Lab Assignment 2: Speech Sound Classification

Phoneme Extraction and Classification from Speech Signals

Objective: The objective of this experiment is to process a speech signal, extract specific phonemes, and visualize their waveforms while labeling them. In this experiment, we will:

1. Load a speech signal from the LJ Speech dataset.
2. Preprocess the audio (convert to mono, resample to 16kHz).
3. Use a pre-trained deep learning model (Wav2Vec2) to recognize phonemes.
4. Extract a phoneme segment from the speech signal based on time intervals.
5. Label and visualize the extracted phoneme by matching it with the recognized phonemes.

This experiment will help understand how deep learning-based speech models process spoken language and how phonemes can be visualized from continuous speech.

Expected Outcome: By the end of this experiment, students should be able to:

1. Successfully load and preprocess a speech signal.
2. Run the Wav2Vec2 model to recognize phonemes in the speech signal.
3. Extract a specific phoneme segment from the waveform using time indexing.
4. Label the extracted phoneme by aligning it with recognized phonemes.
5. Visualize the phoneme waveform with its corresponding label in a plot.

Example Output

1. Recognized Phonemes: T EH S T IH NG W AH N T UW (This represents "TESTING ONE TWO" in phonetic format.)
2. Extracted Phoneme between specific time interval and Waveform Plot.

Tools & Libraries to be used

- **Python** for implementation
- **Torchaudio** for loading and processing speech signals
- **Librosa** for visualization
- **Wav2Vec2** (Pre-trained model) for phoneme recognition
- **Matplotlib** for waveform plotting