University of Central Punjab

Department of Applied Computing Technologies

Fall Term 2025

Title:	Assignment 4	**Subject:**	Speech Processing
Course Books:	Speech and Audio Signal Processing — Ben Gold & Nelson Morgan Introduction to Digital Speech Processing — Lawrence Rabiner & Ronald Schafer Signal Processing of Speech — F.J. Owens		
Instructor:	Muhammad Yasin Nasir	**Section:**	AIDE4313-F25-BS-AI-F23-AE1 & AE2
Assign Date:	15-01-2026	**Submission Date:**	21-01-2026 11:59 pm
Name:		**Reg. No:**	

CLO	CLO Statement	Bloom's Taxonomy Level	Complexity Level	PLO
CLO3	Analyze the structure and functioning of basic speech processing models (HMM, GMM) for isolated word recognition.	(Analyze)	C4	3
CLO4	Evaluate and compare modern approaches to speech recognition and synthesis.	(Evaluate)	C5	4

Total Points: 70 (CLO3=20, CLO4=50)

Instructions:

01. Questions are SELF EXPLANATORY. Understanding the Question is part of Solution.
02. Error in Question will be advantageous to Student.
03. Read the Questions carefully before attempting.
04. Attempt All Questions in a Precise Fashion.
05. Manage Your Time. Therefore, there will be no extension in the deadline.
06. Plagiarism may lead to F grade, so be honest.

Announcement:

01. There will be quiz on 22-01-2026 for AE2 and 22-01-2025 for AE1 during lecture.

Submission Method:

01. Assignment must be submitted individually.
02. **Take print out of the given assignment. Use A4 size paper** (210 × 297 millimeters or 8.27 × 11.69 inches) to print it out. Solve the assignment on the blank A4 pages only.
03. Use any scanner tool to convert it to pdf. Text must have clear visibility after scanning. Blurred pdfs will be rejected.
04. Upload the solved assignment (soft copy) at university portal before the deadline.
05. Submit the assignment (hard from) during the first half hour of the subsequent lecture.

Question 1: Forward HMM and backward HMM **(CLO3 , 20 points)**

Background: An online learning platform wants to model a student's engagement level over time. The engagement level is not directly observable, but the platform can observe user activity. The system is modeled first as a Markov chain and then extended to a Hidden Markov Model (HMM).

Part 1: Markov Chain Model

Hidden States (Engagement Levels):

H = Highly Engaged, M = Moderately Engaged, L = Low Engagement

Initial State Probability: On Day 0, the student is known to be highly engaged.

P(H0) =1, P(M0) =0, P(L0) =0

Today ↓ / Tomorrow →	H	M	L
H	0.6	0.3	0.1
M	0.3	0.4	0.3
L	0.2	0.3	0.5

Part 2: Hidden Markov Model (HMM)

Observations (User Activity):

V = Watched video

Q = Attempted quiz

I = Inactive

State ↓ / Observation →	V	Q	I
H	0.5	0.4	0.1
M	0.4	0.3	0.3
L	0.1	0.2	0.7

Observation Sequence (4 days):Day 0: V, Day 1: Q, Day 2: I,

Let: Q = (H_0, M_1, L_2)

Task A: Use forward HMM to find the probability of the above observation sequence.

Task B: Use backward HMM to find the probability of the above observation sequence.

Task C: Give the comparison of both values from forward HMM and backward HMM.

Question 2: Whisper Model **(CLO4,50 points)**

Whisper is an end-to-end Transformer-based automatic speech recognition model designed to handle multiple speech tasks within a single architecture.

Explain the Whisper model in detail by addressing the following points in a structured answer:

Describe the overall purpose and key capabilities of the Whisper model and explain why it is considered more robust than traditional ASR systems.

1. Explain the encoder–decoder Transformer architecture used in Whisper, highlighting the role of self-attention and cross-attention mechanisms.
2. Discuss the importance of log-Mel spectrograms as input features, including how they are generated and why they are preferred over raw audio.
3. Explain how Whisper performs multilingual speech recognition and speech translation using special control tokens, and how automatic language identification is achieved.
4. Briefly outline the data, computational, and methodological requirements for training or fine-tuning a pretrained Whisper model, and compare full fine-tuning with parameter-efficient fine-tuning approaches such as LoRA.

Finish Line 😊