15/20

International Institute of Information Technology, Hyderabad (Deemed to be University)
Subject: Speech Technology (CSE971) (Fall-2019) First Mid Semester Examination
Maximum Time: 1.5 Hours Time: 4:00 PM - 5:30 PM Max. Marks: 20

Roll No: 20171099 Programme: CLD Date of Exam.: 21-09-2019

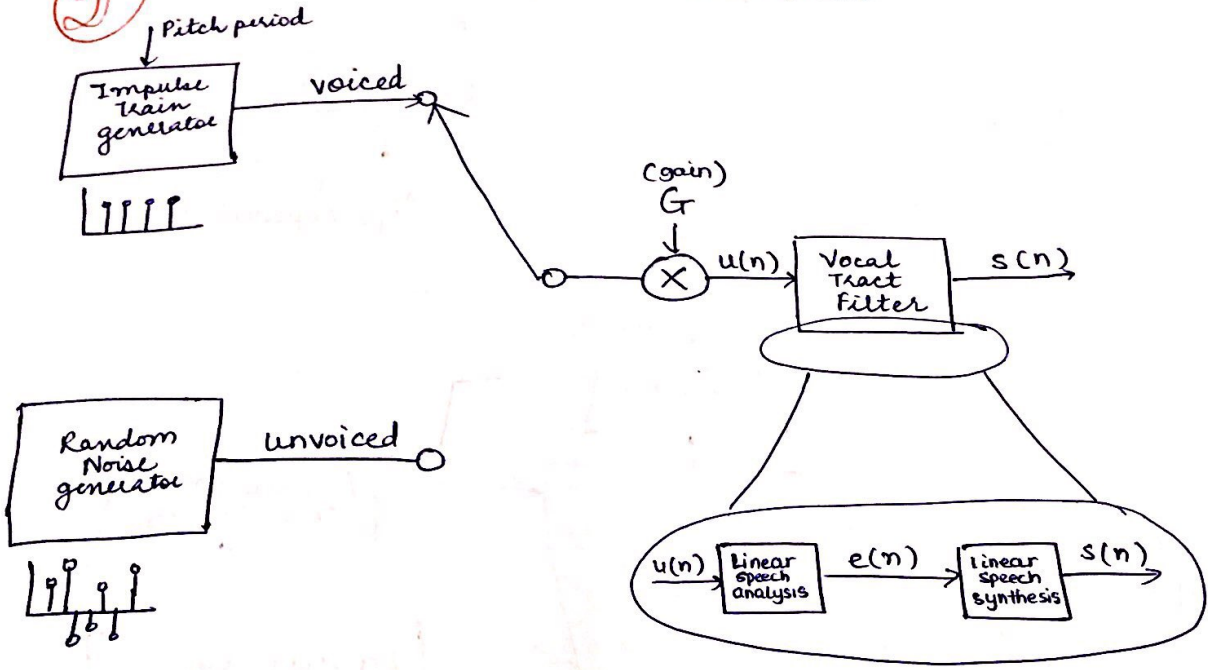
Room No.: 205 Seat No.: C46 Invigilators Signature:

Answer	1	2	3	4	5	6	7	8	9	10	Total
Maximum Marks	2	2	2	2	2	2	2	2	2	2	20
Marks Obtained	2	2	2	2	0	2	1	2	2	0	15/20

Note: Answer all the 10 questions. Answer for each question carries 2 marks. Answers to each of the questions has to be written only in the appropriate pages. All the answers are limited to a maximum of two pages. Last two pages may be used for rough work.

Questions

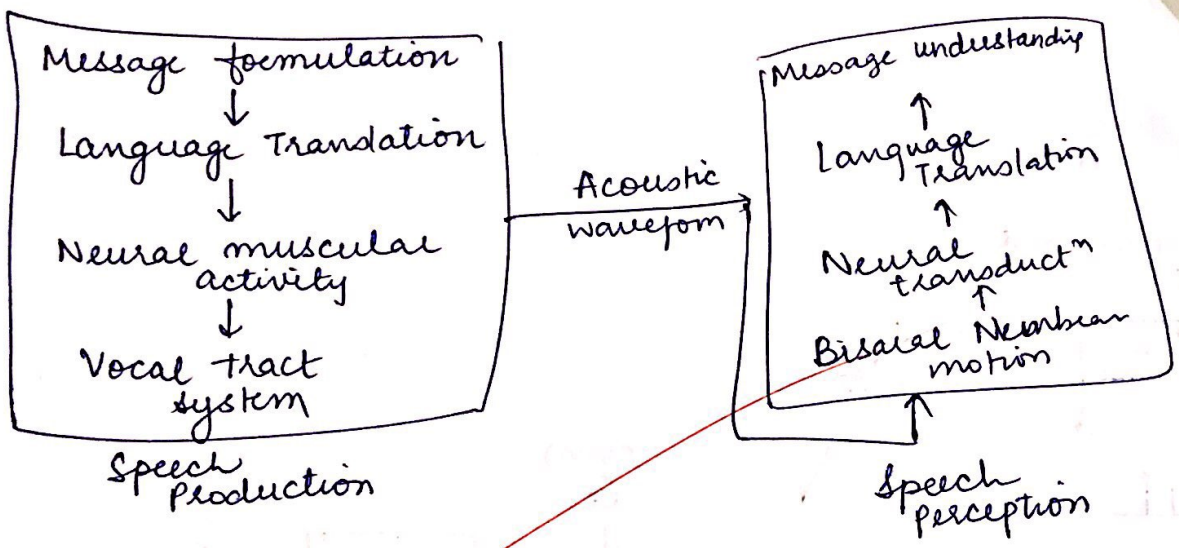
(2) 1. Describe the various components along with their functions of multilingual speech-to-speech translation system.

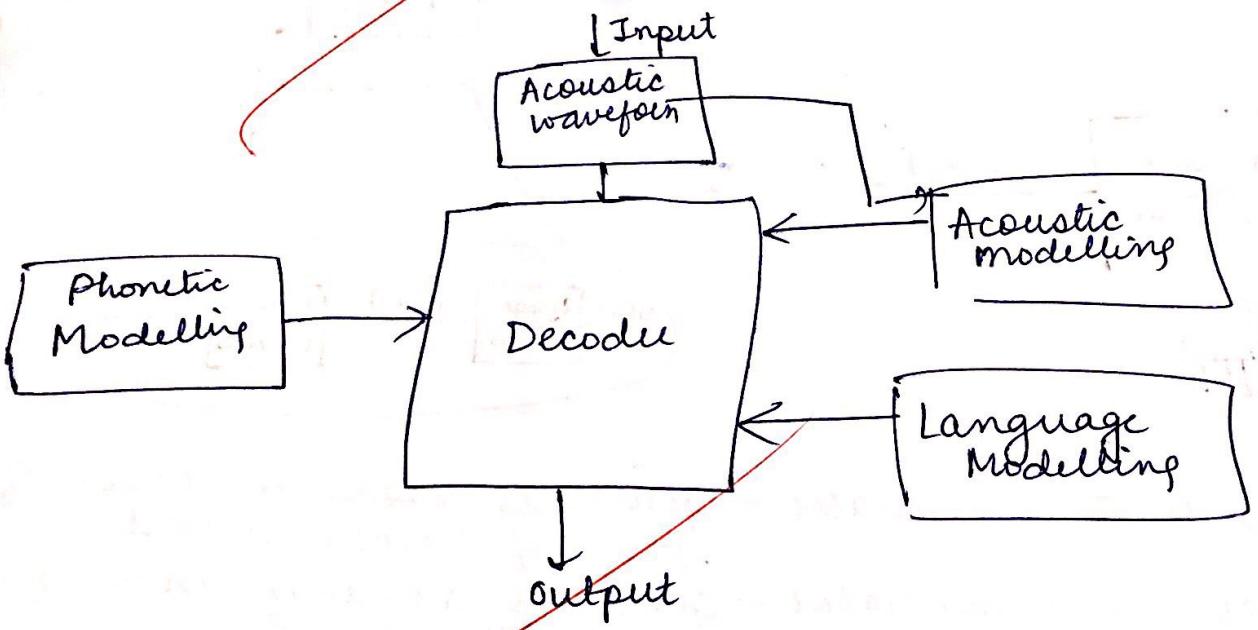


- Impulse Train generator – generates ~~voice~~ a train of pulse if voiced sound
- Random Noise generator – generates random noise for unvoiced sounds
- Gain is provided to intensify signal.
- Vocal tract filter
  - Analysis – takes in input and generates (speech excitation / residual)
  - Synthesis – Residual is taken to produce output

multilingual speech to speech Translator

PTO





Language modelling — Models language using n-grams

Acoustic modelling — depends on accent, ~~space~~ gap b/w syllables, duratⁿ of each syllable etc.

2. Tabulate the different types of sounds in Indian languages based on vowels, place and manner of articulation.

POA	MOA: UnVoiced		Voiced		Vowels	Semi-Vowel	Fricative
	UnAspirated	Aspirated	Unaspirated	Aspirated			
~~Palatal~~ Velar	/ka/ क	/kʰa/ ख	/ga/ ग	/gʰa/ घ	ङ /kna/	–	ह /ha/
Palatal	/cha/ च	/cʰa/ छ	/ja/ ज	/jʰa/ झ	ञ /chna/	/ya/ य	/sha/ श
Alveolar	/Ta/ ट	/Tʰa/ ठ	/Da/ ड	/Dʰa/ ढ	ण /Tna/	/ra/ र	/shna/ ष
Dental	/ta/ त	/tʰa/ थ	/da/ द	/dʰa/ ध	न /na/	/la/ ल	/sa/ स
Bilabial	/pa/ प	/pʰa/ फ	/ba/ ब	/bʰa/ भ	म /ma/	/wa/ व	–

3. Discuss the spotting subword units based approach to speech signal-to-symbol transformation along with its limitations in comparison with segmentation and labeling based approach.

(2)

Spotting subword units based approach

① Detection of anchor points in the continuous speech

② Assigning labels to segments around the anchor points

Segmentation and labelling approach

① Extracting segments from the continuous speech

② Assigning labels to segments using classifier.

Limitation of spotting subword approach —

① region around anchor points may or may not belong to the actual syllable.

Hence this may lead to —

(a) missing information

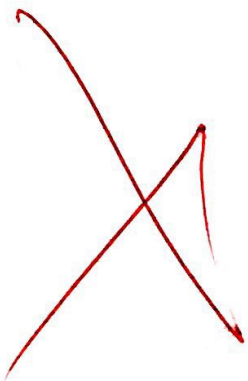
(b) getting wrong information if other syllable is also taken in the region which will result in wrong classification.

4. Discuss the various issues in the development of a phonetic engine.

(2)

Speech Science

1. Choice of subword units
2. Number of subword units
3. Frequency of occurrence of subword units in diff languages
4. Variability in characteristics of subword units
5. Acoustic similarity of subword units
6. Classification model for subword units
7. Representation of subword units
8. Compression of large dimensional patterned vectors
9. Recognition of subword-units in continuous speech.
10. Recognition of subword-units in multiple languages.



5. Describe the typical mathematical model for the simulation of a biological neuron. Further, describe the delta learning law used for the updation of connection strengths between the two neurons in artificial neural network models.

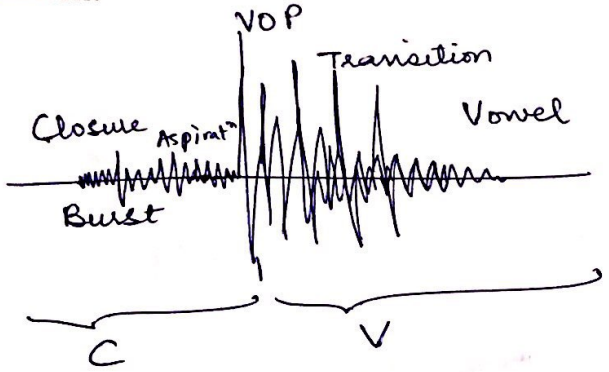
0

~~Ans~~

Delta learning law is a supervised technique where the initial weight is near to zero.

6. Describe the significant events in a syllable unit. Further describe, how these events are attributed to the place and manner of articulation.

(2)



CV unit

(1) Closure – when sound is produced with air restriction

(2) ~~Burst~~ Aspiration – when sound is produced with expulsion of air (more than usual) as in b^h, k^h etc

(3) Burst – when the stricture between the articulators in opened suddenly to produce sound

(4) VOP – Vowel Onset Point — i.e. where vowel begins

(5) Transition – quasi-periodic formants vowels

7. Explain the suitable duration of a Consonant-Vowel (CV) unit which is sufficient to capture the characteristics of a CV unit. Using this information, discuss a method of obtaining fixed dimensional pattern vectors from varying duration CV units.

(1)

25msec 40msec 65 ⇒ 10×39 = 390 ms of V

A suitable duration of a syllable or CV unit is 25-30 ~~m~~seconds because our vocal tract changes every 25-30 ~~m~~second — hence it is better to capture the consistency in one unit and differences in vocal tract in diff units.

A method of ~~fixed~~ obtaining fixed dimensional pattern vectors is to take frames for the total duration — and use overlapping frames — to increase redundancy — with a certain frame shift.
Hence the total no. of frames ~~can~~ —

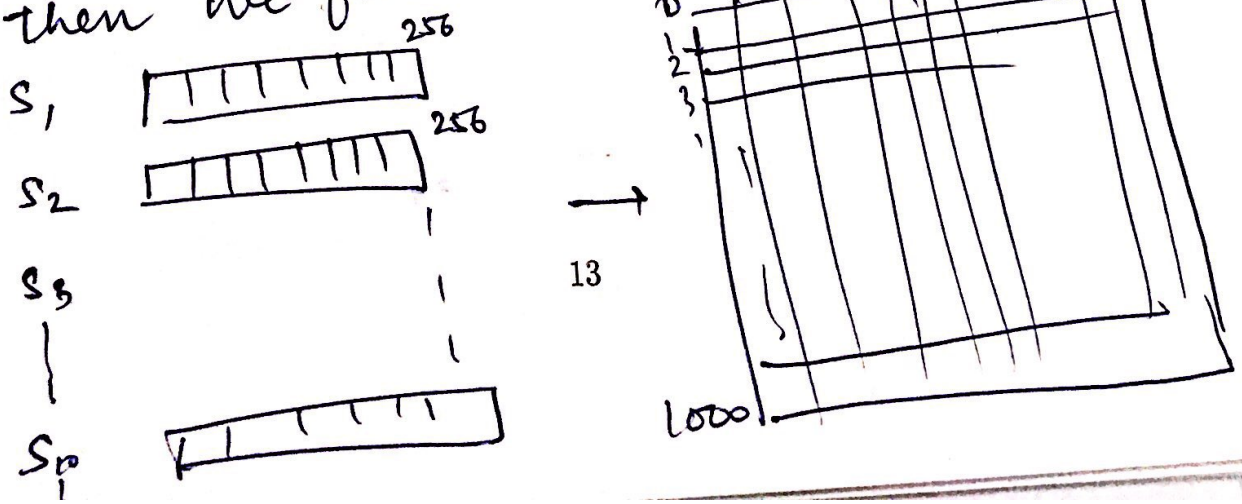
$$\#\text{Frames} = \frac{\text{Total time} - \text{Frame size}}{\text{Frame shift}} + 1$$

So we begin by taking durations in powers of 2.

$2^1 = 2$ $2^2 = 4$ — — — $2^8 = 256$ $2^9 = 512$

So if we have duration of 480sec we will consider 512.

then we find similar samples — such that



then we find the mean ...
for each class.

Sample

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\Sigma = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

And find the probability/likelihood of an unknown phoneme belonging to a class — using Gaussian distribution —

$$P(x/\text{Vowel(class)}) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}[x-\mu]^T \Sigma^{-1} [x-\mu]}$$

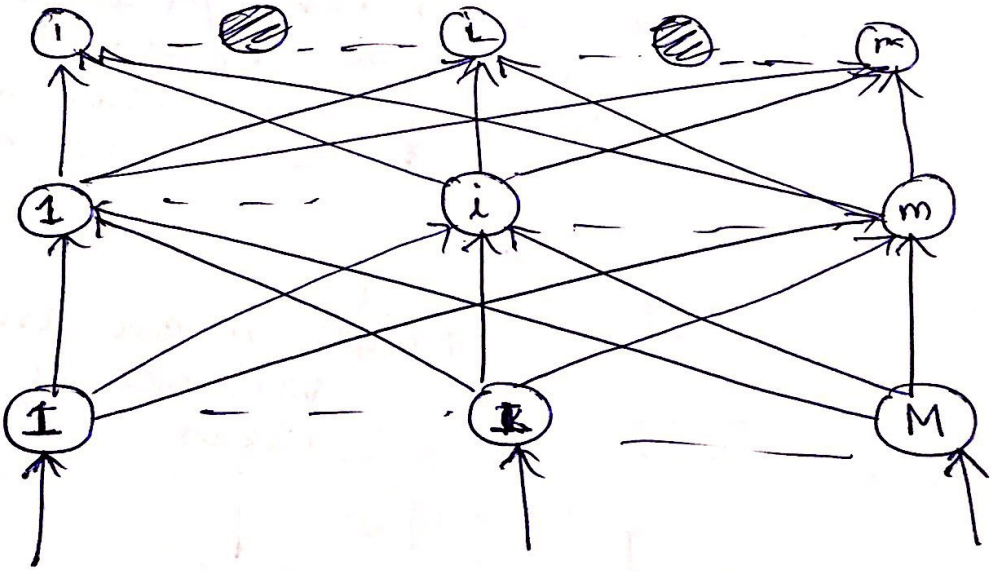
And ~~is~~ whichever

$P(x/V_1)$ is highest — we consider x to belong to V_1.

8. There are five short vowels Indian languages namely /a/,/i/,/u/,/e/,/o/. Describe how the following classification models are used for recognition of these five short vowels along with their advantages and disadvantages. (i) Hidden Markov Models (HMMs), (ii) Multilayer feedforward neural network (MLFFNN) models.

(2)

(ii)



Advantages —

(a) non-linear processing

(b) Pattern classifier

(c) generalised delta learning

~~(d) and~~

Disadvantages —

(a) Slow learning

(i) HMM

3 problems –

(a) Evaluation – likelihood of observed sequence being generated by model

Solution: Forward algorithm

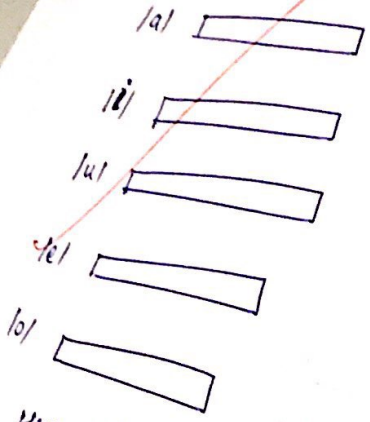
(b) Decoding – finding optimal state sequence

Solution: Viterbi algorithm

(c) Training – estimating parameter values to maximise $P(O/\lambda)$

where O – observation sequence

λ – ~~no~~trained model.



Hence find –

$P(x/a)$

$P(x/i)$

$P(x/u)$

$P(x/e)$

$P(x/o)$

Find mean and variance of each class

Check new samples class by finding class that gives max. probability

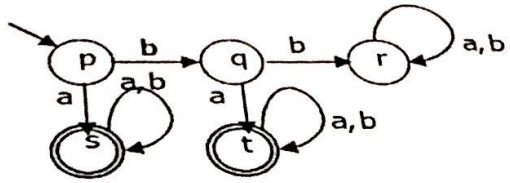
Advantages – computationally efficient, easy to use

Disadvantages – no training method has been known.

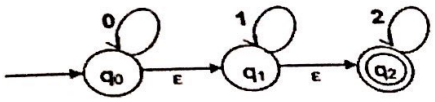
9. What is the significance of FST's in ASR system.

a) Minimize the finite state atomata (FSA) give below.

(2)

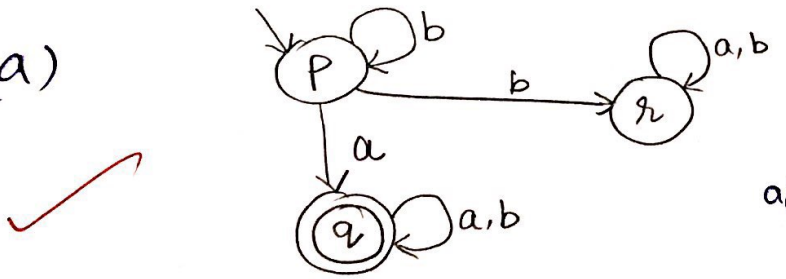


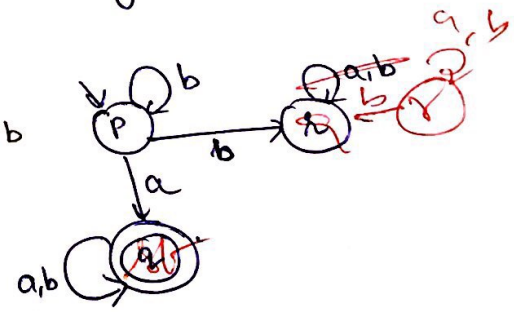
b) Convert the give NFA to Deterministic FSA and mention the final states of the optimized automata.



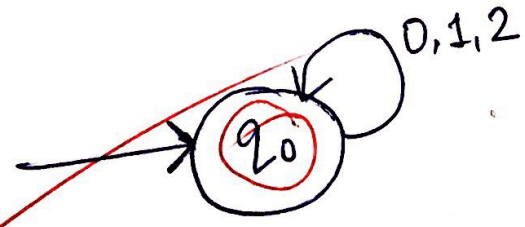
Significance of FST – outputs a new string from the alphabet set from the output set and probabilitys.

(a)





(b)



Final states – q_0

10. Consider HMM model is trained with 5 vowel sounds. Formants F_1, F_2 are considered as observations (2 dimensional vector). After the training their initial probabilities are $\prod_1$, $\prod_2$, $\prod_3$, $\prod_4$, $\prod_5$ as 0.2, 0.3 ,0.25, 0.1, 0.15 respectively. Means of each state are μ_1 = [0.1 0.7] μ_2 = [0.5 0.3] μ_3 = [0.9 0.1] μ_4 = [0.2 0.2] μ_5 = [0.9 0.9]and diagonal covariance of [0.2 0.2] is considered for all the states. Transition probabilities of the network are given in the table below.

	/æ/	/e/	/ɪ/	/ɑ:/	/u:/
/æ/	0.35	0.22	0.1	0.18	0.25
/e/	0.22	0.3	0.2	0.25	0.18
/ɪ/	0.25	0.25	0.25	0.2	0.25
/ɑ:/	0.20	0.4	0.2	0.1	0.1
/u:/	0.1	0.1	0.1	0.3	0.4

Find the top 2 vowel sequences corresponding to the observation sequence (using viterbi algorithm) given below. [Hint: Use Gaussian mixture model to find the output probabilities and Inverse of a diagonal matrix is obtained by replacing each element in the diagonal.]
[1 2] [2 3] [1 1]

$$P(x/aV) = \frac{1}{\sqrt{2\pi|\Sigma|^2}} e^{-\frac{1}{2}[x-\mu]^T \Sigma^{-1} [x-\mu]}$$