LAB ASSIGNMENT – 1

Team Members :

1) G. Vineeth Reddy -AIE23008

2) G. Rishik Reddy -AIE23109

3) M. Naveen Kumar -AIE23168

INTRODUCTION TO SPEECH PROCESSING

Title of the Experiment:

Introduction to Speech Processing using LJ Speech Dataset.

Dataset Description:

The LJ Speech Dataset is a publicly available speech corpus consisting of short audio clips of a single speaker reading passages from books.

The dataset is provided in WAV format with a sampling rate of 22.05 kHz. It is widely used for speech and audio signal processing experiments.

This dataset is suitable for this experiment because it contains clean, high-quality speech signals.

Objectives:

1. To record and analyze a speech signal using a microphone.

2. To understand basic operations on speech signals such as slicing, normalization, amplification, and sampling.

3. To visualize voiced, unvoiced, and silence regions in speech.

Procedure and Code Explanation:

The speech signal is first loaded using Python libraries such as librosa and soundfile.

The waveform is plotted to observe amplitude variations over time.

Basic information such as sample rate, number of samples, and duration is calculated.

The signal is sliced to extract the first 2 seconds and normalized to control amplitude.

Amplification and de-amplification are performed by scaling the signal.

Up-sampling and down-sampling are carried out to study sampling effects.

Voiced, unvoiced, and silence regions are identified visually based on waveform energy.

Outputs:

- Original waveform plotted against time.

- Display of sample rate, total samples, and duration.

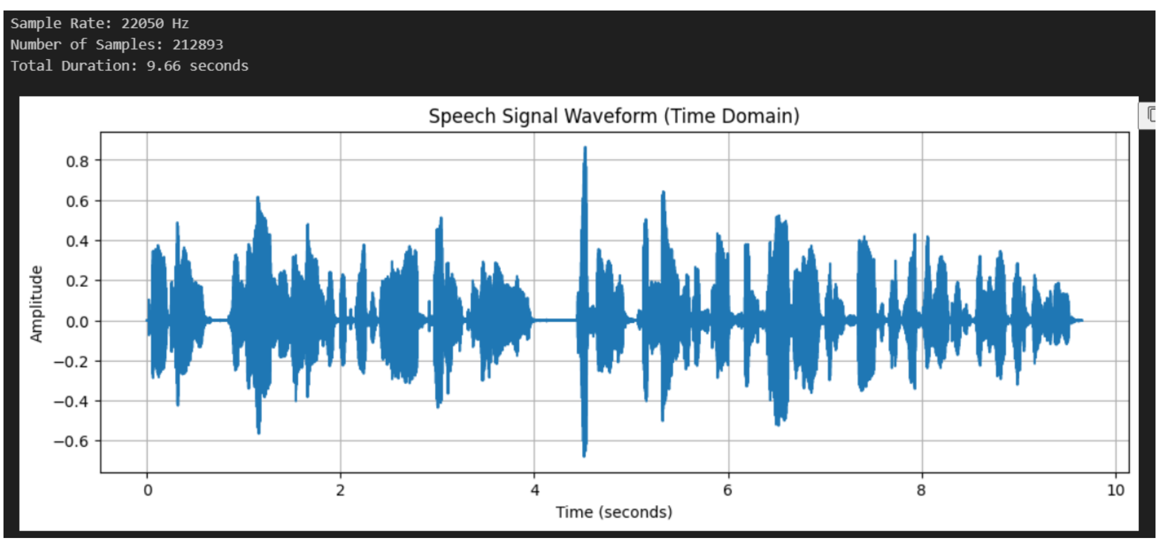
- Sliced and normalized waveform plots.

- Amplified and de-amplified signal plots.

- Up-sampled and down-sampled waveform plots.

- Visual identification of voiced, unvoiced, and silence regions.

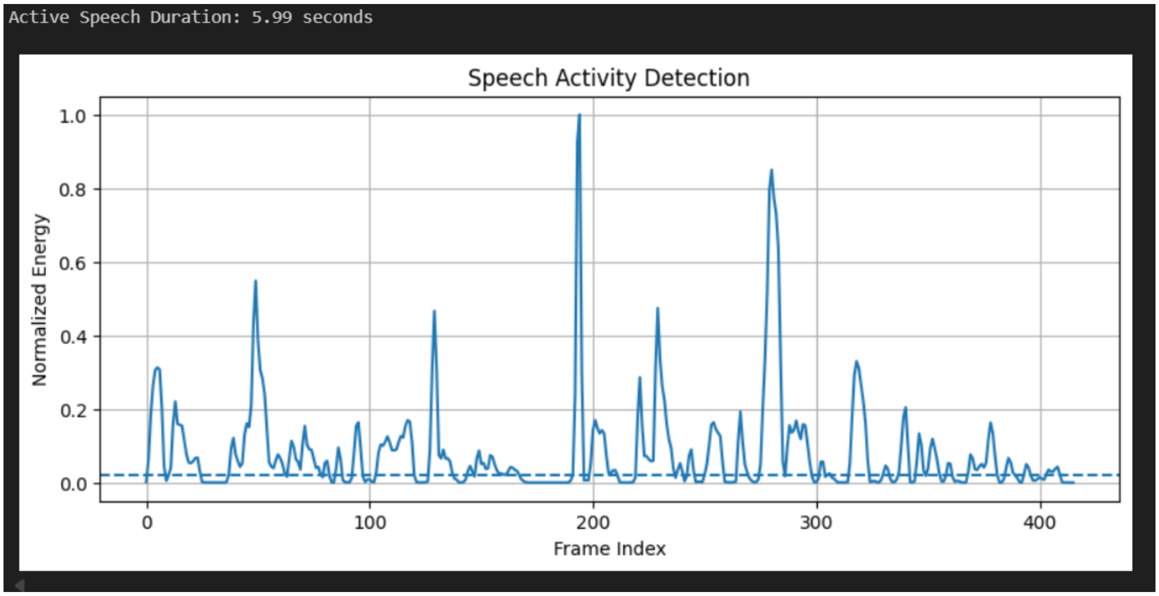
1)Speech Signal Waveform (Time Domain)



This plot shows the raw speech signal amplitude varying with time.
High peaks indicate voiced speech sounds, while low-amplitude regions represent silence or

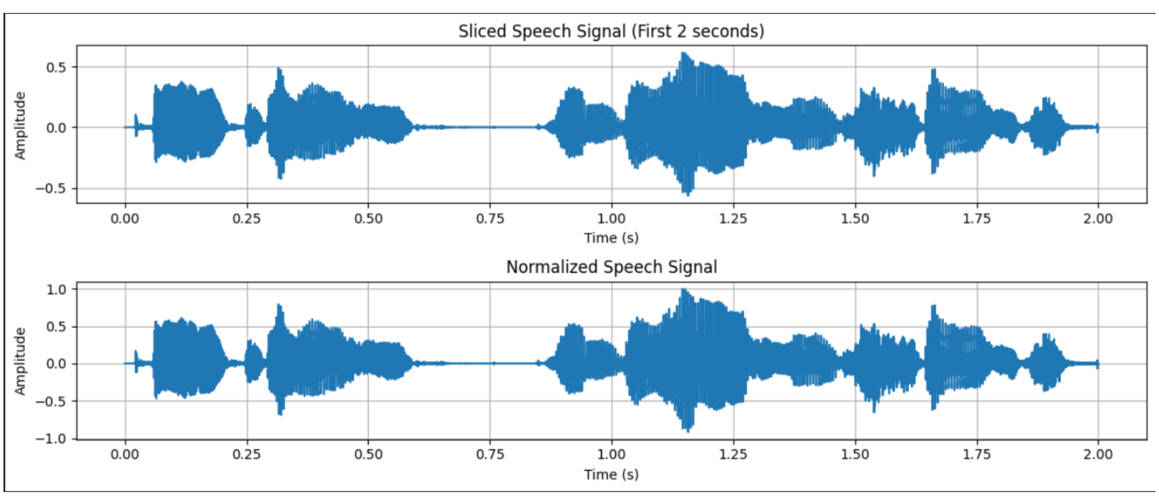
pauses.
It helps in understanding how speech energy changes over the duration of the signal.

2. Speech Activity Detection (Energy-Based)



This graph shows the normalized short-time energy of the speech signal across frames. Frames with energy above the threshold correspond to active speech, while those below indicate silence.
Using this method, the active speech duration is estimated as 5.99 seconds.

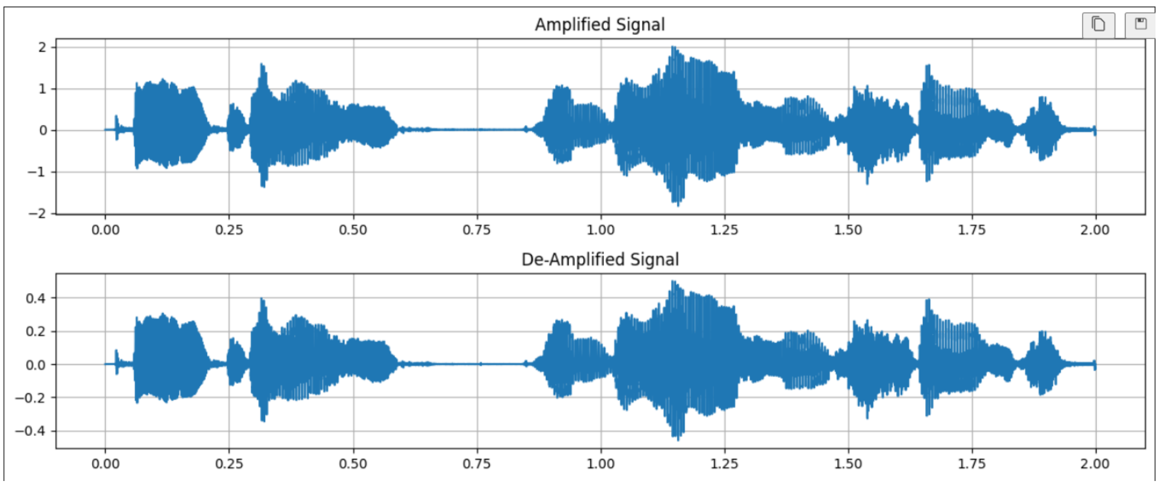
3.Sliced Speech Signal (First 2 Seconds) and Normalized Speech Signal



This waveform represents the first 2 seconds of the original speech signal extracted using slicing.
Slicing allows us to analyze a specific portion of speech instead of the entire signal.
It is useful for focused analysis and faster processing.

The normalized signal scales the amplitude so that it lies between −1 and +1.
Normalization prevents signal distortion and ensures uniform amplitude levels for processing.
It is commonly used before further signal manipulation or feature extraction.

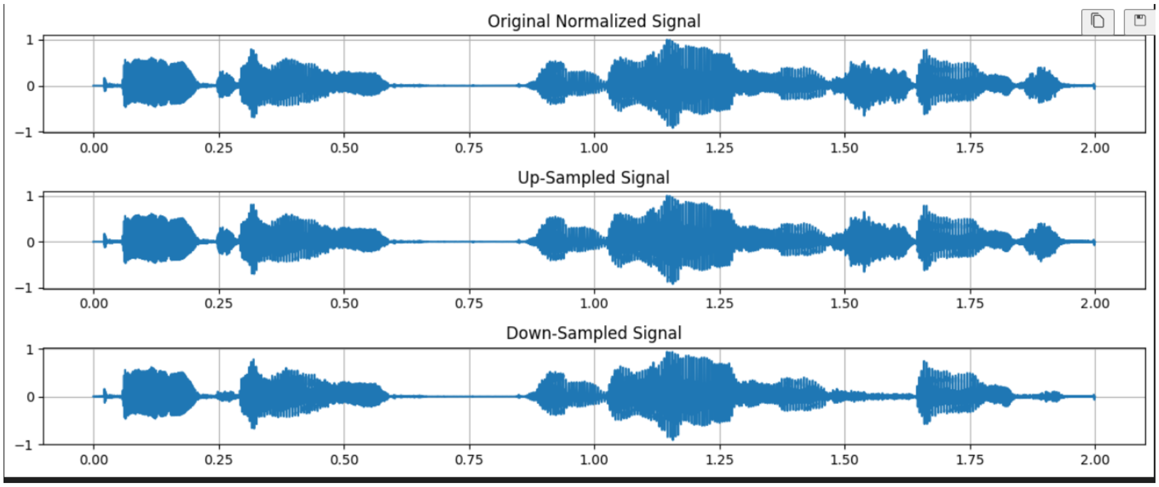
4)Amplified Signal and De-Amplified Signal



In the amplified signal, the amplitude is increased by multiplying the signal with a factor greater than one.
This makes the speech louder and more prominent in the waveform.
Amplification is useful when the original signal is weak or low in volume.

The de-amplified signal is obtained by reducing the signal amplitude using a scaling factor less than one.
This results in a lower amplitude waveform without changing the speech content.
It is helpful to study the effect of reduced signal strength.

5)Original Normalized Signal , Up-Sampled Signal , Down-Sampled Signal.

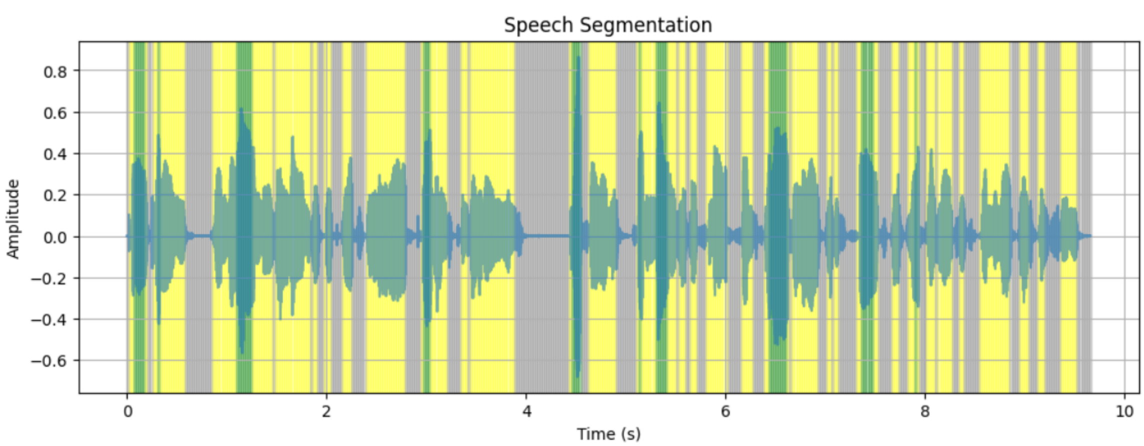


i)This plot shows the original speech signal after normalization.
The waveform shape remains the same, but the amplitude range is controlled.
It serves as a reference for comparing up-sampled and down-sampled signals.

ii) Up-sampling increases the number of samples in the signal by interpolation.
This results in a smoother waveform with higher time resolution.
It is commonly used when converting signals to a higher sampling rate.

iii) Down-sampling reduces the number of samples by keeping every alternate sample.
The waveform appears less smooth and may lose fine details.
This demonstrates the effect of lowering the sampling rate.

6)Speech Segmentation



This plot shows the segmentation of speech into voiced, unvoiced, and silence regions.
Voiced regions have high energy and regular patterns, while silence regions have very low

amplitude.
Segmentation helps in identifying different speech components for further analysis.

Conclusion

In this experiment, basic speech signal processing operations were performed and analyzed using a real speech sample. By observing the waveform and applying operations like slicing, normalization, amplification, sampling, and segmentation, the characteristics of speech signals were clearly understood. This lab provided hands-on experience with speech processing techniques and helped in understanding how speech signals are represented and manipulated in the time domain.