Lab List for ICE-3208

Q1. Audio Loading and Signal Analysis

Using Python, read a given speech .wav file and perform:

a) Print sampling rate, number of samples, duration, max/min amplitude.

b) Plot the waveform with proper axis labeling.

c) Compute and plot magnitude spectrum using FFT.

Q2. Framing and Windowing (From Scratch)

Write Python code to segment a speech signal into frames:

a) Frame size = 25 ms, frame shift = 10 ms.

b) Apply Hamming window to each frame.

c) Show total number of frames and plot two sample frames (before & after window).

Q3. STFT and Spectrogram Generation

Generate a spectrogram of the given speech signal:

a) Implement STFT using frame-wise FFT (rfft).

b) Convert magnitude to dB and plot the spectrogram.

c) Write one practical application of spectrogram in speech processing.

Q4. MFCC Extraction

Implement MFCC extraction:

a) Implement the steps: pre-emphasis → framing → FFT → Mel filterbank → log → DCT.

b) Extract 13 MFCCs per frame and plot MFCC heatmap.

c) Mention why MFCC is effective for speech-related tasks (2–3 lines).

Q5. Pitch Detection Using Autocorrelation

Estimate pitch (F0) from speech:

a) Implement autocorrelation-based pitch estimation per frame.

b) Apply voiced/unvoiced detection using energy threshold.

c) Plot pitch contour (F0 vs time) for voiced frames.

Q6. Speech Enhancement by Spectral Subtraction

Noise reduction task:

a) Add white noise to the speech at SNR = 10 dB (or given SNR).

b) Implement spectral subtraction for enhancement.

c) Compute SNR before and after enhancement and report improvement.

Q7. DTW-Based Isolated Word Recognition

Build a small word recognizer:

a) Prepare dataset of 4 words with ≥ 3 training and ≥ 2 testing samples each.

b) Extract MFCC sequences and classify using DTW distance.

c) Report accuracy and draw confusion matrix.

Q8. Voice Activity Detection (VAD) and Segmentation

Implement rule-based VAD:

a) Compute short-time energy and zero crossing rate per frame.

b) Detect speech regions and show them on a plot.

c) Save detected speech regions as separate .wav segments.

Q9. Speaker Identification using Simple Embedding

Speaker ID task (≥ 3 speakers):

a) Compute embedding per utterance (mean MFCC + delta mean).

b) Train a classifier (KNN / SVM) and evaluate performance.

c) Report accuracy and classification report.

Q10. Real-Time Command Recognition

Develop a simple real-time recognizer:

a) Record 2 seconds audio from microphone for 2 commands (e.g., “start”, “stop”).

b) Extract MFCC-based feature and classify using nearest template / DTW.

c) Show predicted command and write one limitation.