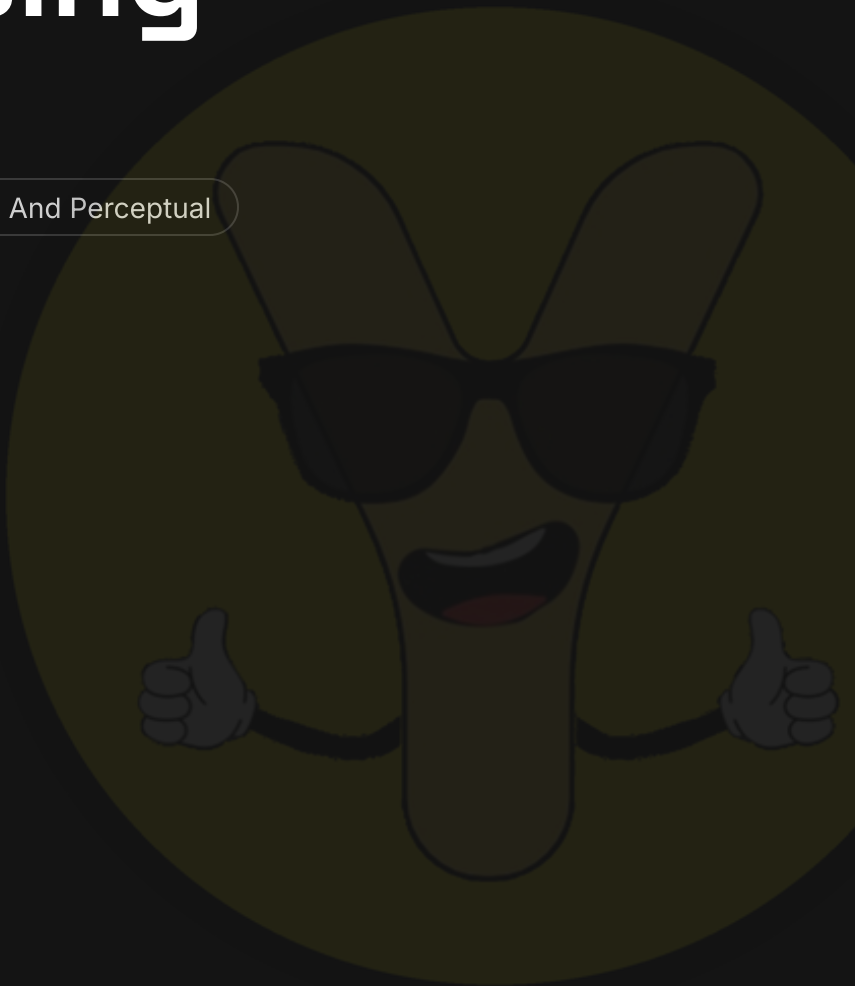
AKTU · 3RD YEAR · SEMESTER 5 · BCAI052

Natural Language Processing

Unit 5 — Speech Analysis and Real World NLP

- Features, Feature Extraction And Pattern Comparison Techniques
- Speech Distortion Measures - Mathematical And Perceptual
- Real World NLP Challenges - Information Extraction and Question Answering
- Dialog Engines
- Optimization, Parallelization and batch processing



DECODED BY

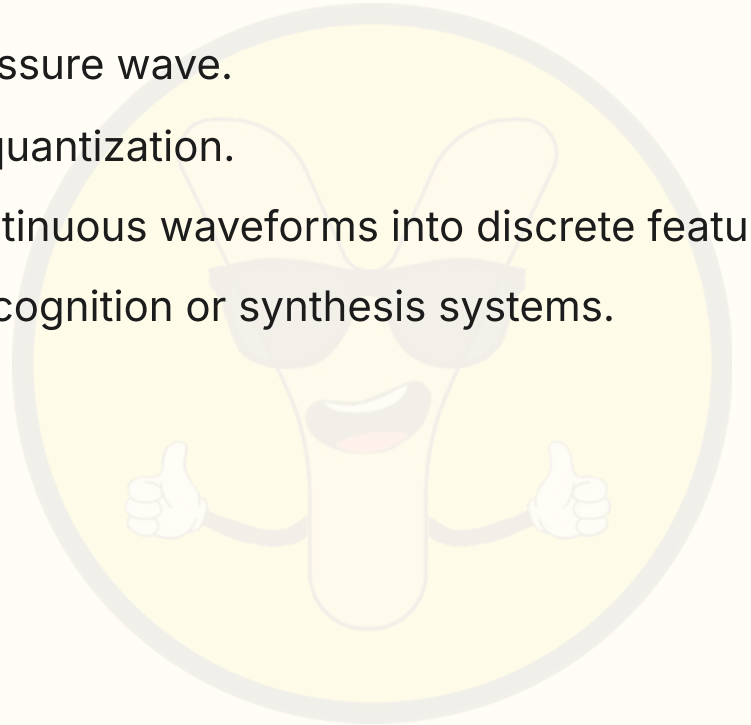
Chirag Jain

Founder, Ycotes · Gold Medalist in Artificial Intelligence, IIT Madras · taught 10,000+ students

Introduction to Speech Analysis

Speech analysis is the process of extracting meaningful information from acoustic signals.

- Speech signal is a time-varying pressure wave.
- Digitization involves sampling and quantization.
- Analysis focuses on converting continuous waveforms into discrete feature vectors.
- Core goal: Represent speech for recognition or synthesis systems.

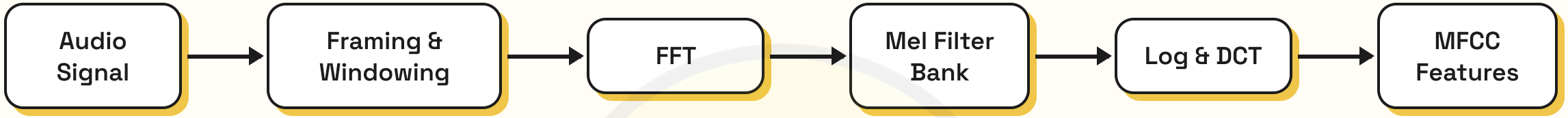


Feature Extraction Techniques

Feature extraction reduces the dimensionality of raw audio while preserving linguistic content.

- Filter Bank Analysis: Mimics human auditory system using band-pass filters.
- Linear Predictive Coding (LPC): Models the vocal tract as an all-pole filter.
- Mel-Frequency Cepstral Coefficients (MFCC): Most common; uses Mel-scale to mimic human pitch perception.
- Steps: Pre-emphasis, Framing, Windowing, FFT, Mel-filter bank, Log, DCT.

MFCC Extraction Pipeline



The standard pipeline for extracting MFCC coefficients from raw speech.

Pattern Comparison Techniques

Comparing two speech patterns requires handling temporal variations.

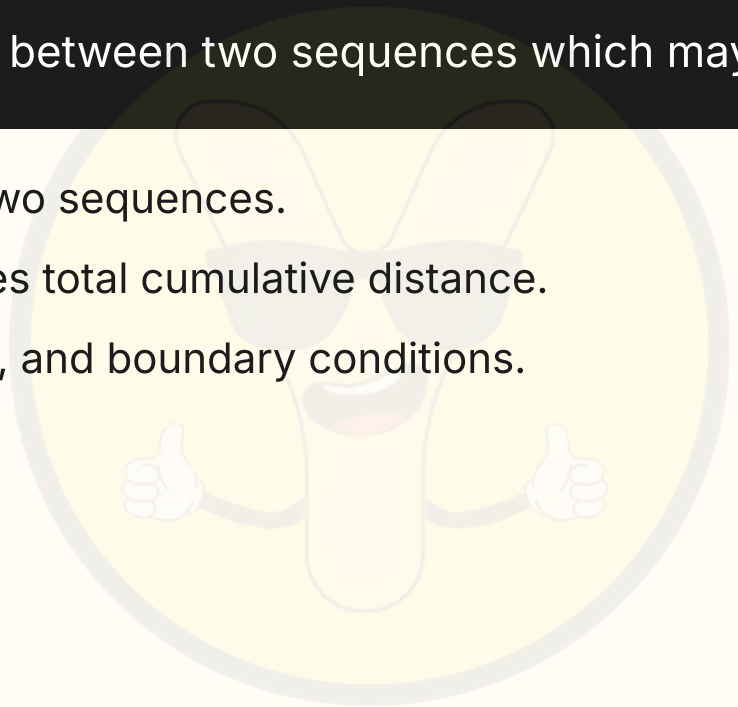
- Dynamic Time Warping (DTW): Aligns two sequences that vary in speed.
- Hidden Markov Models (HMM): Statistical approach for modeling temporal sequences.
- Vector Quantization (VQ): Maps continuous feature vectors to a discrete codebook.
- Goal: Minimize the distance between test and reference templates.

Dynamic Time Warping (DTW)

DTW ALGORITHM

An algorithm to measure similarity between two sequences which may vary in time or speed.

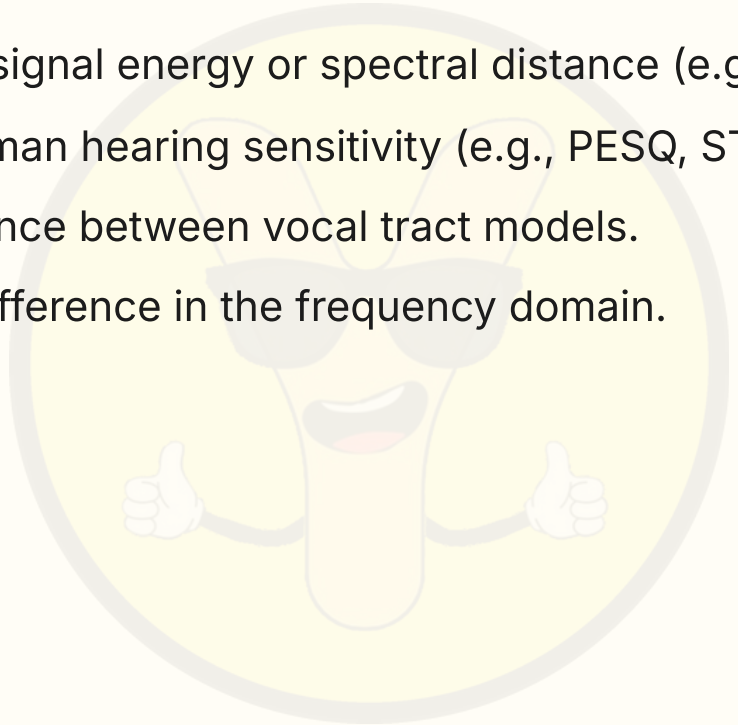
- Constructs a cost matrix between two sequences.
- Finds the optimal path that minimizes total cumulative distance.
- Constraint: Monotonicity, continuity, and boundary conditions.



Speech Distortion Measures

Distortion measures quantify the difference between original and processed speech.

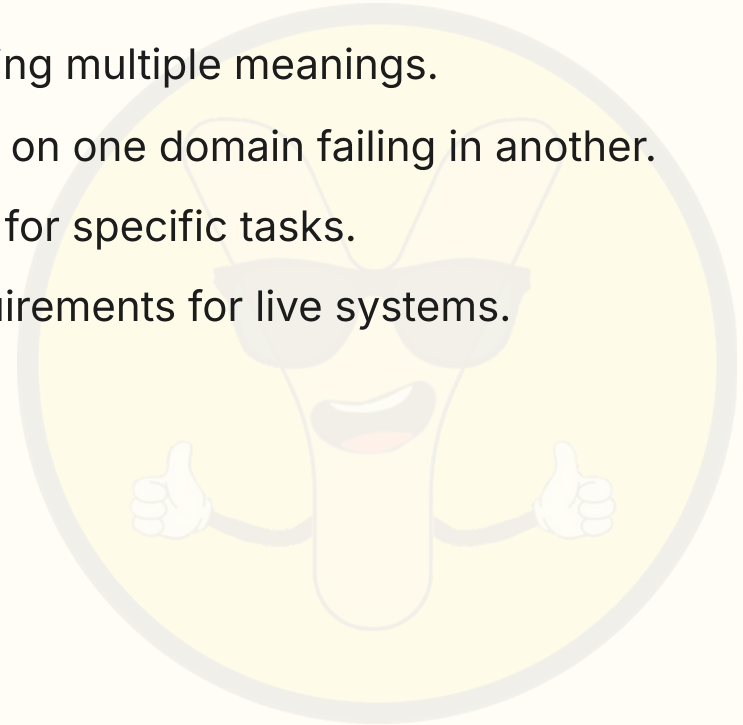
- Mathematical Measures: Based on signal energy or spectral distance (e.g., Euclidean distance).
- Perceptual Measures: Based on human hearing sensitivity (e.g., PESQ, STOI).
- LPC Distance: Measures the difference between vocal tract models.
- Log Spectral Distance: Measures difference in the frequency domain.



Real-World NLP Challenges

Deploying NLP in real-world scenarios introduces significant complexity.

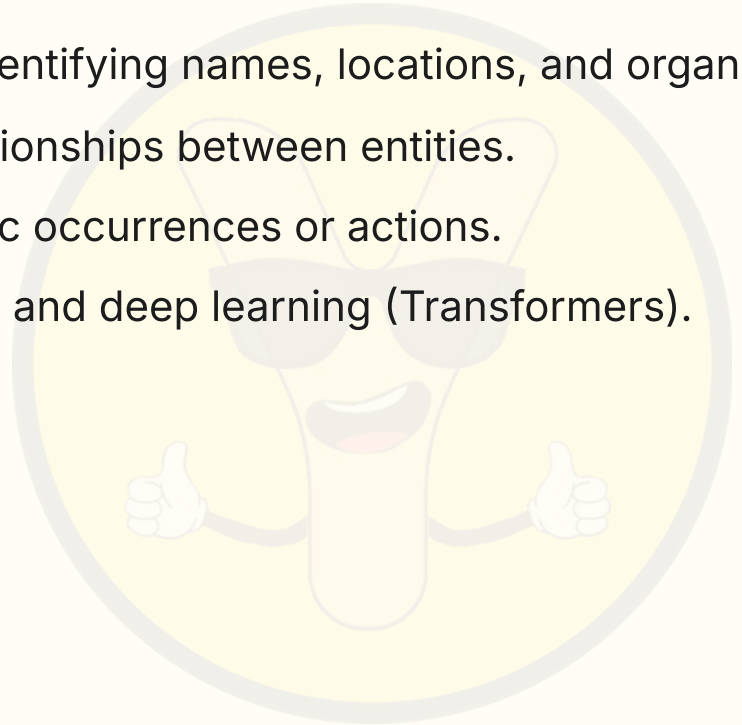
- Ambiguity: Words and phrases having multiple meanings.
- Domain Adaptation: Models trained on one domain failing in another.
- Data Sparsity: Lack of labeled data for specific tasks.
- Real-time constraints: Latency requirements for live systems.



Information Extraction (IE)

IE transforms unstructured text into structured data.

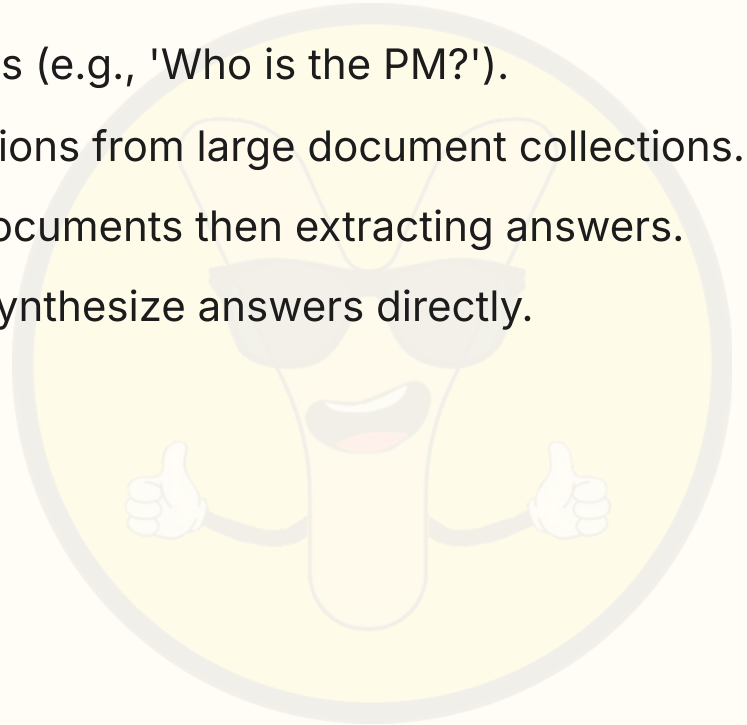
- Named Entity Recognition (NER): Identifying names, locations, and organizations.
- Relation Extraction: Identifying relationships between entities.
- Event Extraction: Identifying specific occurrences or actions.
- Techniques: Rule-based, statistical, and deep learning (Transformers).



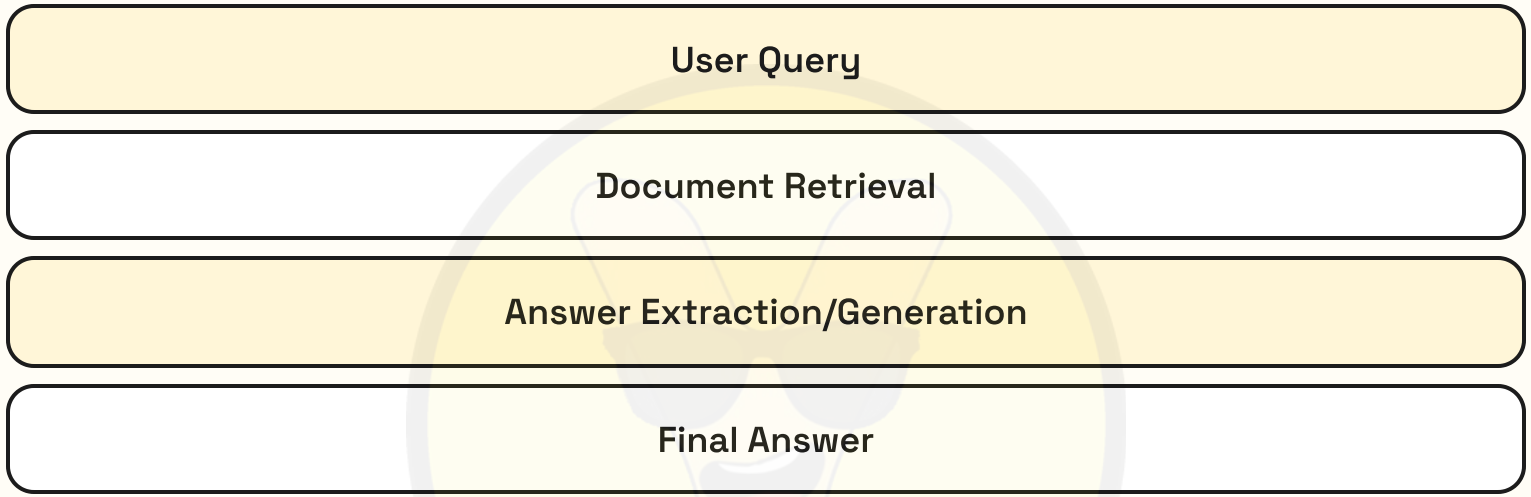
Question Answering (QA) Systems

QA systems provide direct answers to user queries.

- Factoid QA: Simple factual questions (e.g., 'Who is the PM?').
- Open-domain QA: Answering questions from large document collections.
- Retrieval-based: Finding relevant documents then extracting answers.
- Generative-based: Using LLMs to synthesize answers directly.



QA System Architecture

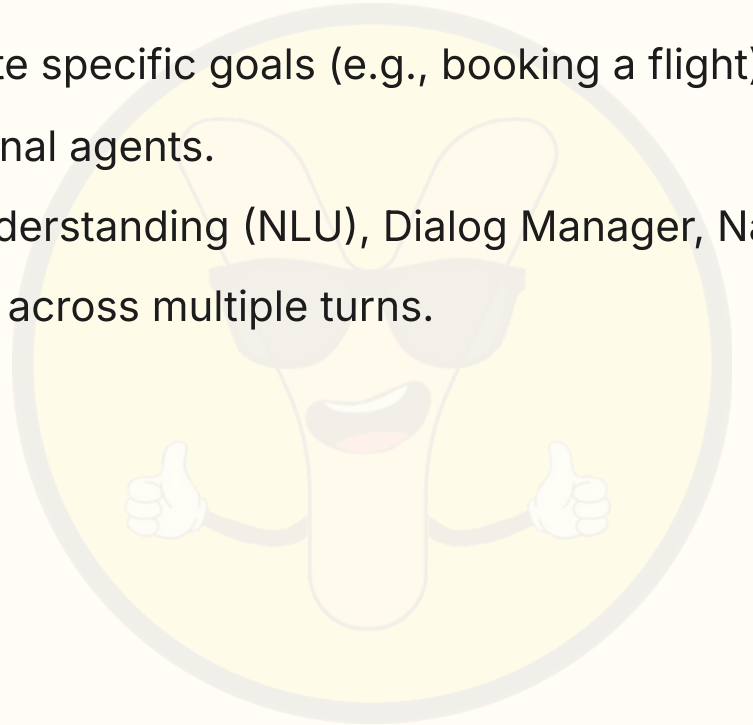


High-level architecture of a modern Question Answering system.

Dialog Engines

Dialog engines manage the state and flow of human-computer interaction.

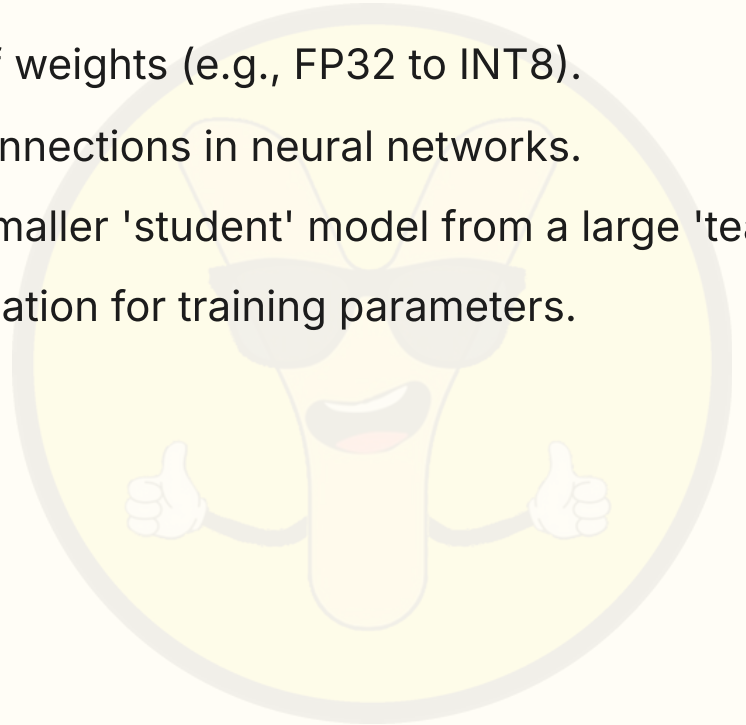
- Task-oriented: Designed to complete specific goals (e.g., booking a flight).
- Chit-chat: Open-ended conversational agents.
- Components: Natural Language Understanding (NLU), Dialog Manager, Natural Language Generation (NLG).
- State Tracking: Maintaining context across multiple turns.



Optimization in NLP

Optimization is critical for deploying large models efficiently.

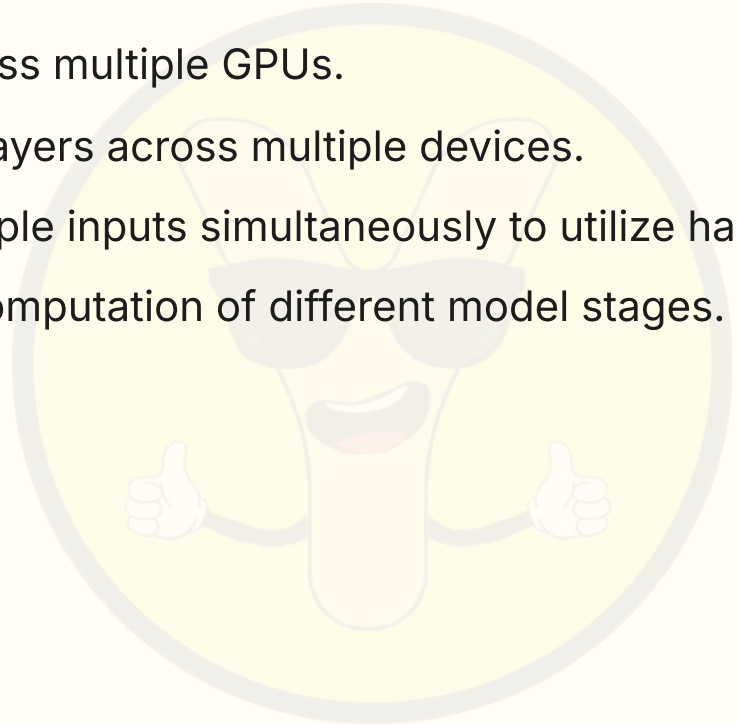
- Quantization: Reducing precision of weights (e.g., FP32 to INT8).
- Pruning: Removing unnecessary connections in neural networks.
- Knowledge Distillation: Training a smaller 'student' model from a large 'teacher'.
- Gradient Descent: Standard optimization for training parameters.



Parallelization and Batch Processing

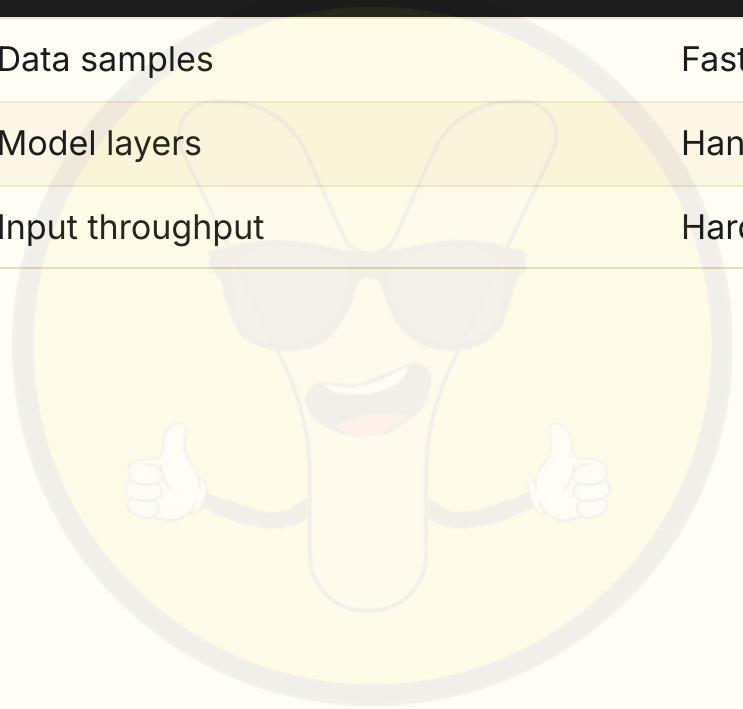
Techniques to improve throughput and training speed.

- Data Parallelism: Splitting data across multiple GPUs.
- Model Parallelism: Splitting model layers across multiple devices.
- Batch Processing: Processing multiple inputs simultaneously to utilize hardware vectorization.
- Pipeline Parallelism: Overlapping computation of different model stages.



Comparison of Parallelization Techniques

Technique	Focus	Benefit
Data Parallelism	Data samples	Faster training
Model Parallelism	Model layers	Handles large models
Batch Processing	Input throughput	Hardware efficiency



Exam corner

- **Q1** Explain the steps involved in MFCC feature extraction with a block diagram.
- **Q2** What is Dynamic Time Warping (DTW)? Explain its role in speech recognition.
- **Q3** Differentiate between mathematical and perceptual speech distortion measures.
- **Q4** Discuss the key challenges in building real-world Information Extraction systems.
- **Q5** Explain the architecture of a typical Task-Oriented Dialog Engine.
- **Q6** How do quantization and pruning help in optimizing NLP models?
- **Q7** Define Data Parallelism and Model Parallelism in the context of training large models.

Unit 5 done. You are exam-ready.

Revise the Exam corner once more, then open Unit 1 on theycotes.com. Every AKTU subject, unit by unit, free. Notes by Chirag Jain, Ycotes.

