 reduces slightly, towards 50% for this 2-frame observation sequence.
The third state could not have been occupied to produce this observation type, according to the initial B matrix, and so with occupation likelihood of zero it remains at zero probability for this observation type. [10]

Module: EEEM.034 SPEAKER AND SPEECH RECOGNITION
Year: 2010/2011
Examiner: J Kittler
Special Requirements: None
QUESTION No.:

1. (a) Explain what an authenticator is. List the major authenticator types and give an example for each category.

[15%]

(b) Identify advantages and disadvantages of voice biometrics.

[15%]

(c) Draw a bloc diagram for a speaker recognition system, describing each component and its function.

[15%]

(d) Sketch a receiver operating curve and explain its purpose.

[10%]

(e) A speaker model is characterised by a covariance matrix $\Phi = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$. A speech utterance to be tested against the model has covariance matrix $\Sigma = \Phi + \begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}$. Determine the range of permissible values of a for Σ to remain a covariance matrix.

[25%]

(f) Using the Bhattacharrya distance

$$J = ln \frac{|\frac{1}{2}(\Phi + \Sigma)|}{\sqrt{|\Phi||\Sigma|}}$$

as a matching criterion, determine whether an access claim posed by the speaker verification problem defined by the covariance matrices Φ and Σ given in part 1e) with $a = 2$ will be accepted against a threshold $t = 0.3$.

[20%]

Module: EEEM034 SPEAKER AND SPEECH RECOGNITION
Year: 2010/2011
Examiner: J Kittler
Special Requirements: None
SOLUTION No.: 1

1. (a) An authenticator is a device which allows an authorised access to a secure site or service.

   The main types of authenticators are

   - Secret (password, obscure name)
   - Token (active, such as synchronised password generation, or passive, such as smart card password storage)
   - ID (inalterable, such as fingerprint, face, hand, iris, or alterable, such as voice, signature, keystroke)

   [15%]

   (b) Advantages and disadvantages of voice biometrics

   - Advantages (used by humans, user friendly, natural interface, conveying emotion, can be used over telephone, convenient, ubiquitous, inexpensive, provides challenge-response security)
   - Disadvantages (mainly performance, required length of speech utterance, requires controlled environment, open to spoofing, template aging problem, privacy concerns)

   6 examples from the first group and 4 examples from the second group for full mark.

   [15%]

   (c) A bloc diagram for a speaker recognition system has the following components.

   - Microphone
   - Preprocessing stage (signal filtering and conditioning)
   - Silence detection to segment out voice containing data
   - Feature extraction stage (computation of speech descriptors, such as MFCC coefficients)
   - Classifier (matching between input utterance and speaker model)

   [15%]

(d) A receiver operating curve (ROC) plots the relationship between false acceptances and false rejections in a biometric identity verification system. ROC is generated by changing the acceptance threshold and measuring these two performance measures for each threshold setting. An operating point for a biometric verification system is selected from the ROC curve. The curve is normally computed on an independent validation set to avoid an optimistic bias in performance assessment. The system is then tested on an independent test set at the operating point threshold. [10%]

(e) For Σ to be a valid covariance matrix, it must be a positive definite matrix. This condition can easily be checked by, for instance, eigenvalue analysis of the matrix. Its eigenvalues must be non-negative. Thus from eigenvalues of Σ, we can determine an admissible range for a. The eigenvalues can be found by setting

$$\left|\Phi + \begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}\right| = 0$$

which leads to a quadratic equation

$$(3-\lambda)(2-\lambda) - a^2 = 0$$

The solution for the smaller eigenvalue is

$$\lambda = \frac{1}{2}[5 - \sqrt{25 - 4(6-a^2)}]$$

The eigenvalue will be non-negative, if

$$25 - 4(6-a^2) < 25$$

Thus a should be in the interval $a\epsilon[-2.44, 2.44]$ [25%]

(f) The numerator in the Bhattacharrya distance formula is given as

$$|\frac{1}{2}(\Phi + \Sigma)|$$

Substituting $a = 2$ we get

$$\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} = 5$$

The denominator gives

$$\sqrt{|\Phi||\Sigma|} = \sqrt{6 \times 2} = 3.46$$

The log of the ratio of the numerator and denominator gives $J = 0.367$
Thus the claim will be rejected, as the value exceed threshold $t = 0.3$. [20%]