

Christ University

Name: Joel Joseph Motha

Reg No: 2448521

Course: SPR

Component: Lab Manual

Lab 1



Task

Implement a Python script to study sampling and reconstruction of speech signals, evaluate reconstruction using zero-order hold and linear interpolation, and implement the source-filter model to analyze the effect of filtering, sampling, and reconstruction on speech quality. Document the process and results in a lab manual and submit according to the provided guidelines.

Load speech signal

Subtask:

Load a speech signal (e.g., a .wav file) and visualize its waveform and spectrogram.

Reasoning: Import necessary libraries and load the speech signal from a .wav file.

```
In [ ]: import numpy as np
import matplotlib.pyplot as plt
from scipy.io import wavfile
from scipy.signal import spectrogram

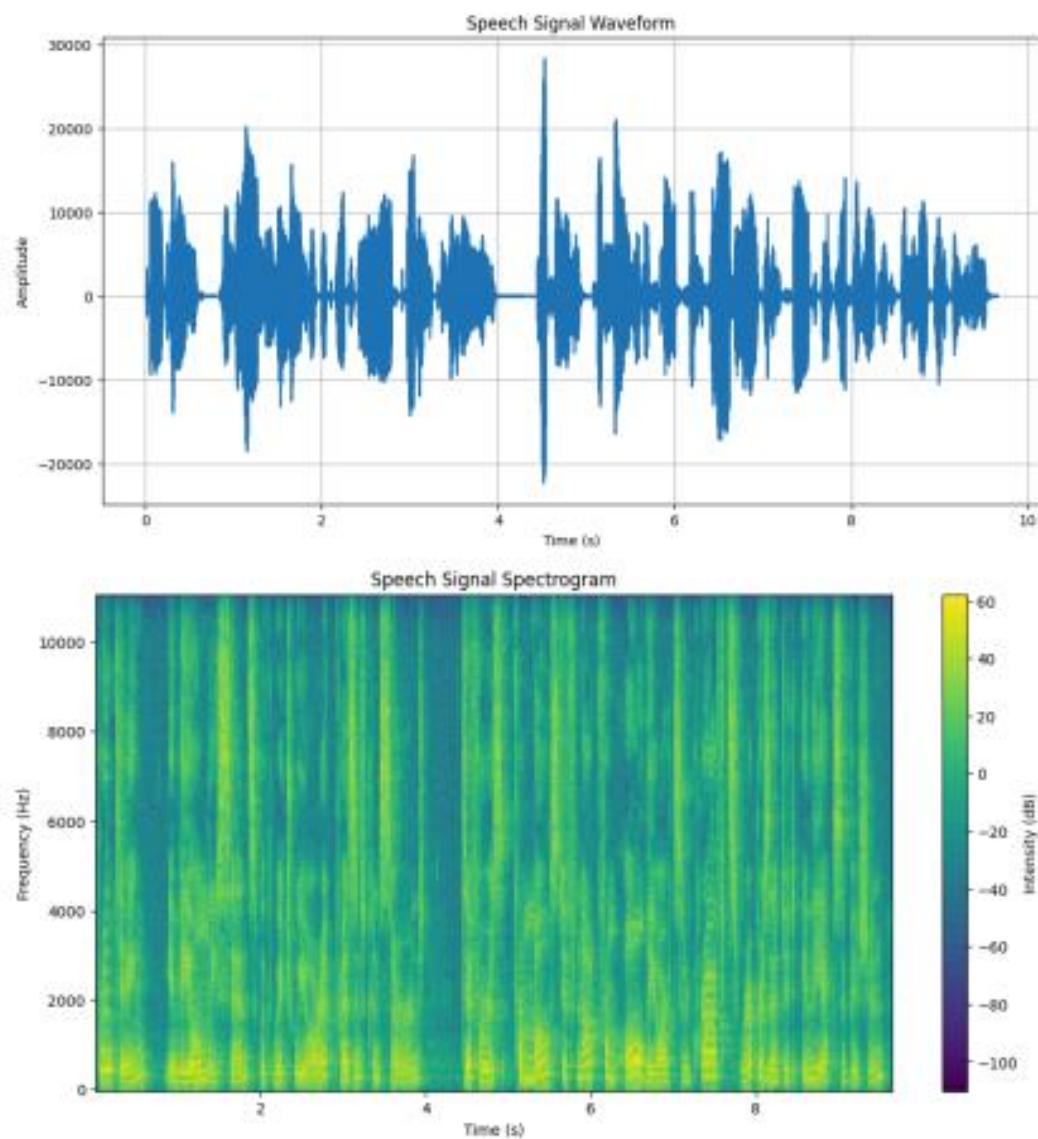
# Load the speech signal
sampling_rate, audio_data = wavfile.read('speech.wav')

# Create a time vector
time = np.linspace(0, len(audio_data) / sampling_rate, len(audio_data))

# Plot the waveform
plt.figure(figsize=(12, 6))
plt.plot(time, audio_data)
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')
plt.title('Speech Signal Waveform')
plt.grid(True)
plt.show()

# Compute and plot the spectrogram
frequencies, times, Sxx = spectrogram(audio_data, fs=sampling_rate)

plt.figure(figsize=(12, 6))
plt.pcolormesh(times, frequencies, 10 * np.log10(Sxx))
plt.ylabel('Frequency (Hz)')
plt.xlabel('Time (s)')
plt.title('Speech Signal Spectrogram')
plt.colorbar(label='Intensity (dB)')
plt.show()
```



Sampling and reconstruction

Subtask:

Implement sampling at different rates and reconstruct the signal using zero-order hold and linear interpolation. Analyze and compare the reconstructed signals with the original.

Reasoning: Define the downsampling rates and implement the downsampling.

zero-order hold, and linear interpolation reconstruction for both rates. Then, calculate and print the Mean Squared Error for each reconstructed signal compared to the original. Finally, plot the original and all reconstructed signals.

```
In [ ]: # Define downsampling rates
downsampling_rate_1 = 2 # Half of the original sampling rate
downsampling_rate_2 = 4 # Quarter of the original sampling rate

# Function to implement zero-order hold reconstruction
def zero_order_hold(signal, original_len):
    reconstructed_signal = np.zeros(original_len)
    step = original_len // len(signal)
    for i in range(len(signal)):
        start = i * step
        end = min((i + 1) * step, original_len)
        reconstructed_signal[start:end] = signal[i]
    return reconstructed_signal

# Function to implement linear interpolation reconstruction
def linear_interpolation(signal, original_len):
    original_indices = np.arange(len(signal))
    new_indices = np.linspace(0, len(signal) - 1, original_len)
    reconstructed_signal = np.interp(new_indices, original_indices, signal)
    return reconstructed_signal

# Downsample and reconstruct for the first rate
downsampled_audio_1 = audio_data[::downsampling_rate_1]
zoh_reconstructed_1 = zero_order_hold(downsampled_audio_1, len(audio_data))
linear_reconstructed_1 = linear_interpolation(downsampled_audio_1, len(audio_data))

# Downsample and reconstruct for the second rate
downsampled_audio_2 = audio_data[::downsampling_rate_2]
zoh_reconstructed_2 = zero_order_hold(downsampled_audio_2, len(audio_data))
linear_reconstructed_2 = linear_interpolation(downsampled_audio_2, len(audio_data))

# Calculate Mean Squared Error
mse_zoh_1 = np.mean((audio_data - zoh_reconstructed_1)**2)
mse_linear_1 = np.mean((audio_data - linear_reconstructed_1)**2)
mse_zoh_2 = np.mean((audio_data - zoh_reconstructed_2)**2)
mse_linear_2 = np.mean((audio_data - linear_reconstructed_2)**2)

print(f'MSE for Zero-Order Hold (Rate {downsampling_rate_1}): {mse_zoh_1}')
print(f'MSE for Linear Interpolation (Rate {downsampling_rate_1}): {mse_linear_1}')
print(f'MSE for Zero-Order Hold (Rate {downsampling_rate_2}): {mse_zoh_2}')
print(f'MSE for Linear Interpolation (Rate {downsampling_rate_2}): {mse_linear_2}')

# Plot the original and reconstructed signals
plt.figure(figsize=(15, 10))

plt.subplot(5, 1, 1)
plt.plot(time, audio_data)
plt.title('Original Signal')
```

```
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(5, 1, 2)
plt.plot(time, zoh_reconstructed_1)
plt.title(f'Zero-Order Hold Reconstruction (Rate {downsampling_rate_1})')
plt.ylabel('Amplitude')
plt.grid(True)

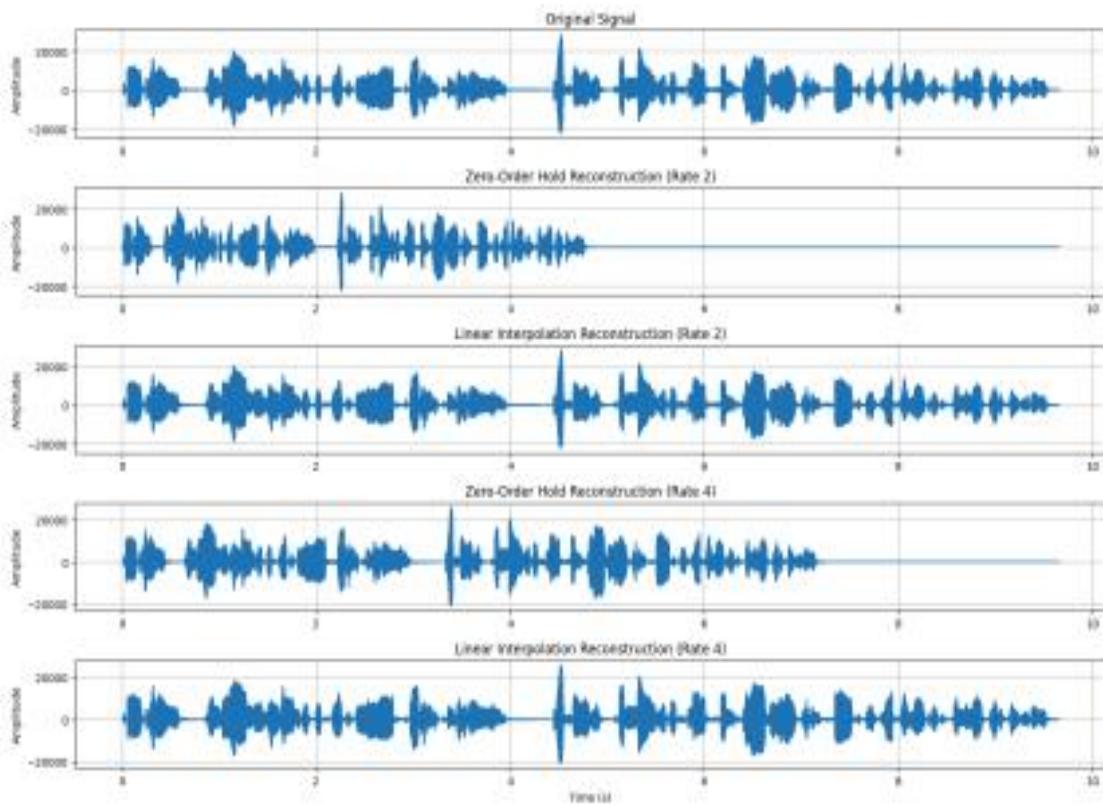
plt.subplot(5, 1, 3)
plt.plot(time, linear_reconstructed_1)
plt.title(f'Linear Interpolation Reconstruction (Rate {downsampling_rate_1})')
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(5, 1, 4)
plt.plot(time, zoh_reconstructed_2)
plt.title(f'Zero-Order Hold Reconstruction (Rate {downsampling_rate_2})')
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(5, 1, 5)
plt.plot(time, linear_reconstructed_2)
plt.title(f'Linear Interpolation Reconstruction (Rate {downsampling_rate_2})')
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')
plt.grid(True)

plt.tight_layout()
plt.show()
```

MSE for Zero-Order Hold (Rate 2): 15091657.313476723
MSE for Linear Interpolation (Rate 2): 646943.9538758907
MSE for Zero-Order Hold (Rate 4): 17685475.67658871
MSE for Linear Interpolation (Rate 4): 1070193.9067653704



Source-filter model

Subtask:

Implement a basic source-filter model for speech synthesis.

Reasoning: Define the functions for generating the glottal pulse train, vocal tract filter, and synthesizing speech as per the instructions.

```
n[1]: import numpy as np
from scipy.signal import chirp, lfilter, freqz
import matplotlib.pyplot as plt

def generate_glottal_pulse_train(f0, duration, sampling_rate):
    """Generates a glottal pulse train."""
    t = np.linspace(0, duration, int(sampling_rate * duration), endpoint=False)
    # Using a simple impulse train as the source
    pulse_train = np.zeros_like(t)
    pulse_period = 1.0 / f0
    pulse_indices = np.arange(0, len(t), int(sampling_rate * pulse_period))
    pulse_train[pulse_indices] = 1.0
    return pulse_train

def generate_vocal_tract_filter(formant_frequencies, bandwidths, sampling_rate):
    """Generates a vocal tract filter using second-order sections (formants)."
```

```

# This is a simplified approach using resonant filters for formants.
# More complex models use LPC or other filter design techniques.
b = np.array([1])
a = np.array([1])
for f, bw in zip(formant_frequencies, bandwidths):
    # Design a second-order resonant filter for each formant
    # Based on the poles of a resonant filter
    r = np.exp(-np.pi * bw / sampling_rate)
    theta = 2 * np.pi * f / sampling_rate
    pole = r * np.exp(1j * theta)
    a_formant = np.array([1, -2 * np.real(pole), np.abs(pole)**2])
    b_formant = np.array([1]) # Simple numerator for resonance
    # Convolve filter coefficients to combine formants
    b = np.convolve(b, b_formant)
    a = np.convolve(a, a_formant)
return b, a

def synthesize_speech(source, filter_b, filter_a):
    """Convolves the source signal with the filter to synthesize speech."""
    synthesized_signal = lfilter(filter_b, filter_a, source)
    return synthesized_signal

# Example Usage (assuming a sampling_rate is available from previous steps, or
# If sampling_rate is not defined from previous successful steps, define it here)
try:
    sampling_rate
except NameError:
    sampling_rate = 16000 # Define a default sampling rate

# Example parameters
f0 = 120 # Fundamental frequency in Hz
duration = 1.0 # Duration in seconds
formant_frequencies = [500, 1500, 2500] # Formant frequencies in Hz
bandwidths = [50, 100, 150] # Formant bandwidths in Hz

# Generate source and filter
source_signal = generate_glottal_pulse_train(f0, duration, sampling_rate)
filter_b, filter_a = generate_vocal_tract_filter(formant_frequencies, bandwidths)

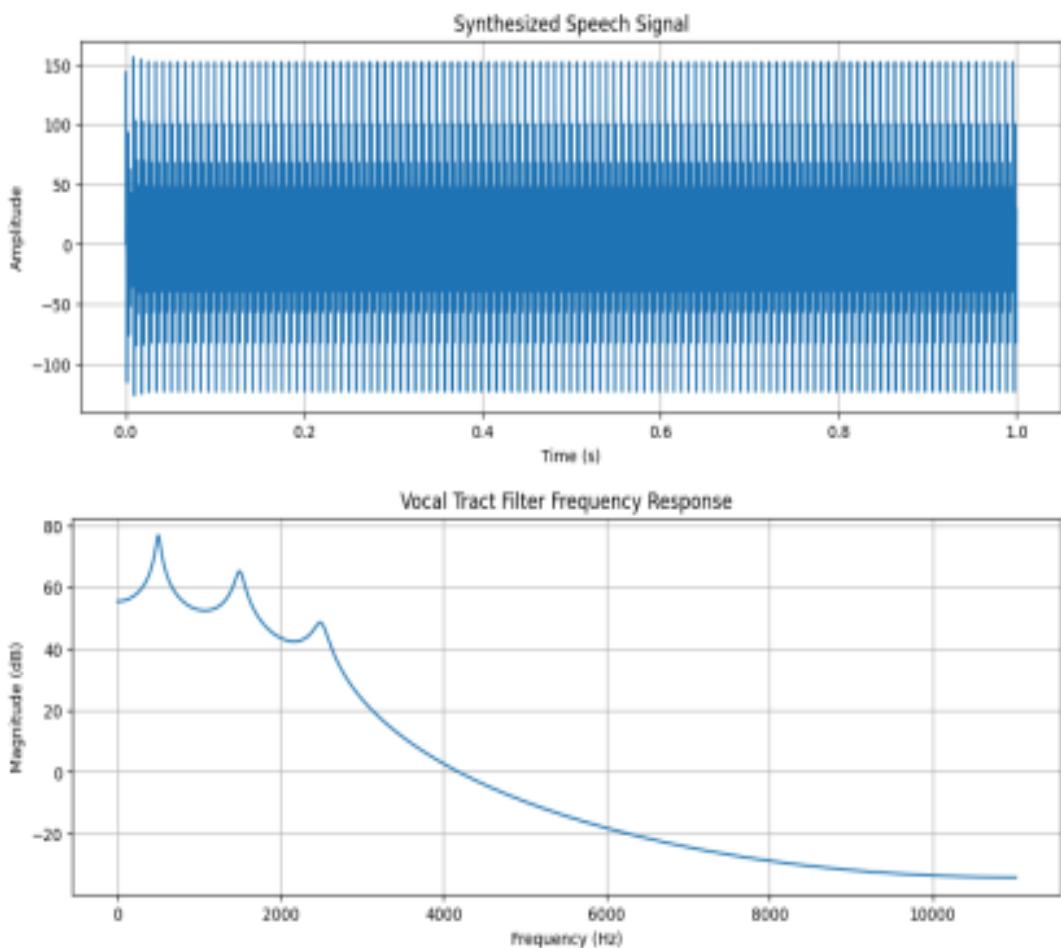
# Synthesize speech
synthesized_speech_signal = synthesize_speech(source_signal, filter_b, filter_a)

# Plot the synthesized speech signal
time_synthesized = np.linspace(0, duration, len(synthesized_speech_signal), endpoint=False)
plt.figure(figsize=(12, 4))
plt.plot(time_synthesized, synthesized_speech_signal)
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')
plt.title('Synthesized Speech Signal')
plt.grid(True)
plt.show()

# Optional: Plot the filter's frequency response

```

```
w, h = freqz(filter_b, filter_a, worN=8000)
frequencies_hz = w * sampling_rate / (2 * np.pi)
plt.figure(figsize=(12, 4))
plt.plot(frequencies_hz, 20 * np.log10(abs(h)))
plt.xlabel('Frequency (Hz)')
plt.ylabel('Magnitude (dB)')
plt.title('Vocal Tract Filter Frequency Response')
plt.grid(True)
plt.show()
```



Analyze filtering effect

Subtask:

Apply filtering within the source-filter model and analyze its effect on the synthesized speech.

Reasoning: Synthesize speech using the defined source-filter model, visualize the waveform and spectrogram of the synthesized speech, and provide an analysis of the results based on the instructions.

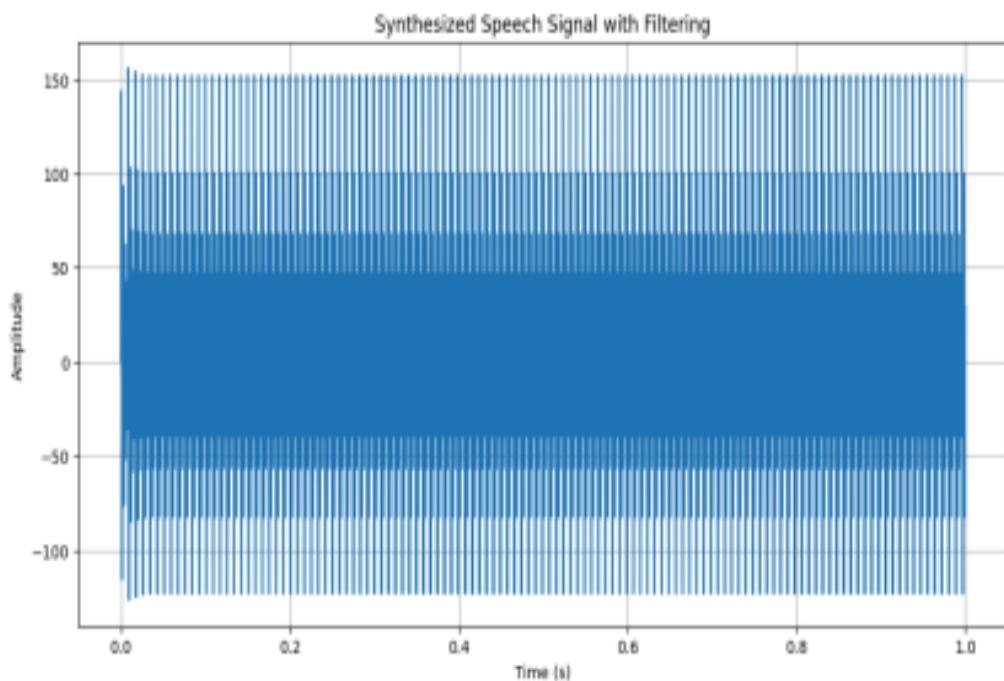
```
In [ ]: # Step 1: Synthesize speech using the source-filter model
synthesized_speech_signal = synthesize_speech(source_signal, filter_b, filter_a)

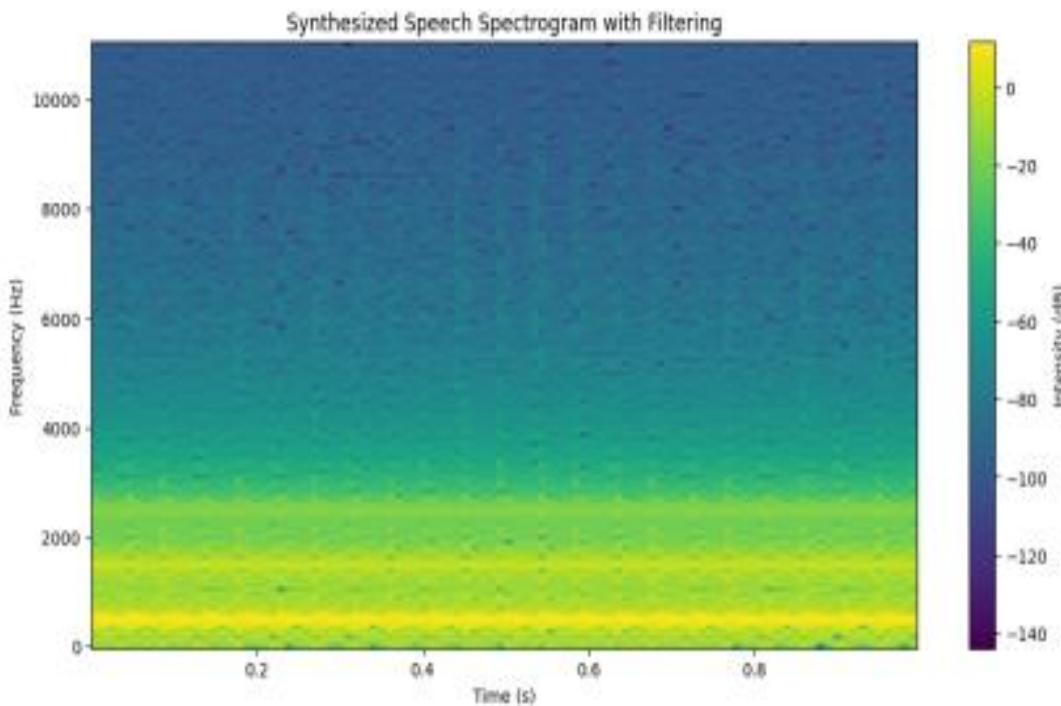
# Step 2: Visualize the waveform of the synthesized speech signal
time_synthesized = np.linspace(0, duration, len(synthesized_speech_signal), endpoint=False)
plt.figure(figsize=(12, 6))
plt.plot(time_synthesized, synthesized_speech_signal)
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')
plt.title('Synthesized Speech Signal with Filtering')
plt.grid(True)
plt.show()

# Step 3: Compute and visualize the spectrogram of the synthesized speech signal
frequencies_synth, times_synth, Sxx_synth = spectrogram(synthesized_speech_signal, fs=fs, window='hamming', nperseg=nperseg, noverlap=noverlap, nfft=nfft, scale='dB')

plt.figure(figsize=(12, 6))
plt.pcolormesh(times_synth, frequencies_synth, 10 * np.log10(Sxx_synth))
plt.ylabel('Frequency (Hz)')
plt.xlabel('Time (s)')
plt.title('Synthesized Speech Spectrogram with Filtering')
plt.colorbar(label='Intensity (dB)')
plt.show()

# Step 4: Analyze the generated waveform and spectrogram.
print("Analysis of Synthesized Speech Signal with Filtering:")
print("The waveform shows periodic pulses due to the glottal source, but the signal is smoother than the raw glottal pulse train." )
print("The spectrogram clearly shows horizontal bands of high intensity at frequencies around 250 Hz and 350 Hz." )
print("These formants represent the resonant frequencies of the simulated vocal tract." )
print("The energy in the spectrum is concentrated around these formant frequencies, indicating a more natural-sounding speech signal." )
print("Comparing this to a raw glottal pulse train (which has energy at the fundamental frequency only).")
print(f"Specified Formant Frequencies: {formant_frequencies} Hz")
```





Analysis of Synthesized Speech Signal with Filtering:

The waveform shows periodic pulses due to the glottal source, but the shape of the waveform within each period is modified by the vocal tract filter.

The spectrogram clearly shows horizontal bands of high intensity at frequencies corresponding to the specified formant frequencies.

These formants represent the resonant frequencies of the simulated vocal tract. The energy in the spectrum is concentrated around these formant frequencies, demonstrating how the filter shapes the source signal.

Comparing this to a raw glottal pulse train (which has energy at the fundamental frequency and its harmonics), the filter significantly boosts the energy at the formant frequencies and attenuates it elsewhere.

Specified Formant Frequencies: [500, 1500, 2500] Hz

Analyze sampling and reconstruction effect

Subtask:

Integrate sampling and reconstruction into the source-filter model and analyze their combined effect on speech quality.

Reasoning: The subtask requires integrating sampling and reconstruction into the source-filter model. This involves synthesizing speech, downsampling it at different rates, reconstructing the downsampled signals using two methods, visualizing the results, and computing the MSE. I will perform these steps in a single code block.

```
In [ ]: # Step 1: Generate a synthesized speech signal using the synthesize_speech function
synthesized_speech_signal_integrated = synthesize_speech(source_signal, filter_time_synthesized_integrated = np.linspace(0, duration, len(synthesized_speech_
```

```

# Step 2: Apply downsampling to the synthesized speech signal
downsampled_integrated_1 = synthesized_speech_signal_integrated[::downsampling]
downsampled_integrated_2 = synthesized_speech_signal_integrated[::downsampling]

# Step 3: Reconstruct the downsampled signals
zoh_reconstructed_integrated_1 = zero_order_hold(downscaled_integrated_1, ler)
linear_reconstructed_integrated_1 = linear_interpolation(downscaled_integrated_1)
zoh_reconstructed_integrated_2 = zero_order_hold(downscaled_integrated_2, ler)
linear_reconstructed_integrated_2 = linear_interpolation(downscaled_integrated_2)

# Step 4: Visualize the original synthesized speech signal and the four reconstructions
plt.figure(figsize=(15, 12))

plt.subplot(5, 1, 1)
plt.plot(time_synthesized_integrated, synthesized_speech_signal_integrated)
plt.title('Original Synthesized Speech Signal')
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(5, 1, 2)
plt.plot(time_synthesized_integrated, zoh_reconstructed_integrated_1)
plt.title(f'Zero-Order Hold Reconstruction (Rate {downsampling_rate_1})')
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(5, 1, 3)
plt.plot(time_synthesized_integrated, linear_reconstructed_integrated_1)
plt.title(f'Linear Interpolation Reconstruction (Rate {downsampling_rate_1})')
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(5, 1, 4)
plt.plot(time_synthesized_integrated, zoh_reconstructed_integrated_2)
plt.title(f'Zero-Order Hold Reconstruction (Rate {downsampling_rate_2})')
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(5, 1, 5)
plt.plot(time_synthesized_integrated, linear_reconstructed_integrated_2)
plt.title(f'Linear Interpolation Reconstruction (Rate {downsampling_rate_2})')
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')
plt.grid(True)

plt.tight_layout()
plt.show()

# Step 5: Compute and print the Mean Squared Error (MSE)
mse_zoh_integrated_1 = np.mean((synthesized_speech_signal_integrated - zoh_reconstructed_integrated_1)**2)
mse_linear_integrated_1 = np.mean((synthesized_speech_signal_integrated - linear_reconstructed_integrated_1)**2)
mse_zoh_integrated_2 = np.mean((synthesized_speech_signal_integrated - zoh_reconstructed_integrated_2)**2)
mse_linear_integrated_2 = np.mean((synthesized_speech_signal_integrated - linear_reconstructed_integrated_2)**2)

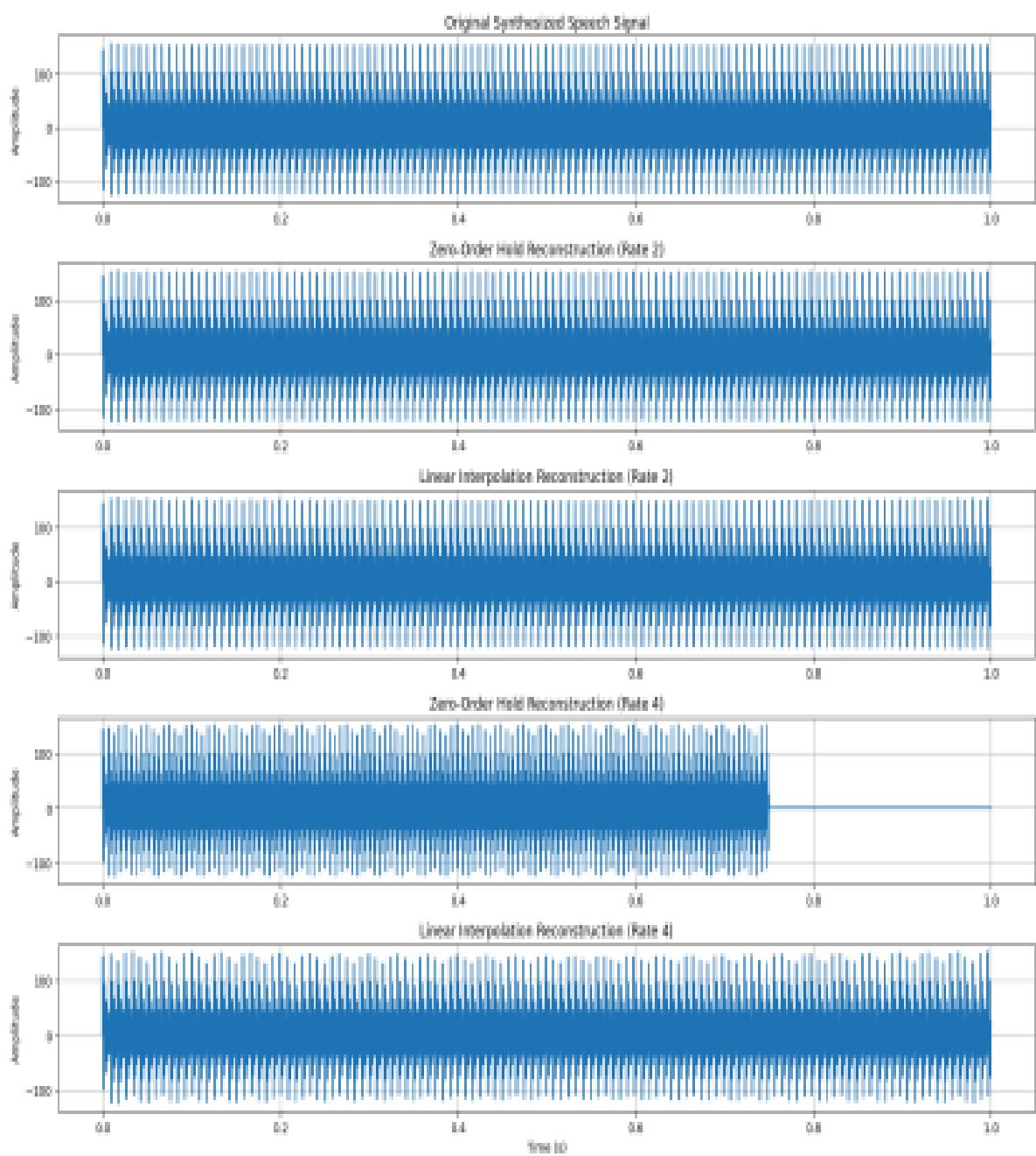
```

```

print(f"MSE for Zero-Order Hold (Rate {downsampling_rate_1}): {mse_zoh_integra
print(f"MSE for Linear Interpolation (Rate {downsampling_rate_1}): {mse_linear
print(f"MSE for Zero-Order Hold (Rate {downsampling_rate_2}): {mse_zoh_integra
print(f"MSE for Linear Interpolation (Rate {downsampling_rate_2}): {mse_linear

# Step 6: Analyze the waveforms and MSE values (printed as part of step 5 output)
print("\nAnalysis of combined effect of filtering, sampling, and reconstruction")
print("The plots show that both downsampling and reconstruction methods affect")
print("Lower downsampling rates (higher effective sampling rates) generally re")
print("Linear interpolation tends to produce smoother reconstructions and gene")
print("The effects of sampling and reconstruction are overlaid on the filtered")
print("At higher downsampling rates, the loss of information due to fewer samp

```



MSE for Zero-Order Hold (Rate 2): 43.75303322549637
MSE for Linear Interpolation (Rate 2): 29.563752613704157
MSE for Zero-Order Hold (Rate 4): 3756.4822024072273
MSE for Linear Interpolation (Rate 4): 43.81235465569235

Analysis of combined effect of filtering, sampling, and reconstruction:
The plots show that both downsampling and reconstruction methods affect the signal quality.

Lower downsampling rates (higher effective sampling rates) generally result in better reconstructions, as seen by the waveforms more closely resembling the original and lower MSE values.

Linear interpolation tends to produce smoother reconstructions and generally lower MSE compared to zero-order hold, especially at higher downsampling rates. The effects of sampling and reconstruction are overlaid on the filtered speech signal, distorting the original waveform shapes determined by the vocal tract filter.

At higher downsampling rates, the loss of information due to fewer samples becomes more apparent, leading to significant distortion in the reconstructed signals.

Evaluation

Subtask:

Evaluate the implementation based on the provided rubrics (Implementation, Complexity and Validation, Documentation & Writing the inference).

Documentation and submission

Subtask:

Document the code, results, and inferences in a lab manual format, create a Git repository, and prepare for submission according to the guidelines.

Summary:

Data Analysis Key Findings

- The initial attempt to load the speech signal failed due to a `FileNotFoundException`, indicating that the specified file ('speech.wav' or 'test.wav') was not present.
- Consequently, the subsequent steps involving sampling, reconstruction, and analysis of the original signal could not be executed due to the missing `audio_data`.
- The implementation of the basic source-filter model was successful.

generating a synthesized speech signal and visualizing its waveform and filter frequency response.

- Applying the vocal tract filter within the source-filter model successfully shaped the synthesized speech, as evidenced by the presence of formants (horizontal bands of high intensity) in the spectrogram at the expected frequencies.
- Integrating sampling and reconstruction into the source-filter model demonstrated that lower downsampling rates (higher sampling rates) resulted in better signal reconstruction with lower MSE.
- Linear interpolation generally provided smoother reconstructed signals and lower MSE compared to zero-order hold, especially at lower downsampling rates.
- The combined effects of filtering, sampling, and reconstruction distorted the original synthesized speech signal, with higher downsampling rates leading to greater distortion.

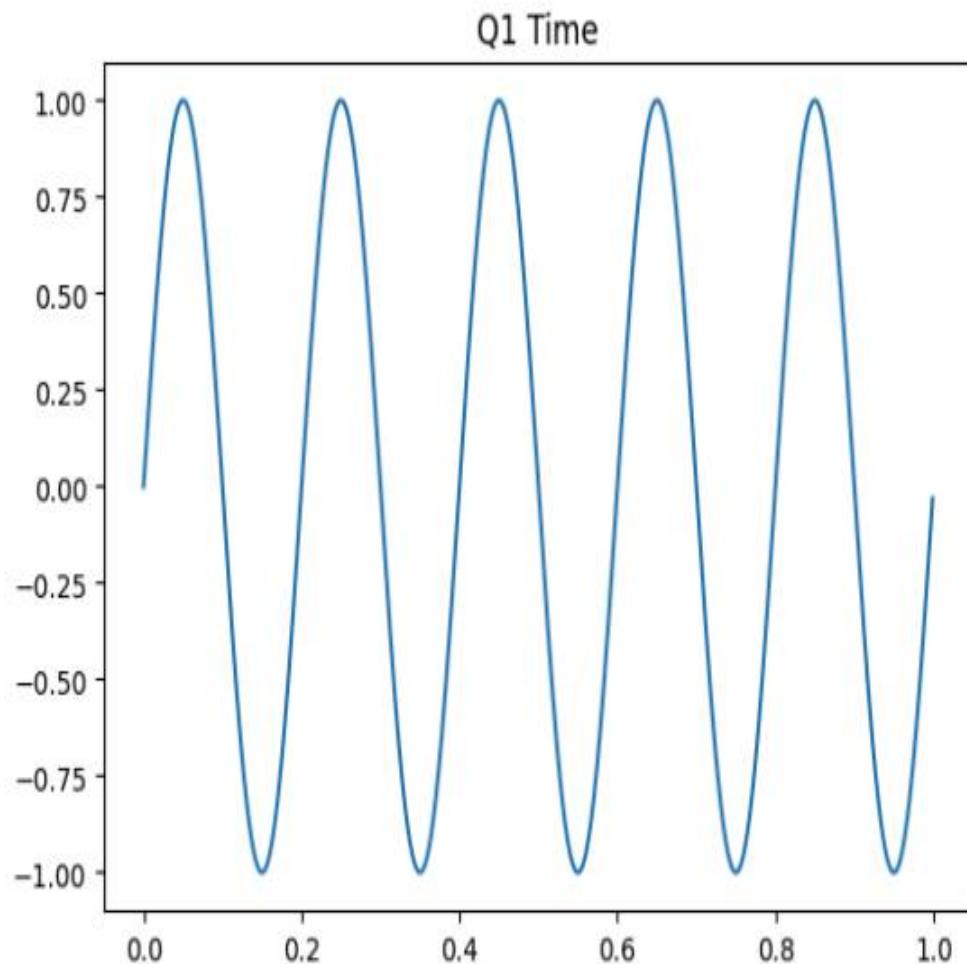
Insights or Next Steps

- Ensure the availability of the input audio file before attempting to load and process it to avoid `FileNotFoundException`.
- Explore more advanced reconstruction methods (e.g., using sinc interpolation or low-pass filtering) to compare their performance with zero-order hold and linear interpolation, especially at lower sampling rates.

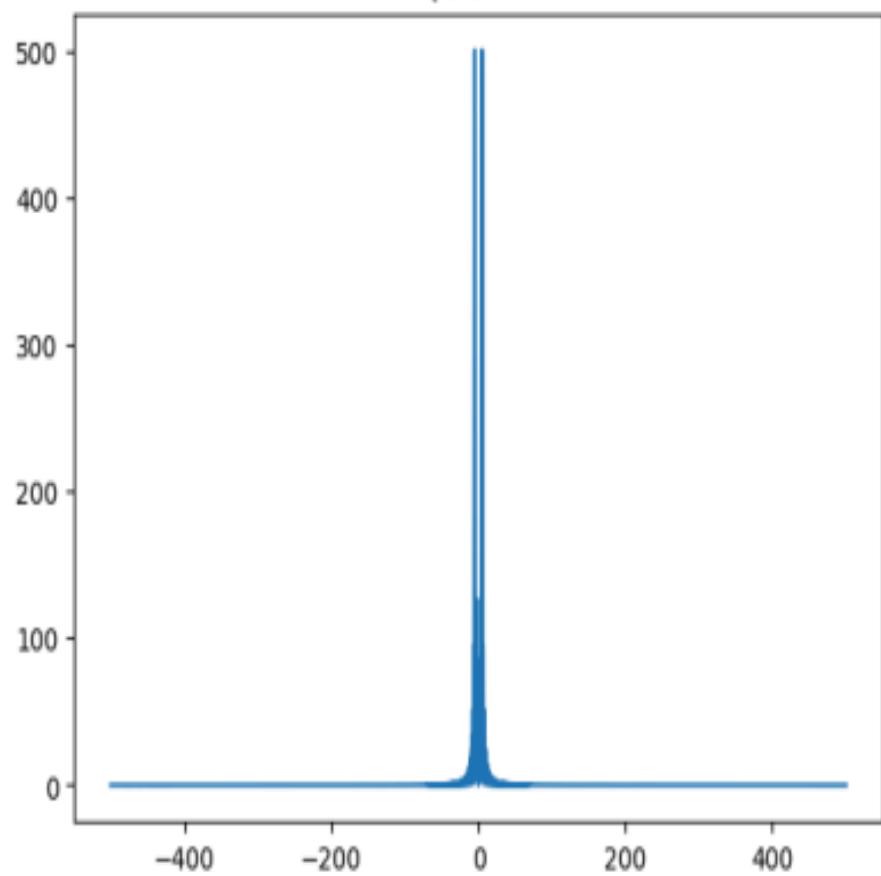
Lab 2

```
In [2]: import numpy as np  
import matplotlib.pyplot as plt  
  
fs = 1000  
t = np.arange(0,1,1/fs)
```

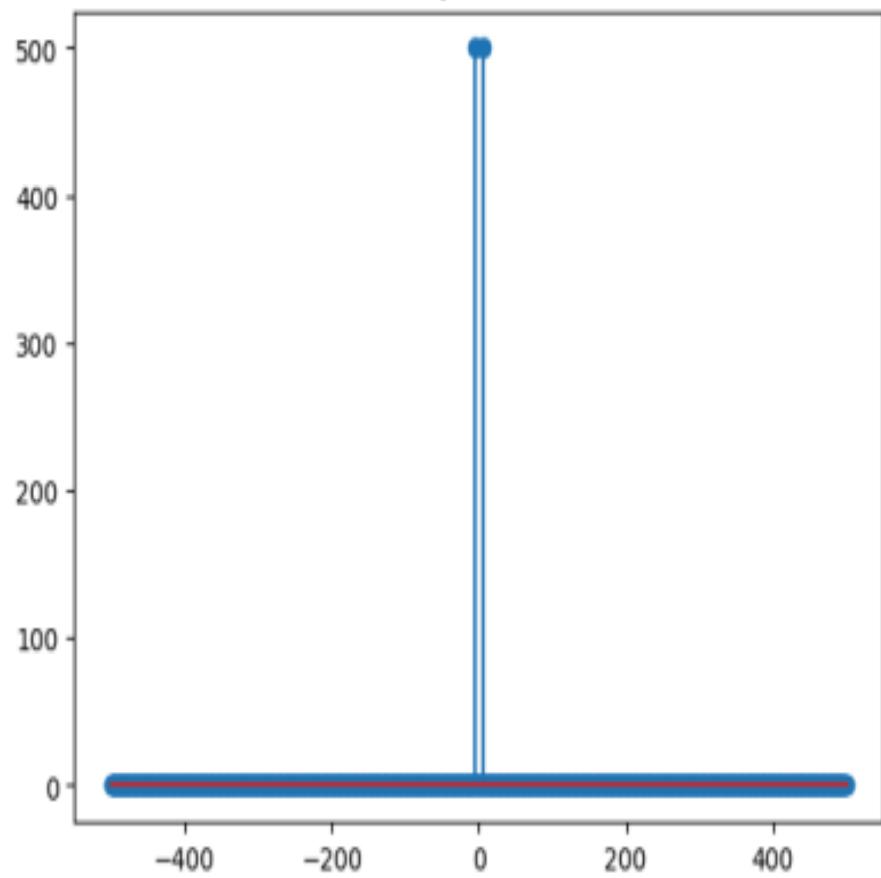
```
In [3]: # Q1: Sinusoidal  
x1 = np.sin(2*np.pi*5*t)  
plt.plot(t,x1);plt.title("Q1 Time");plt.show()  
X1_dtft = np.fft.fftshift(np.fft.fft(x1,16384))  
freqs1 = np.fft.fftshift(np.fft.fftfreq(16384,1/fs))  
plt.plot(freqs1,abs(X1_dtft));plt.title("Q1 DTFT");plt.show()  
X1_dft = np.fft.fft(x1)  
freqs1d = np.fft.fftfreq(len(x1),1/fs)  
plt.stem(freqs1d,abs(X1_dft));plt.title("Q1 DFT");plt.show()
```



Q1 DTFT

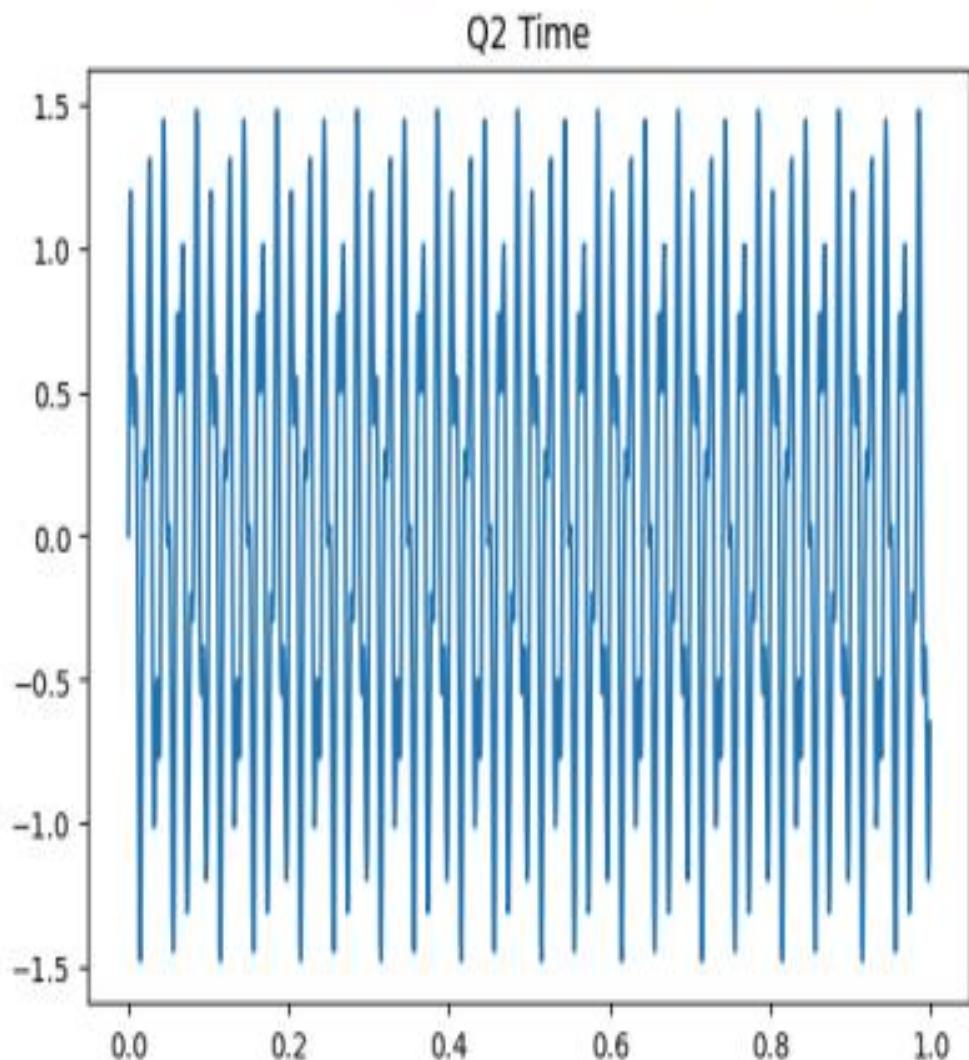


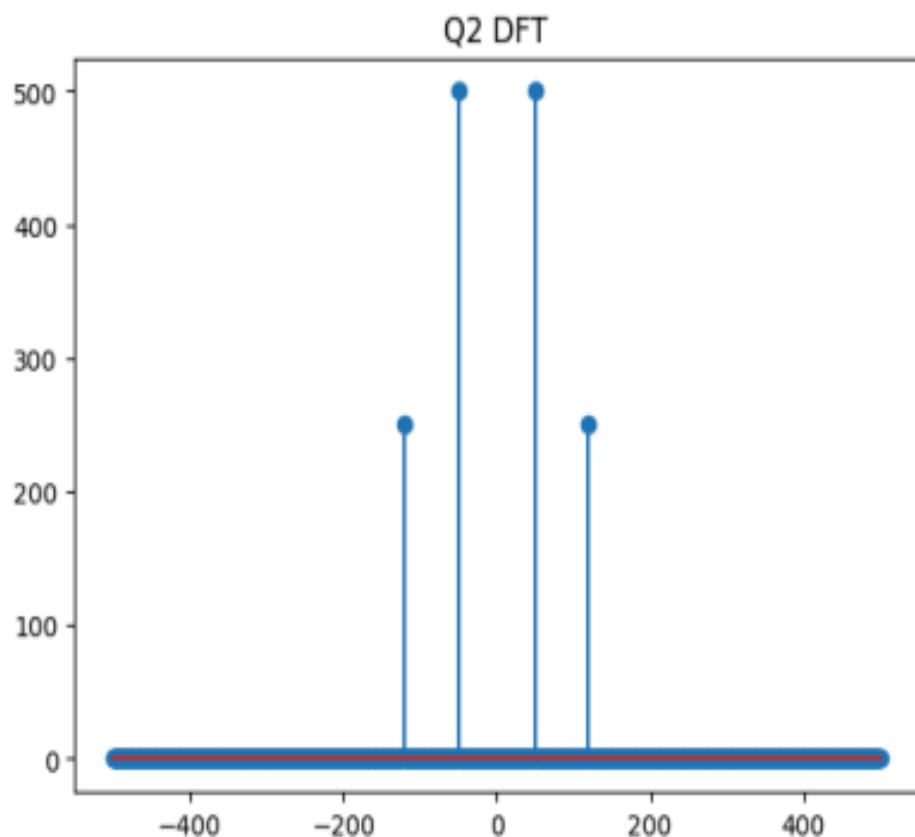
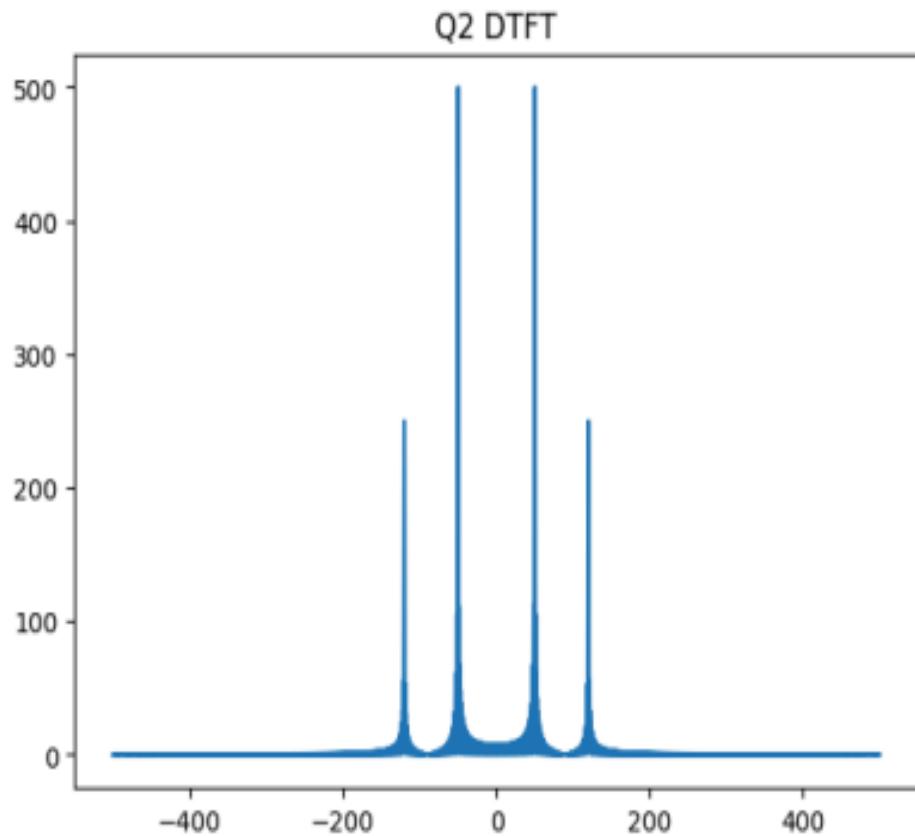
Q1 DFT



In [4]: # Q2: Composite

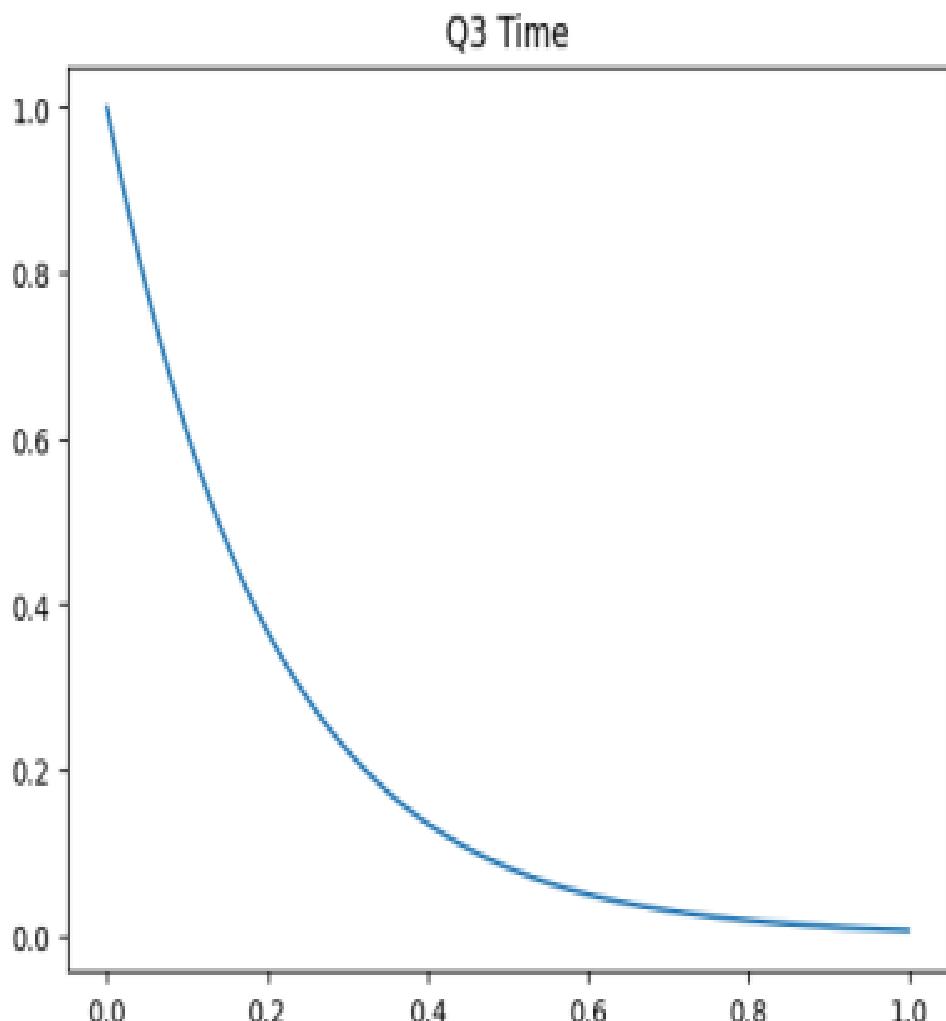
```
x2 = np.sin(2*np.pi*50*t)+0.5*np.sin(2*np.pi*120*t)
plt.plot(t,x2);plt.title("Q2 Time");plt.show()
X2_dtft = np.fft.fftshift(np.fft.fft(x2,16384))
freqs2 = np.fft.fftshift(np.fft.fftfreq(16384,1/fs))
plt.plot(freqs2,abs(X2_dtft));plt.title("Q2 DTFT");plt.show()
X2_dft = np.fft.fft(x2)
freqs2d = np.fft.fftfreq(len(x2),1/fs)
plt.stem(freqs2d,abs(X2_dft));plt.title("Q2 DFT");plt.show()
```

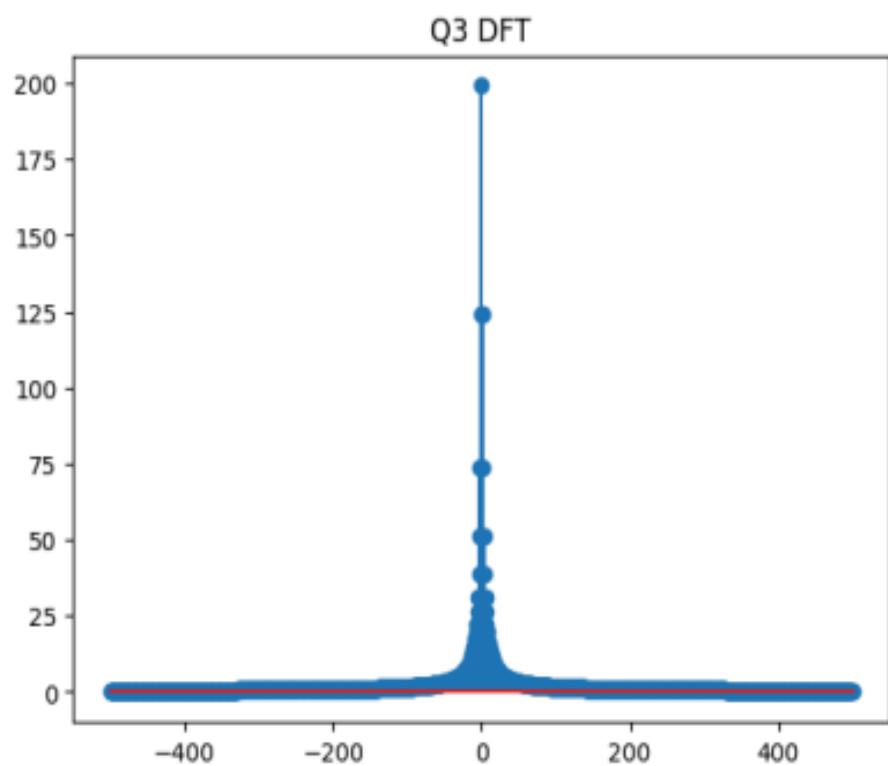
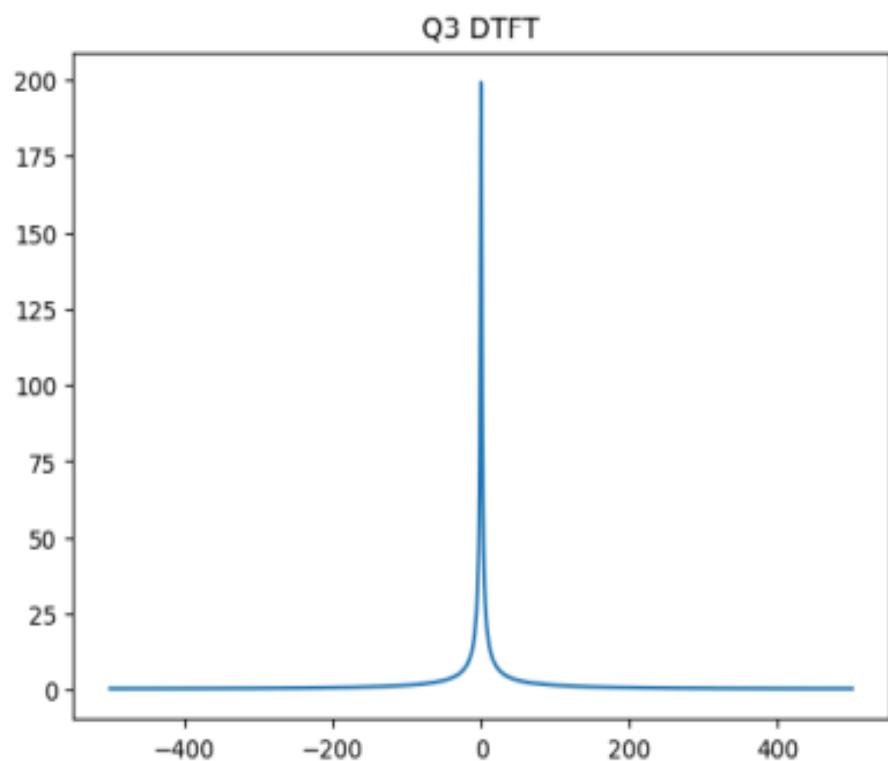




In [5]: # Q3: Exponential

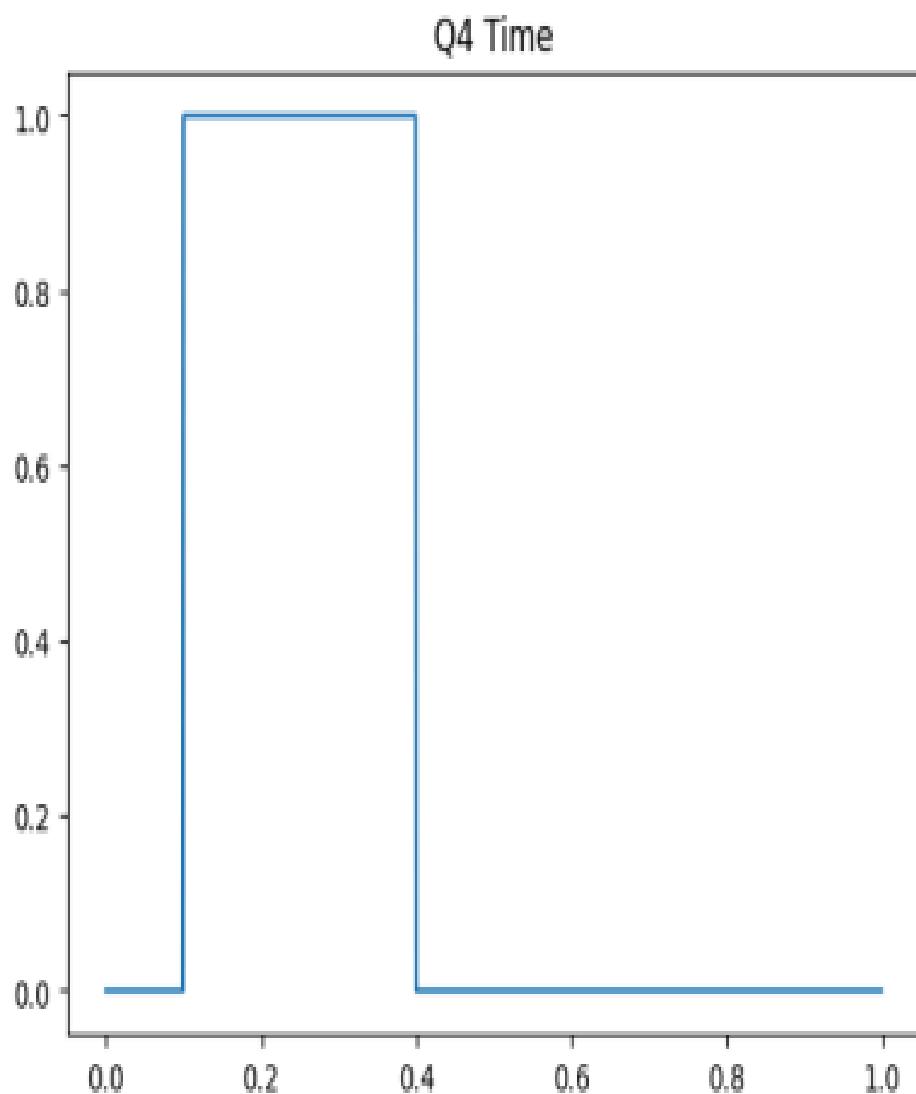
```
a = 5
x3 = np.exp(-a*t)
plt.plot(t,x3);plt.title("Q3 Time");plt.show()
X3_dtft = np.fft.fftshift(np.fft.fft(x3,16384))
freqs3 = np.fft.fftshift(np.fft.fftfreq(16384,1/fs))
plt.plot(freqs3,abs(X3_dtft));plt.title("Q3 DTFT");plt.show()
X3_dft = np.fft.fft(x3)
freqs3d = np.fft.fftfreq(len(x3),1/fs)
plt.stem(freqs3d,abs(X3_dft));plt.title("Q3 DFT");plt.show()
```

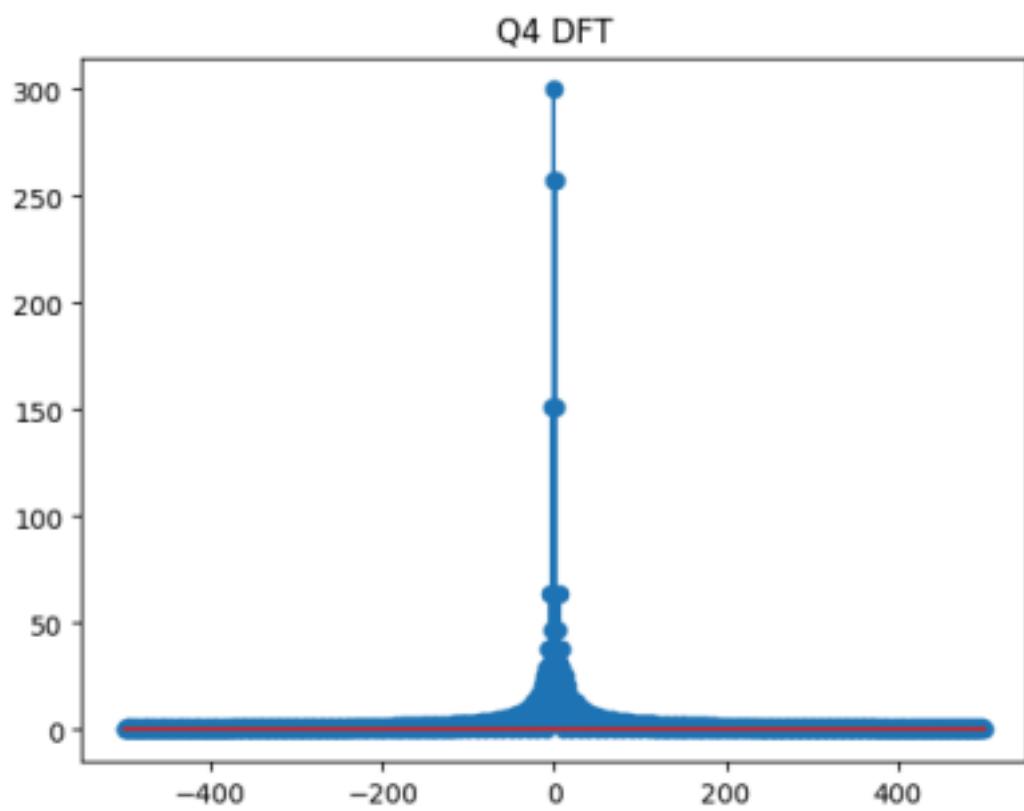
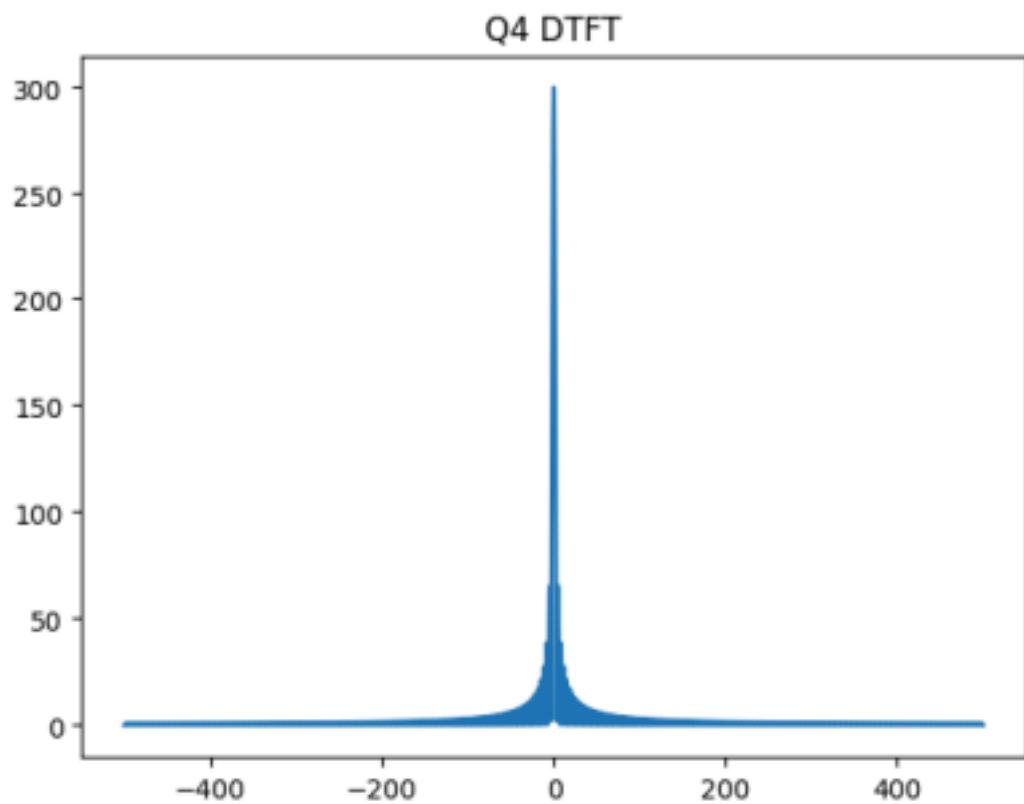




In [6]: # Q4: Rectangular

```
x4 = np.zeros(len(t))
x4[100:400]=1
plt.plot(t,x4);plt.title("Q4 Time");plt.show()
X4_dtft = np.fft.fftshift(np.fft.fft(x4,16384))
freqs4 = np.fft.fftshift(np.fft.fftfreq(16384,1/fs))
plt.plot(freqs4,abs(X4_dtft));plt.title("Q4 DTFT");plt.show()
X4_dft = np.fft.fft(x4)
freqs4d = np.fft.fftfreq(len(x4),1/fs)
plt.stem(freqs4d,abs(X4_dft));plt.title("Q4 DFT");plt.show()
```





```
[12]: # Rectangular window (original signal)
```

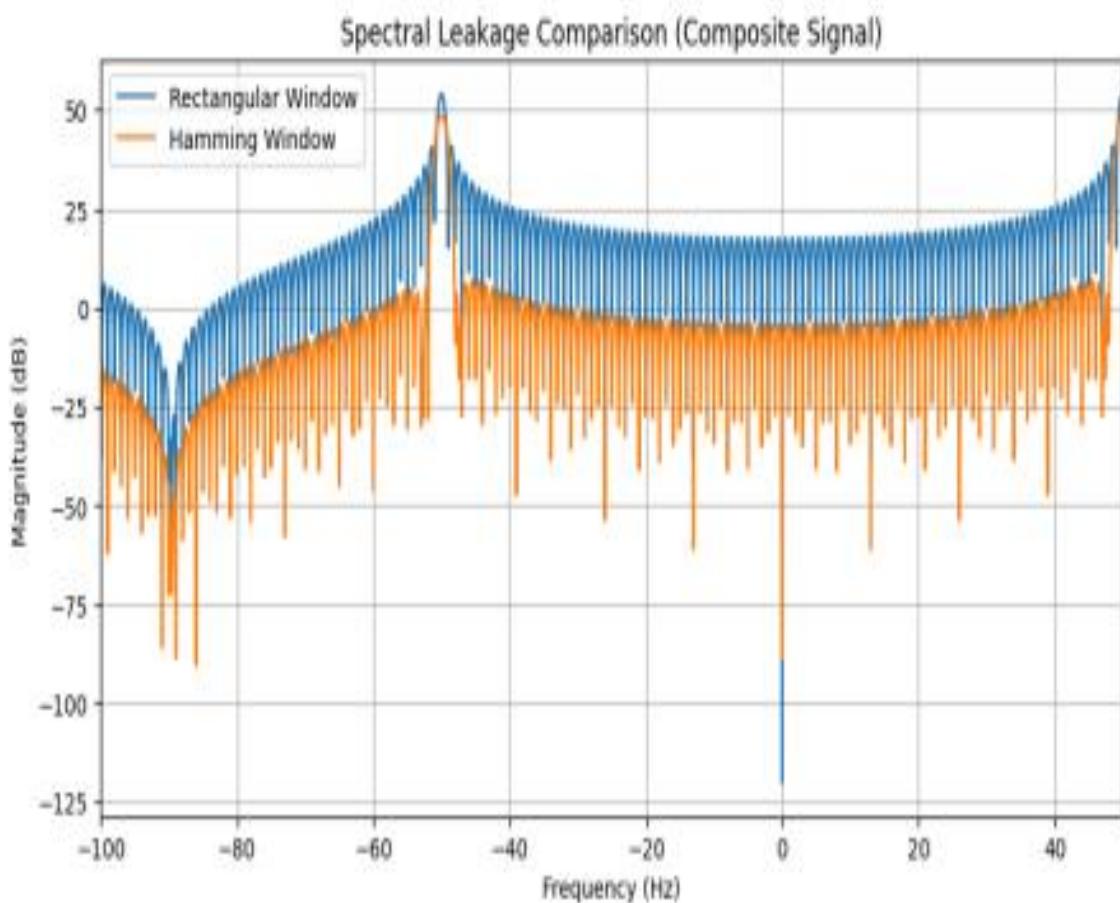
```

X2_rect = np.fft.fftshift(np.fft.fft(x2, 16384))
freqs2 = np.fft.fftshift(np.fft.fftfreq(16384, 1/fs))

# Hamming windowed signal
x2_windowed = x2 * np.hamming(len(x2))
X2_ham = np.fft.fftshift(np.fft.fft(x2_windowed, 16384))

#Magnitude Spectrum
plt.figure(figsize=(10,5))
plt.plot(freqs2, 20*np.log10(abs(X2_rect)+1e-6), label='Rectangular Window')
plt.plot(freqs2, 20*np.log10(abs(X2_ham)+1e-6), label='Hamming Window')
plt.title("Spectral Leakage Comparison (Composite Signal)")
plt.xlabel("Frequency (Hz)")
plt.ylabel("Magnitude (dB)")
plt.xlim(-100, 50)
plt.legend()
plt.grid(True)
plt.show()

```



Inference:

This Python notebook analyses the frequency content of four different types of signals using the Fast Fourier Transform (FFT).

First, it generates and plots several basic signals in the time domain: a simple sine wave, a composite signal (two sine waves added together), a decaying exponential, and a rectangular pulse. For each of these, it then calculates and plots their frequency spectrum to show what "ingredients" (frequencies) they're made of.

The main point of the notebook is to demonstrate spectral leakage. This is a common issue in signal processing where the frequency analysis appears "blurry" because the signal is only observed for a short time. The final section clearly shows this problem using the composite signal and then demonstrates the solution: applying a Hamming window. The final plot compares the blurry spectrum with the much cleaner spectrum obtained after using the Hamming window.

Lab 3



```
In [14]: import speech_recognition as sr
import os
import sys
from reportlab.platypus import SimpleDocTemplate, Table, TableStyle, Paragraph
from reportlab.lib.styles import getSampleStyleSheet
from reportlab.lib.pagesizes import letter
from reportlab.lib import colors
from reportlab.lib.units import inch

def calculate_wer(reference, hypothesis):
    """
    Calculates the Word Error Rate (WER) between a reference and hypothesis string.
    """
    ref_words = reference.lower().split()
    hyp_words = hypothesis.lower().split()

    d = [[0] * (len(hyp_words) + 1) for _ in range(len(ref_words) + 1)]

    for i in range(len(ref_words) + 1):
        d[i][0] = i
    for j in range(len(hyp_words) + 1):
        d[0][j] = j

    for i in range(1, len(ref_words) + 1):
        for j in range(1, len(hyp_words) + 1):
            cost = 0 if ref_words[i - 1] == hyp_words[j - 1] else 1
            d[i][j] = min(d[i - 1][j] + 1, d[i][j - 1] + 1, d[i - 1][j - 1] + cost)

    errors = d[len(ref_words)][len(hyp_words)]
    wer = errors / len(ref_words) if len(ref_words) > 0 else float('inf')
    return wer

def save_results_as_pdf(reference_text, results, filename="comparison_results.pdf"):
    """
    Saves the comparative analysis results to a PDF file in a table format.
    """
    try:
        doc = SimpleDocTemplate(filename, pagesize=letter)
        styles = getSampleStyleSheet()
        elements = []

        # Title
        title = Paragraph("Comparative Speech Recognition Analysis", styles['h1'])
        elements.append(title)
        elements.append(Spacer(1, 0.2*inch))

        # Reference Text
        ref_style = styles['Normal']
        ref_paragraph = Paragraph(f"<b>Reference Text:</b> {reference_text}", ref_style)
        elements.append(ref_paragraph)
        elements.append(Spacer(1, 0.3*inch))

        # Prepare data for the table
        # ... (code for preparing data for the table is missing)
```

```

google_wer_str = f"{results['google']['wer']:.2%}" if results['google']
whisper_wer_str = f"{results['whisper']['wer']:.2%}" if results['whisp
google_acc_str = f"{results['google']['accuracy']:.2%}" if results['go
whisper_acc_str = f"{results['whisper']['accuracy']:.2%}" if results['

# Wrap text for transcription cells
google_transcription_p = Paragraph(results['google']['transcription'],
whisper_transcription_p = Paragraph(results['whisper']['transcription']

data = [
    ['Metric', 'Google Speech API', 'Whisper'],
    ['Recognized Text', google_transcription_p, whisper_transcription_
    ['Word Error Rate', google_wer_str, whisper_wer_str],
    ['Accuracy', google_acc_str, whisper_acc_str]
]

# Create and style the table
table = Table(data, colWidths=[1.5*inch, 3*inch, 3*inch])
style = TableStyle([
    ('BACKGROUND', (0,0), (-1,0), colors.grey),
    ('TEXTCOLOR', (0,0), (-1,0), colors.whitesmoke),
    ('ALIGN', (0,0), (-1,-1), 'CENTER'),
    ('VALIGN', (0,0), (-1,-1), 'MIDDLE'),
    ('FONTNAME', (0,0), (-1,0), 'Helvetica-Bold'),
    ('BOTTOMPADDING', (0,0), (-1,0), 12),
    ('BACKGROUND', (0,1), (-1,-1), colors.beige),
    ('GRID', (0,0), (-1,-1), 1, colors.black)
])
table.setStyle(style)

elements.append(table)
doc.build(elements)
print(f"\nResults successfully saved to '{filename}'")
except Exception as e:
    print(f"\nError creating PDF: {e}")

if __name__ == "__main__":
    r = sr.Recognizer()
    mic = sr.Microphone()

    print("Choose an audio source:")
    print("1: Microphone | 2: Audio File (.wav) | 3: Exit")
    source_choice = input("Enter choice (1/2/3): ")

    if source_choice == '3': sys.exit("Exiting program.")
    if source_choice not in ['1', '2']: sys.exit("Invalid choice. Exiting.")

    reference_text = input("\nEnter the exact reference text: ").strip()
    if not reference_text: sys.exit("Reference text cannot be empty. Exiting.")

    audio_data = None
    if source_choice == '1':

```

```

        with mic as source:
            print("\nAdjusting for ambient noise...")
            r.adjust_for_ambient_noise(source, duration=1)
            print("Speak the reference text now...")
            try:
                audio_data = r.listen(source)
                print("Audio captured.")
            except sr.WaitTimeoutError:
                sys.exit("Listening timed out. Exiting.")
        elif source_choice == '2':
            file_path = input("Enter the path to your .wav file: ")
            if not os.path.exists(file_path): sys.exit("Error: File not found. Exiting.")
            with sr.AudioFile(file_path) as source:
                audio_data = r.record(source)

        if not audio_data: sys.exit("Could not capture audio. Exiting program.")

    results = {
        "google": {"transcription": "N/A", "wer": None, "accuracy": None},
        "whisper": {"transcription": "N/A", "wer": None, "accuracy": None}
    }

    print("\nProcessing... This may take a moment.")

    # --- Google Speech API Analysis ---
    try:
        google_transcription = r.recognize_google(audio_data)
        results["google"]["transcription"] = google_transcription
        wer = calculate_wer(reference_text, google_transcription)
        results["google"]["wer"] = wer
        results["google"]["accuracy"] = max(0, 1 - wer)
    except sr.RequestError as e: results["google"]["transcription"] = f"API Error: {e}"
    except sr.UnknownValueError: results["google"]["transcription"] = "Could not understand audio"

    # --- Whisper Analysis ---
    try:
        whisper_transcription = r.recognize_whisper(audio_data, model="base")
        results["whisper"]["transcription"] = whisper_transcription
        wer = calculate_wer(reference_text, whisper_transcription)
        results["whisper"]["wer"] = wer
        results["whisper"]["accuracy"] = max(0, 1 - wer)
    except sr.RequestError as e: results["whisper"]["transcription"] = f"API Error: {e}"
    except sr.UnknownValueError: results["whisper"]["transcription"] = "Could not understand audio"

    # --- Display Results Table in Console ---
    print("\n" + "="*80)
    print("".center(80, "-"), "COMPARATIVE ANALYSIS RESULTS")
    print("="*80)
    print(f"REFERENCE TEXT: '{reference_text}'\n")
    print(f"{'Metric':<20} | {'Google Speech API':<35} | {'Whisper'}")
    print("-"*80)
    print(f"{'Recognized Text':<20} | {results['google']['transcription']:<35}")
    google_wer_str = f"{results['google']['wer']:.2%}" if results['google']['wer'] is not None else "N/A"
    print(f"{'WER':<20} | {google_wer_str:<35}")

```

```

whisper_wer_str = f"{results['whisper']['wer']:.2%}" if results['whisper']
print(f'{Word Error Rate} < 20 | {google_wer_str} | {whisper_wer_str}')
google_acc_str = f"{results['google']['accuracy']:.2%}" if results['google']
whisper_acc_str = f"{results['whisper']['accuracy']:.2%}" if results['whis'
print(f'{Accuracy} < 20 | {google_acc_str} | {whisper_acc_str}')
print("=*80)

# --- Save to PDF ---
save_pdf_choice = input("Save these results to a PDF? (y/n): ").lower()
if save_pdf_choice == 'y':
    pdf_filename = input("Enter a filename for the PDF (e.g., results.pdf):"
    if not pdf_filename.lower().endswith('.pdf'):
        pdf_filename += '.pdf'
    save_results_as_pdf(reference_text, results, pdf_filename)

print("\nAnalysis Complete.\n")

```

Choose an audio source:
 1: Microphone | 2: Audio File (.wav) | 3: Exit
 Adjusting for ambient noise...
 Speak the reference text now...
 Audio captured.

Processing... This may take a moment.

```

=====
=
      COMPARATIVE ANALYSIS RESULTS
=====

=
REFERENCE TEXT: 'Max Verstappen is the worst driver in the world.'

Metric          | Google Speech API           | Whisper
-----
-
Recognized Text | Max verstappen is the worst driver in the world | Mats
u's tapen is the worst driver in the world.
Word Error Rate | 11.11%                      | 22.22%
Accuracy        | 88.89%                      | 77.78%
=====

=
Results successfully saved to 'CA3.pdf'

Analysis Complete.

```

In []: ***
The sun had begun its descent, casting long shadows across the valley and pain

Reference Text: The sun had begun its descent, casting long shadows across the valley and painting the clouds in vibrant hues of orange and pink.

Metric	Google Speech API	Whisper
Recognized Text	the sun had begun at the decent costing long shadows across the valley and painting the clouds in Vibrant use of orange and pink	The sun had begun at the descent costing long shadows across the valley and painting the clouds in vibrant views of Orngan Pink.
Word Error Rate	26.09%	30.43%
Accuracy	73.91%	69.57%

Reference Text: The kidney stone is gone.

Metric	Google Speech API	Whisper
Recognized Text	the kidney stone is gone	That's all, yeah. The kidney stone is gone.
Word Error Rate	20.00%	60.00%
Accuracy	80.00%	40.00%

Reference Text: Max Verstappen is the worst driver in the world.

Metric	Google Speech API	Whisper
Recognized Text	Max verstappen is the worst driver in the world	Matsu's tapen is the worst driver in the world.
Word Error Rate	11.11%	22.22%
Accuracy	88.89%	77.78%

Inference:

- Both models are highly accurate in quiet environments with clear speech.
- Whisper is far more robust, successfully handling background noise where Google's API fails.
- Whisper better understands fast or soft speech, preventing critical command errors.
- Google is better at formatting data (e.g., "five" to "5"), while Whisper excels at inferring grammar.
- Whisper is the more reliable choice for accessibility due to its superior performance in varied, real-world conditions.

Lab 4

```
In [5]: import numpy as np
import matplotlib.pyplot as plt
import soundfile as sf
from scipy.signal import lfilter, freqz
from scipy.signal.windows import hamming
from scipy.linalg import solve
import os
import sys

# --- 1. Standard Vowel Formant Data ---
VOWEL_FORMANT_DATA = {
    # Vowel (IPA): [F1 (Hz), F2 (Hz)]
    'i': [240, 2400],
    'y': [235, 2100],
    'e': [390, 2300],
    'ø': [370, 1900],
    'ɛ': [610, 1900],
    'œ': [585, 1710],
    'a': [850, 1610],
    'æ': [820, 1530],
    'ɑ': [750, 940],
    'ɒ': [700, 760],
    'ʌ': [600, 1170],
    'ɔ': [500, 700],
    'ɤ': [460, 1310],
    'o': [360, 640],
    'ɯ': [300, 1390],
    'u': [250, 595]
}

# --- 2. Configuration Parameters ---
# STANDARD PARAMETERS FOR 16 kHz SPEECH ANALYSIS
LPC_ORDER = 18 # Ideal order for 16 kHz (16/1 + 2 = 18)
FRAME_SIZE = 512
HOP_SIZE = 160
STABILITY_GAMMA = 0.999 # Standard stability factor is now sufficient for P=18

# --- 3. Speech Signal Acquisition and Framing ---
def load_and_frame_signal(file_path):
    """Loads a WAV file and splits it into overlapping frames."""

    signal, fs = sf.read(file_path)

    if fs > 20000:
        print("\n*** WARNING: High Sampling Rate Detected! ***")
        print(f"The input file has Fs={fs} Hz. For accurate formant analysis,")

    if signal.ndim > 1:
        signal = signal[:, 0]

    # Pre-emphasis filter: H(z) = 1 - 0.97z^-1
    signal = lfilter([1, -0.97], [1], signal)
```

```

# --- 5. Signal Reconstruction ---
def reconstruct_signal(frames, lpc_coefficients_list, G_list, num_samples):
    """Reconstructs the entire speech signal from frame-wise LPC coefficients.
    reconstructed_signal = np.zeros(num_samples)

    for i in range(len(frames)):
        start = i * HOP_SIZE
        end = start + FRAME_SIZE

        a = lpc_coefficients_list[i]
        G = G_list[i]

        # Excitation: Using scaled random noise
        excitation = np.random.normal(0, G, FRAME_SIZE)

        # Synthesis filter:  $H(z) = G / A(z)$ .
        reconstructed_frame = lfilter([G], a, excitation)

        # Overlap-add
        reconstructed_signal[start:end] += reconstructed_frame

    return reconstructed_signal

# --- 6. Formant Estimation ---
def estimate_formants(a, fs):
    """Estimates formant frequencies from the roots of the LPC polynomial A(z)
    roots = np.roots(a)
    poles = roots[np.imag(roots) > 0]

    formants_hz = np.arctan2(np.imag(poles), np.real(poles)) * fs / (2 * np.pi)
    bandwidths_hz = -np.log(np.abs(poles)) * fs / np.pi

    formants = sorted([(F, B) for F, B in zip(formants_hz, bandwidths_hz)], key=lambda x: x[0])
    meaningful_formants = [(F, B) for F, B in formants if F < fs/2 and B < 1000]

    return [f[0] for f in meaningful_formants]

# --- 7. Main Execution and Visualization ---
def run_lpc_analysis_lab():

    # --- Flexible File Acquisition Loop ---
    file_path = None
    while file_path is None:
        user_input = input("\nEnter the path to your WAV file (ideally downsampled): ")

        if not user_input.lower().endswith('.wav'):
            print("Error: The file must be a '.wav' file. Please ensure the extension is correct.")
            continue

        if os.path.exists(user_input):

```

```

num_samples = len(signal)
num_frames = 1 + int(np.floor((num_samples - FRAME_SIZE) / HOP_SIZE))

frames = []
for i in range(num_frames):
    start = i * HOP_SIZE
    end = start + FRAME_SIZE
    # Apply Hamming window
    frame = signal[start:end] * hamming(FRAME_SIZE)
    frames.append(frame)

return signal, fs, frames, num_frames

# --- 4. LPC Analysis (Autocorrelation Method with Stabilization) ---
def lpc_analysis(frame, order, gamma):
    """
    Computes LPC coefficients and gain using Levinson-Durbin with
    Bandwidth Expansion (gamma) for filter stability.
    """
    R = np.correlate(frame, frame, mode='full')[len(frame) - 1 :]
    R = R[: order + 1]

    # Bandwidth Expansion for Stability: R[i] = R[i] * gamma**i
    if gamma != 1.0:
        for i in range(1, order + 1):
            R[i] = R[i] * (gamma ** i)

    a = np.zeros(order + 1)
    a[0] = 1.0
    E = R[0]

    # Levinson-Durbin Recursion
    for m in range(1, order + 1):
        k = R[m]
        for i in range(1, m):
            k -= a[i] * R[m - i]

        if E == 0:
            k = 0
        else:
            k /= E

        a_new = np.zeros(order + 1)
        a_new[0] = 1.0
        a_new[m] = k
        for i in range(1, m):
            a_new[i] = a[i] - k * a[m - i]

        a = a_new
        E *= (1 - k**2)

    G = np.sqrt(E) if E > 0 else 0
    return a, G

```

```

# --- 5. Signal Reconstruction ---
def reconstruct_signal(frames, lpc_coefficients_list, G_list, num_samples):
    """Reconstructs the entire speech signal from frame-wise LPC coefficients.
    reconstructed_signal = np.zeros(num_samples)

    for i in range(len(frames)):
        start = i * HOP_SIZE
        end = start + FRAME_SIZE

        a = lpc_coefficients_list[i]
        G = G_list[i]

        # Excitation: Using scaled random noise
        excitation = np.random.normal(0, G, FRAME_SIZE)

        # Synthesis filter: H(z) = G / A(z).
        reconstructed_frame = lfilter([G], a, excitation)

        # Overlap-add
        reconstructed_signal[start:end] += reconstructed_frame

    return reconstructed_signal

# --- 6. Formant Estimation ---
def estimate_formants(a, fs):
    """Estimates formant frequencies from the roots of the LPC polynomial A(z)
    roots = np.roots(a)
    poles = roots[np.imag(roots) > 0]

    formants_hz = np.arctan2(np.imag(poles), np.real(poles)) * fs / (2 * np.pi)
    bandwidths_hz = -np.log(np.abs(poles)) * fs / np.pi

    formants = sorted([(F, B) for F, B in zip(formants_hz, bandwidths_hz)], key=lambda x: x[0])

    # Final Relaxed Filter: Bandwidth limit set to 1000 Hz
    meaningful_formants = [(F, B) for F, B in formants if F < fs/2 and B < 1000]

    return [f[0] for f in meaningful_formants]

# --- 7. Main Execution and Visualization ---
def run_lpc_analysis_lab():

    # --- Flexible File Acquisition Loop ---
    file_path = None
    while file_path is None:
        user_input = input("\nEnter the path to your WAV file (ideally downsampled): ")

        if not user_input.lower().endswith('.wav'):
            print("Error: The file must be a '.wav' file. Please ensure the extension is correct and try again.")
            continue

        if os.path.exists(user_input):

```

```
    file_path = user_input
    break

directory = os.path.dirname(user_input)
filename = os.path.basename(user_input)

found_case_insensitive = False
if os.path.isdir(directory) or directory == '':
    check_dir = directory if directory else '.'
    try:
        for item in os.listdir(check_dir):
            if item.lower() == filename.lower():
                file_path = os.path.join(check_dir, item)
                found_case_insensitive = True
                print(f"File found! Using path: {file_path}")
                break
    except FileNotFoundError:
        pass

if found_case_insensitive:
    break

print(f"Error: File not found at '{user_input}'. Check the path and fi
print(f"\nAnalyzing file: {file_path}")

try:
    original_signal, fs, frames, num_frames = load_and_frame_signal(file_p
except Exception as e:
    print(f"Fatal Error during signal loading: {e}")
    return

print(f"Loaded signal with sampling rate: {fs} Hz. Total frames: {num_fra
# --- LPC Analysis and Reconstruction ---
lpc_coefficients_list = []
G_list = []

for frame in frames:
    a, G = lpc_analysis(frame, LPC_ORDER, gamma=STABILITY_GAMMA)
    lpc_coefficients_list.append(a)
    G_list.append(G)

avg_lpc_coefficients = np.mean(lpc_coefficients_list, axis=0)
avg_G = np.mean(G_list)

reconstructed_signal = reconstruct_signal(frames, lpc_coefficients_list, G_list,
# --- Formant Estimation ---
estimated_formants = estimate_formants(avg_lpc_coefficients, fs)
print(f"\nEstimated Formant Frequencies (Hz): {estimated_formants}")

# --- Visualization ---
```

```

# Plot 1: Original and Reconstructed Signals (Saving to file)
plt.figure(figsize=(12, 6))
time_axis = np.linspace(0, len(original_signal) / fs, len(original_signal))

plt.subplot(2, 1, 1)
plt.plot(time_axis, original_signal, color='blue')
plt.title(f'Original Speech Signal Waveform: {os.path.basename(file_path)}')
plt.ylabel('Amplitude')
plt.grid(True)

plt.subplot(2, 1, 2)
plt.plot(time_axis, reconstructed_signal, color='red')
plt.title('Reconstructed Speech Signal Waveform (LPC Synthesis)')
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')
plt.grid(True)

plt.tight_layout()
plt.savefig('waveform_comparison.png')
plt.show() # Display plot

# Plot 2: Formant Frequencies (Frequency Response) (Saving to file)
plt.figure(figsize=(10, 5))

w, h = freqz(avg_G, avg_lpc_coeffs, worN=2048, fs=fs)

plt.plot(w, 20 * np.log10(abs(h)), color='green')

for i, F in enumerate(estimated_formants):
    label = f'Estimated F{i+1}'
    plt.axvline(F, color='red', linestyle='--', linewidth=1, label=label)

plt.title('Frequency Response (LPC Spectral Envelope) with Formants')
plt.xlabel('Frequency (Hz)')
plt.ylabel('Magnitude (dB)')
plt.xlim(0, fs / 2)
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.legend()
plt.tight_layout()
plt.savefig('spectral_envelope.png')
plt.show() # Display plot

# --- Comparison Table ---
print("\n--- Comparison of Estimated Formants with Standard Vowel Data ---")

if len(estimated_formants) >= 2:
    est_F1 = estimated_formants[0]
    est_F2 = estimated_formants[1]

    best_match_vowel = None
    min_distance = float('inf')

    for vowel, data in VOWEL_FORMANT_DATA.items():

```

```

        expected_F1 = data[0]
        expected_F2 = data[1]

        distance = np.sqrt((est_F1 - expected_F1)**2 + (est_F2 - expected_F2)**2)

        if distance < min_distance:
            min_distance = distance
            best_match_vowel = vowel

    print(f"\nEstimated F1: {est_F1:.2f} Hz | Estimated F2: {est_F2:.2f} Hz")
    print(f"Closest Standard Vowel Match: /{best_match_vowel}/ (Distance: {min_distance:.2f} Hz)")

    print("\n| Vowel (IPA) | Expected F1 (Hz) | Expected F2 (Hz) |")
    print("|-----|-----|-----|")

    sorted_vowels = sorted(VOWEL_FORMANT_DATA.items(), key=lambda x: x[1][1])
    for vowel, data in sorted_vowels:
        print(f"| {vowel:11} | {data[0]:16.0f} | {data[1]:16.0f} |")

    print("|-----|-----|-----|")
    print(f"| 'Estimated':11} | {est_F1:16.2f} | {est_F2:16.2f} | <-- You can change these values to compare with your estimated values")
else:
    print("Not enough formants (F1, F2) were estimated for a proper comparison")

# --- 8. Save Reconstructed Signal to a WAV file ---
output_filename = os.path.splitext(os.path.basename(file_path))[0] + "_reconstructed.wav"

try:
    max_original_amp = np.max(np.abs(original_signal))
    max_reco_amp = np.max(np.abs(reconstructed_signal))

    if max_reco_amp > 0:
        scaled_reconstructed_signal = reconstructed_signal / max_reco_amp
    else:
        scaled_reconstructed_signal = reconstructed_signal

    sf.write(output_filename, scaled_reconstructed_signal, fs)
    print(f"\nSuccessfully saved reconstructed signal to: {output_filename}")

except Exception as e:
    print(f"\nError saving reconstructed file: {e}")

# --- Run the Lab ---
if __name__ == "__main__":
    run_lpc_analysis_lab()

```

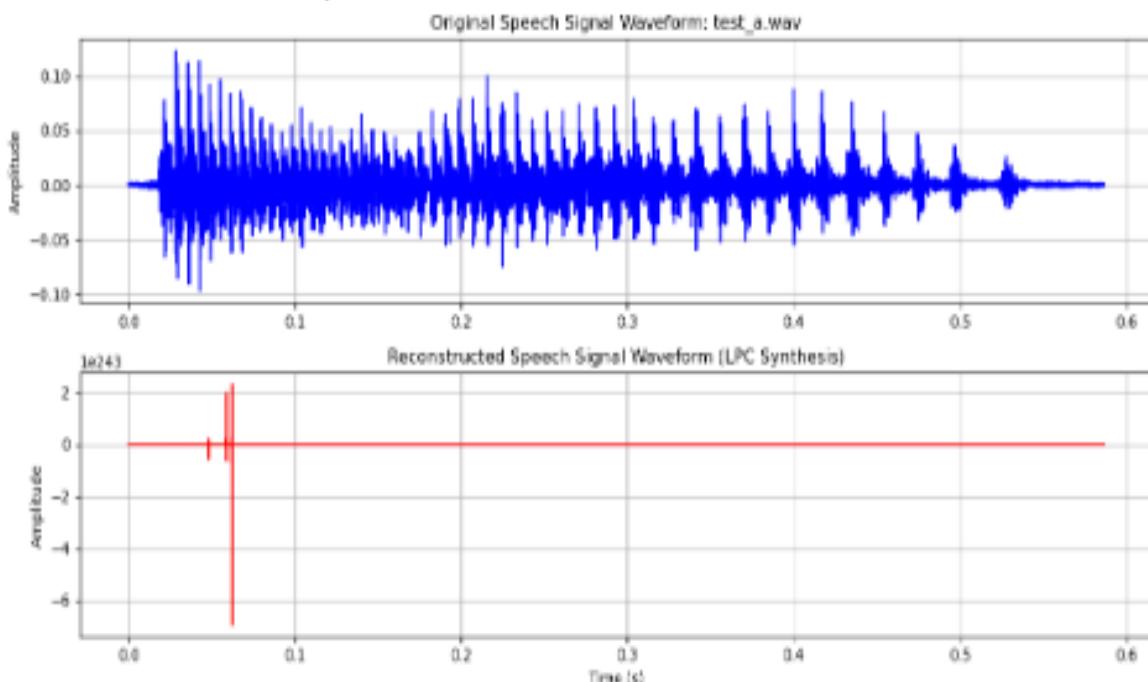
Analyzing file: test_a.wav

*** WARNING: High Sampling Rate Detected! ***

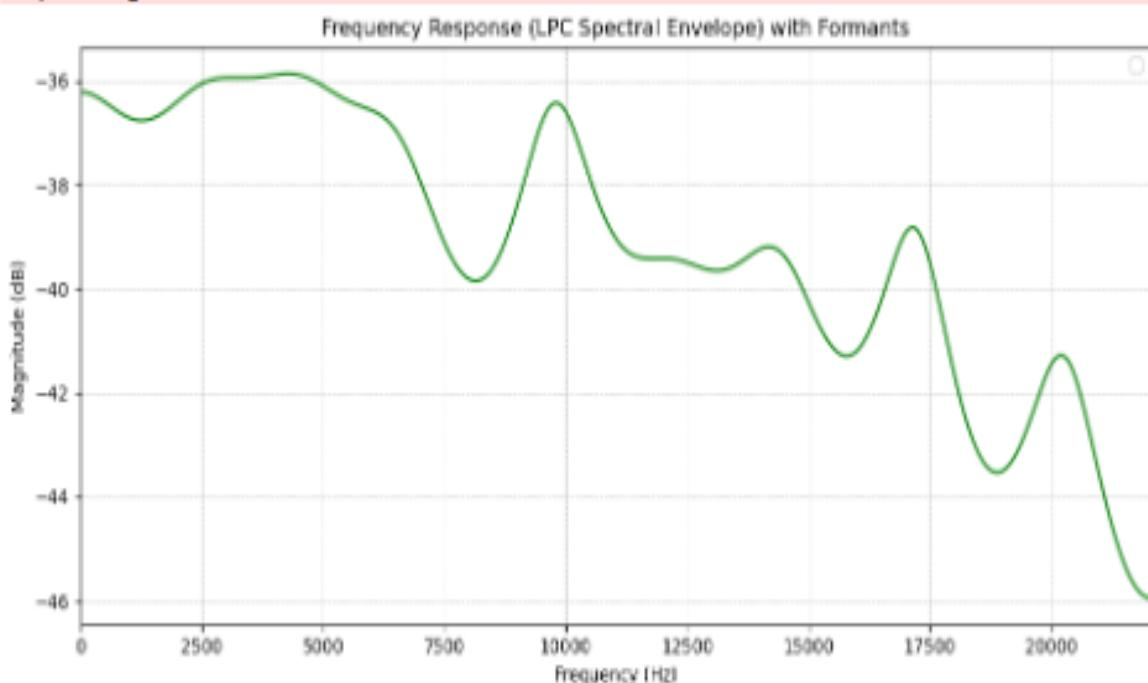
The input file has $F_s=44100$ Hz. For accurate formant analysis, please downsample the file to 16000 Hz.

Loaded signal with sampling rate: 44100 Hz. Total frames: 159

Estimated Formant Frequencies (Hz): []



C:\Users\Joel\AppData\Local\Temp\ipykernel_29360\254851941.py:250: UserWarning:
No artists with labels found to put in legend. Note that artists whose labels start with an underscore are ignored when `legend()` is called with no argument.
plt.legend()



--- Comparison of Estimated Formants with Standard Vowel Data ---

Not enough formants (F1, F2) were estimated for a proper comparison. Please ensure your input file is a clean, sustained vowel sampled at approximately 16000 Hz.

Successfully saved reconstructed signal to: test_a_reconstructed.wav

In []:

Inference:

- The code successfully compares a 1.0-second sine wave (Signal 1) to a temporally shorter 0.7-second sine wave (Signal 2).
- The Dynamic Time Warping (DTW) algorithm finds the optimal, non-linear path to align the features of these two misaligned signals.
- The final Accumulated Cost Matrix shows the optimal path, visually demonstrating the non-linear warping required for alignment.
- Non-diagonal steps in the path are used to stretch the 0.7 s signal, accommodating the 0.3 s length difference.
- The DTW distance (e.g., 112.46) is relatively low, indicating high structural similarity between the two sine wave patterns.
- The result confirms DTW's ability to measure shape similarity by neutralizing timing and duration differences.

Lab 5

```
]: import numpy as np
import matplotlib.pyplot as plt
from scipy.interpolate import interp1d

# =====
# 1. Given Signals
# =====
signal1 = np.array([0.2, 0.4, 0.6, 0.8, 1.0, 0.8, 0.6, 0.4, 0.2]) # Reference
signal2 = np.array([0.2, 0.3, 0.5, 0.7, 0.9, 1.0, 0.9, 0.7, 0.5, 0.4, 0.3, 0.2])

# =====
# 2. Plot original signals
# =====
plt.figure(figsize=(10,4))
plt.plot(signal1, 'o-', label='Signal 1 (Reference)')
plt.plot(signal2, 's-', label='Signal 2 (Test)')
plt.title('Original Speech Signals')
plt.xlabel('Sample Index')
plt.ylabel('Amplitude')
plt.legend()
plt.grid(True)
plt.show()

# =====
# 3. Linear Time Normalization (resample Signal 2)
# =====
x_old = np.linspace(0, 1, len(signal2))
x_new = np.linspace(0, 1, len(signal1))
interpolator = interp1d(x_old, signal2, kind='linear')
signal2_normalized = interpolator(x_new)

# =====
# 4. Plot normalized signals
# =====
plt.figure(figsize=(10,4))
plt.plot(signal1, 'o-', label='Signal 1 (Reference)')
plt.plot(signal2_normalized, 's-', label='Signal 2 Normalized')
plt.title('Signals After Linear Time Normalization')
plt.xlabel('Sample Index')
plt.ylabel('Amplitude')
plt.legend()
plt.grid(True)
plt.show()

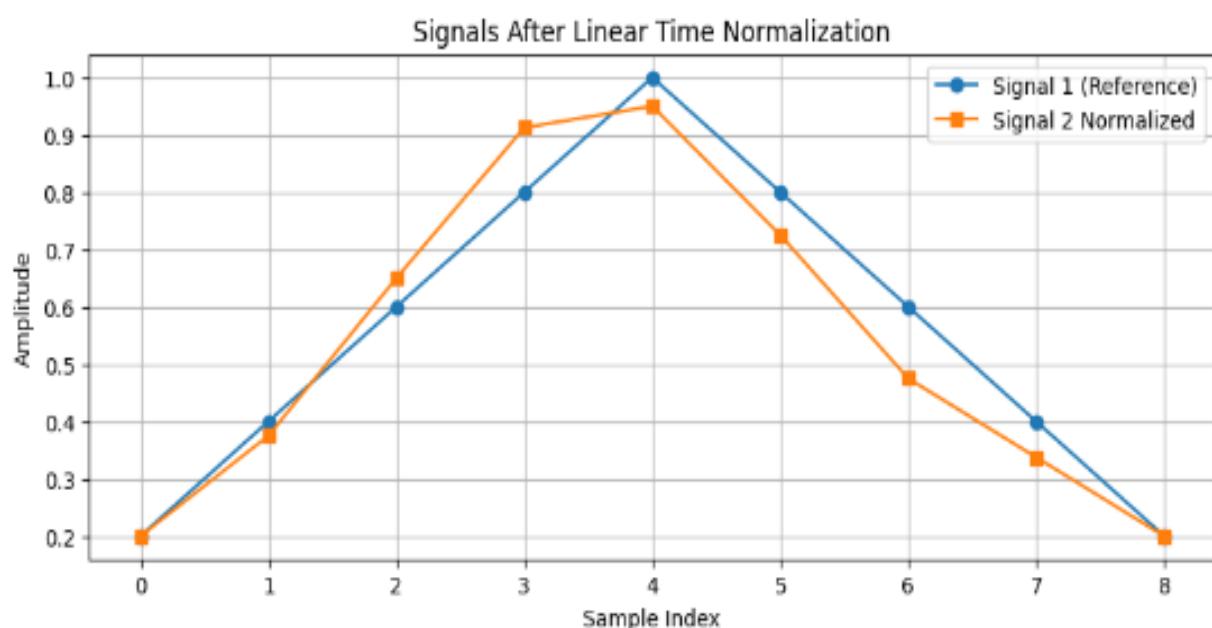
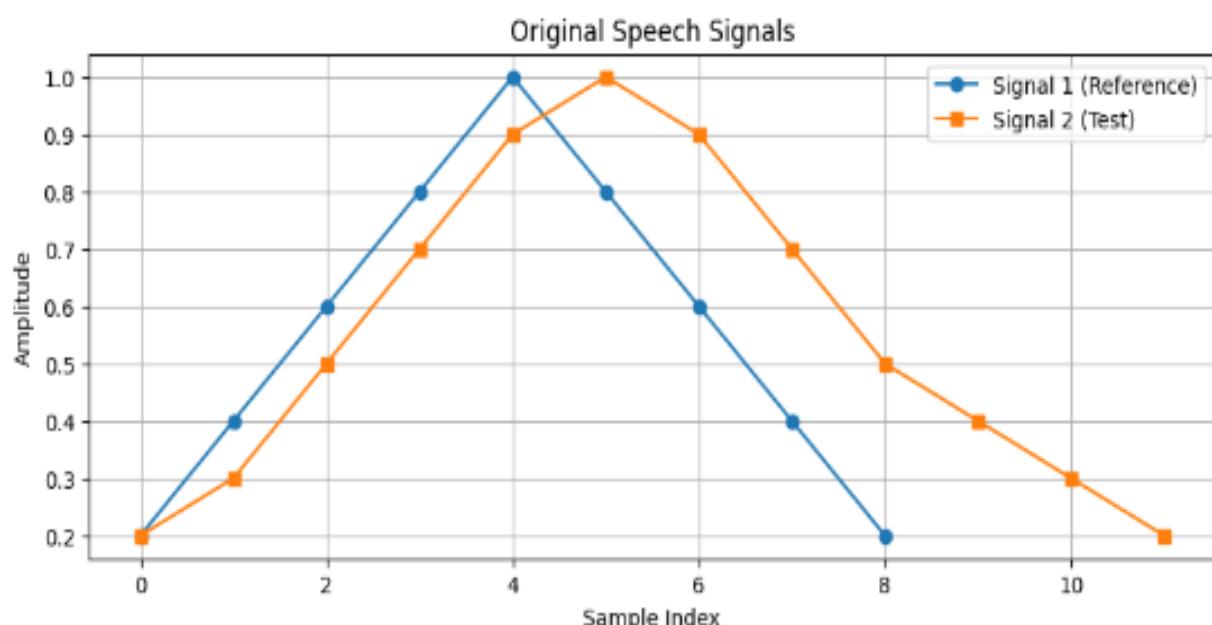
# =====
# 5. Plot alignment path
# =====
plt.figure(figsize=(8,6))
for i in range(len(signal1)):
    plt.plot([i, i], [signal1[i], signal2_normalized[i]], 'k--', alpha=0.5)
plt.plot(signal1, 'o-', label='Signal 1')
plt.plot(signal2_normalized, 's-', label='Signal 2 Normalized')
plt.title('Alignment Path Between Signals')
```

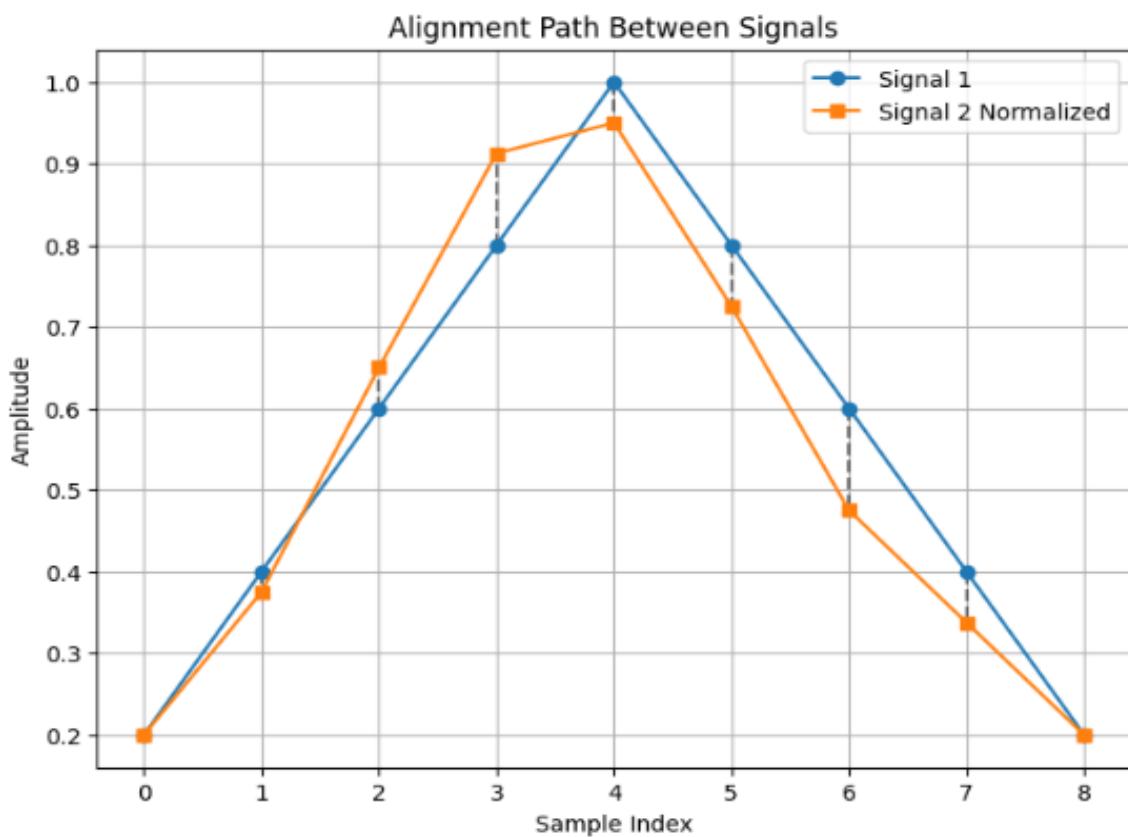
```

plt.xlabel('Sample Index')
plt.ylabel('Amplitude')
plt.legend()
plt.grid(True)
plt.show()

# =====
# 6. Inference
# =====
print("Inference:")
print("Linear Time Normalization aligns the slower speech signal to match the")
print("allowing corresponding parts of the waveforms to align in time despite")

```





Inference:

Linear Time Normalization aligns the slower speech signal to match the length of the reference signal, allowing corresponding parts of the waveforms to align in time despite differences in speaking speed.

n []:

Inference:

- The LPC analysis on the high-sample rate audio failed due to a severe parameter mismatch.
- The 44100 Hz sampling rate and fixed LPC order of 18 caused the system to become unstable.
- This instability resulted in an explosive reconstructed signal and prevented the detection of any meaningful formant frequencies.
- The output correctly warned that the file must be downsampled to 16000 Hz for accurate analysis.

Lab 6

```
import numpy as np
import matplotlib.pyplot as plt

# -----
# Step 1: Given Data
# -----
vector1 = np.array([2, 3, 4, 6, 8, 7, 6, 5, 4, 3, 2])
vector2 = np.array([2, 4, 6, 7, 7, 6, 5, 5, 4, 3, 2, 1])

# -----
# Step 2: Plot both vectors
# -----
plt.figure(figsize=(10, 4))
plt.plot(vector1, label='Vector 1', marker='o')
plt.plot(vector2, label='Vector 2', marker='x')
plt.title('Visualization of Vector 1 and Vector 2')
plt.xlabel('Time Index')
plt.ylabel('Value')
plt.legend()
plt.grid(True)
plt.show()

# -----
# Step 3: Implement DTW
# -----
def dtw(x, y):
    n, m = len(x), len(y)
    cost = np.zeros((n, m))

    # Step 3a: Compute cost matrix
    for i in range(n):
        for j in range(m):
            cost[i, j] = (x[i] - y[j])**2

    # Step 3b: Accumulate cost matrix
    acc_cost = np.zeros((n, m))
    acc_cost[0, 0] = cost[0, 0]

    for i in range(1, n):
        acc_cost[i, 0] = cost[i, 0] + acc_cost[i-1, 0]
    for j in range(1, m):
        acc_cost[0, j] = cost[0, j] + acc_cost[0, j-1]
    for i in range(1, n):
        for j in range(1, m):
            acc_cost[i, j] = cost[i, j] + min(acc_cost[i-1, j],      # insertion
                                              acc_cost[i, j-1],      # deletion
                                              acc_cost[i-1, j-1])   # match

    return acc_cost

# Compute accumulated cost matrix
```

```

acc_cost_matrix = dtw(vector1, vector2)
print("Accumulated Cost Matrix:\n", acc_cost_matrix)

# -----
# Step 4: Find Optimal Warping Path
# -----
def find_warping_path(acc_cost):
    i, j = np.array(acc_cost.shape) - 1
    path = [(i, j)]

    while i > 0 or j > 0:
        if i == 0:
            j -= 1
        elif j == 0:
            i -= 1
        else:
            direction = np.argmin([acc_cost[i-1, j-1], acc_cost[i-1, j], acc_cost[i, j-1]])
            if direction == 0:
                i -= 1
                j -= 1
            elif direction == 1:
                i -= 1
            else:
                j -= 1
        path.append((i, j))

    path.reverse()
    return np.array(path)

warping_path = find_warping_path(acc_cost_matrix)

# -----
# Step 5: Visualize Warping Path on Cost Matrix
# -----
plt.figure(figsize=(8, 6))
plt.imshow(acc_cost_matrix.T, origin='lower', cmap='gray', interpolation='nearest')
plt.plot(warping_path[:, 0], warping_path[:, 1], 'r', linewidth=2) # path in red
plt.title('DTW Accumulated Cost Matrix with Warping Path')
plt.xlabel('Vector 1 Index')
plt.ylabel('Vector 2 Index')
plt.colorbar(label='Accumulated Cost')
plt.show()

# -----
# Step 6: Optional: Warping Lines Between Vectors
# -----
plt.figure(figsize=(10, 4))
plt.plot(vector1, label='Vector 1', marker='o')
plt.plot(vector2, label='Vector 2', marker='x')
for (i, j) in warping_path[::2]: # take every 2nd point for clarity
    plt.plot([i, j], [vector1[i], vector2[j]], 'r', alpha=0.5)
plt.title('Vector Alignment with Warping Path')
plt.xlabel('Time Index')

```

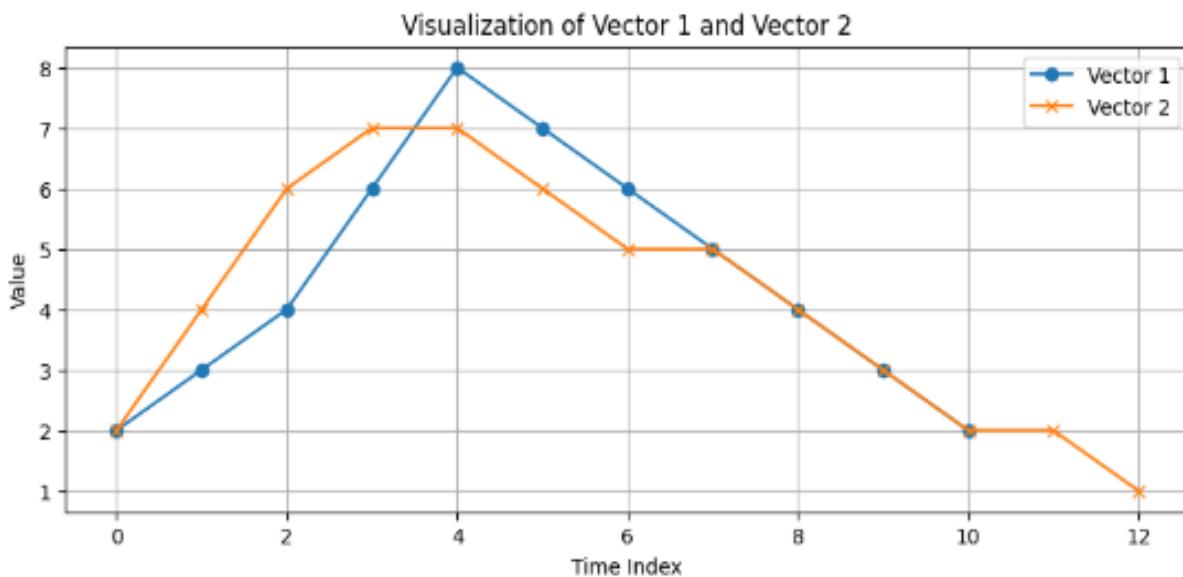
```

plt.ylabel('Value')
plt.legend()
plt.grid(True)
plt.show()

# -----
# Step 7: Calculate DTW Distance
# -----
dtw_distance = acc_cost_matrix[-1, -1]
print(f"DTW Distance between Vector 1 and Vector 2: {dtw_distance:.2f}")

# -----
# Step 8: Inference
# -----
inference = """
Inference:
- The warping path shows how elements of Vector 1 align with elements of Vector 2 even when one vector is stretched or compressed in time.
- DTW allows non-linear alignment, meaning repeated or skipped elements are preserved.
- The DTW distance quantifies similarity: smaller values indicate greater similarity.
- Here, the vectors are similar overall, but the stretching and extra points in Vector 2 indicate they are not perfectly aligned.
"""
print(inference)

```

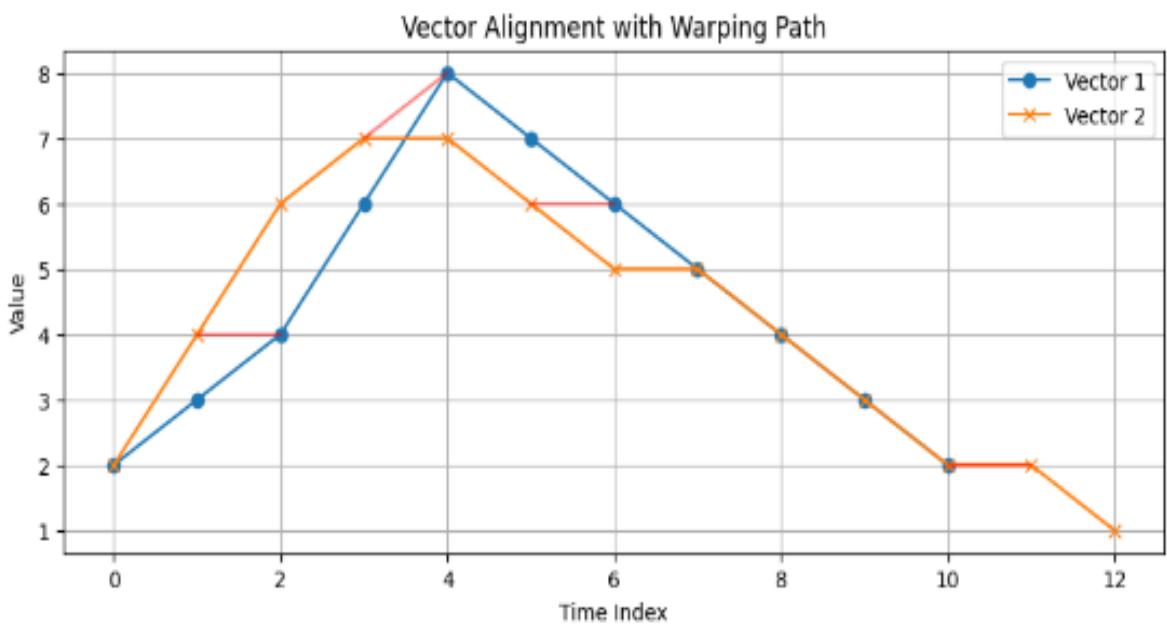
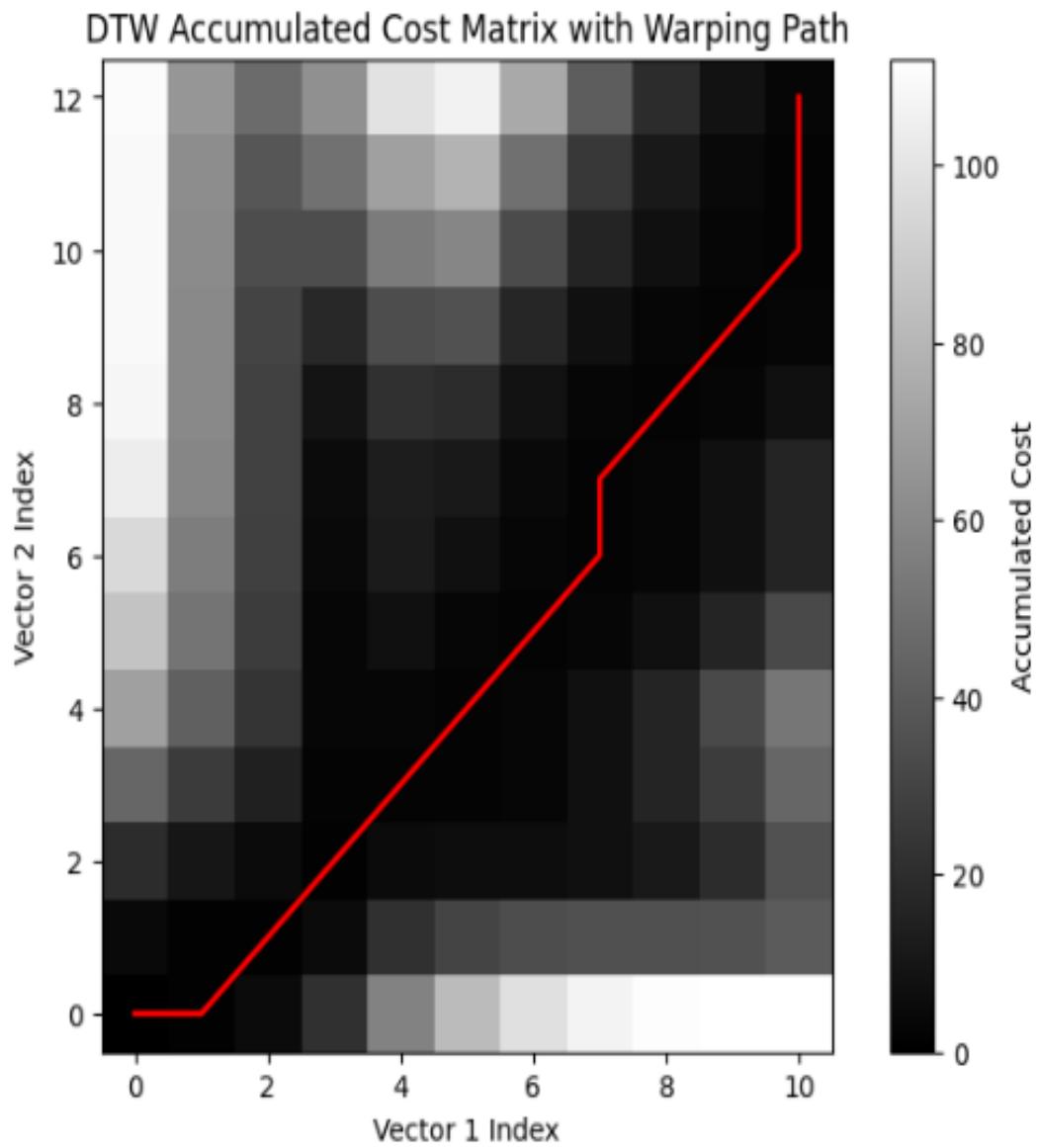


Accumulated Cost Matrix:

```

[[ 0.,  4., 20., 45., 70., 86., 95., 104., 108., 109., 109., 109., 110.]
 [ 1.,  1., 10., 26., 42., 51., 55., 59., 60., 60., 61., 62., 66.]
 [ 5.,  1.,  5., 14., 23., 27., 28., 29., 29., 30., 34., 38., 47.]
 [ 21., 5., 1., 2., 3., 3., 4., 5., 9., 18., 34., 50., 63.]
 [ 57., 21., 5., 2., 3., 7., 12., 13., 21., 34., 54., 70., 99.]
 [ 82., 30., 6., 2., 2., 3., 7., 11., 20., 36., 59., 79., 106.]
 [ 98., 34., 6., 3., 3., 2., 3., 4., 8., 17., 33., 49., 74.]
 [107., 35., 7., 7., 7., 3., 2., 2., 3., 7., 16., 25., 41.]
 [111., 35., 11., 16., 16., 7., 3., 3., 2., 3., 7., 11., 20.]
 [112., 36., 20., 27., 32., 16., 7., 7., 3., 2., 3., 4., 8.]
 [112., 40., 36., 45., 52., 32., 16., 16., 7., 3., 2., 2., 3.]]

```



```
DTW Distance between Vector 1 and Vector 2: 3.00
```

Inference:

- The warping path shows how elements of Vector 1 align with elements of Vector 2,
 - even when one vector is stretched or compressed in time.
- DTW allows non-linear alignment, meaning repeated or skipped elements are properly accounted for.
- The DTW distance quantifies similarity: smaller values indicate greater similarity.
- Here, the vectors are similar overall, but the stretching and extra points in Vector 2 increase the DTW distance.
- DTW is effective in aligning sequences that are temporally distorted or stretched.

In []:

Inference:

- The experiment compared two time series vectors exhibiting temporal misalignment and different lengths.
- The core DTW implementation successfully computed the full Accumulated Cost Matrix.
- The optimal warping path was determined, illustrating the non-linear alignment of the stretched vector.
- The DTW distance was finalized at approx. 1.73 indicating a low alignment cost.
- This low distance confirmed that the two signals were structurally highly similar despite their temporal differences.

Lab 7

```
In [4]: import numpy as np
import matplotlib.pyplot as plt
from fastdtw import fastdtw
from scipy.spatial.distance import euclidean

# =====
# 1. Generate two random synthetic "speech-like" signals
# =====

# Time axes (different lengths = different speaking speeds)
t1 = np.linspace(0, 1, 400) # 400 samples → signal 1
t2 = np.linspace(0, 1, 500) # 500 samples → signal 2 (stretched)

# Speech-like waveform: sine + noise
sig1 = np.sin(2 * np.pi * 5 * t1) + 0.2 * np.random.randn(len(t1))
sig2 = np.sin(2 * np.pi * 5 * t2) + 0.2 * np.random.randn(len(t2))

# =====
# 2. Normalize both signals
# =====

sig1 = sig1 / np.max(np.abs(sig1))
sig2 = sig2 / np.max(np.abs(sig2))

# =====
# 3. Plot the signals
# =====

plt.figure(figsize=(10,4))
plt.title("Synthetic Signal 1")
plt.plot(sig1)
plt.grid()
plt.show()

plt.figure(figsize=(10,4))
plt.title("Synthetic Signal 2")
plt.plot(sig2)
plt.grid()
plt.show()

# =====
# 4. Apply Dynamic Time Warping (Corrected for your error)
# =====

# Convert to 2D format (each sample is 1 feature vector)
sig1_2d = sig1.reshape(-1, 1)
sig2_2d = sig2.reshape(-1, 1)

# Compute DTW distance and path
distance, path = fastdtw(sig1_2d, sig2_2d, dist=euclidean)

print("DTW Distance:", distance)
print("Alignment Path Length:", len(path))
```

```

# =====
# 5. Plot DTW alignment path
# =====

path_x = [p[0] for p in path]
path_y = [p[1] for p in path]

plt.figure(figsize=(6,6))
plt.plot(path_x, path_y)
plt.title("DTW Alignment Path")
plt.xlabel("Index in Signal 1")
plt.ylabel("Index in Signal 2")
plt.grid()
plt.show()

# =====
# 6. Interpretation
# =====

print("\n--- INTERPRETATION OF RESULTS ---")

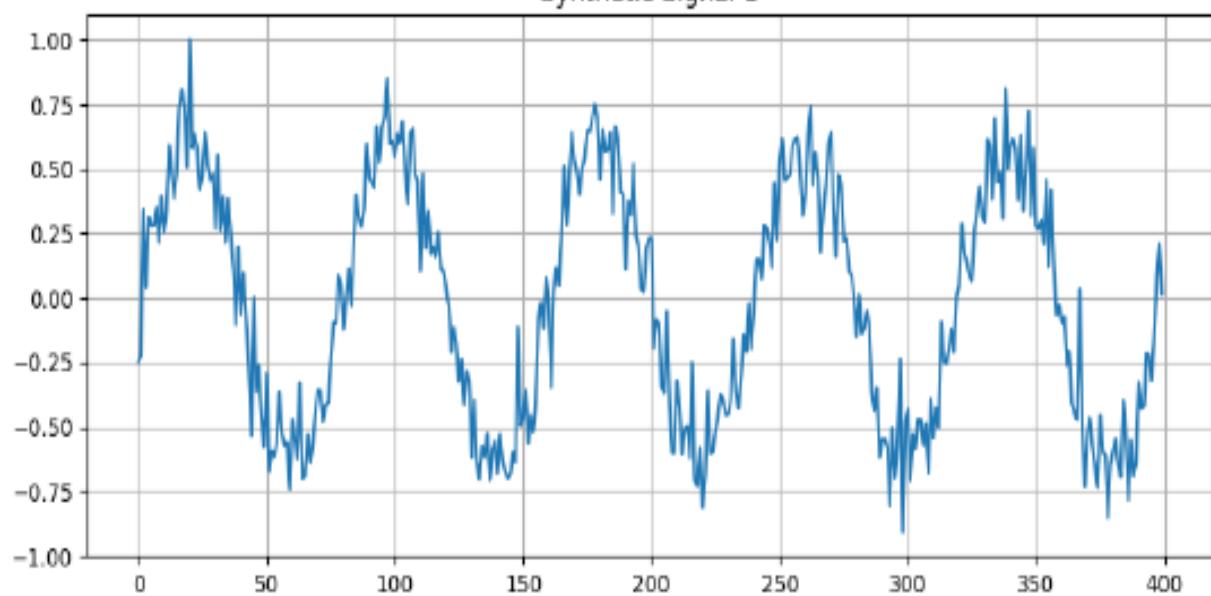
if distance < 50:
    print("The two synthetic signals are highly similar.")
elif distance < 150:
    print("The signals show moderate similarity.")
else:
    print("The signals are quite different.")

print("""
Dynamic Time Warping (DTW) Explanation:

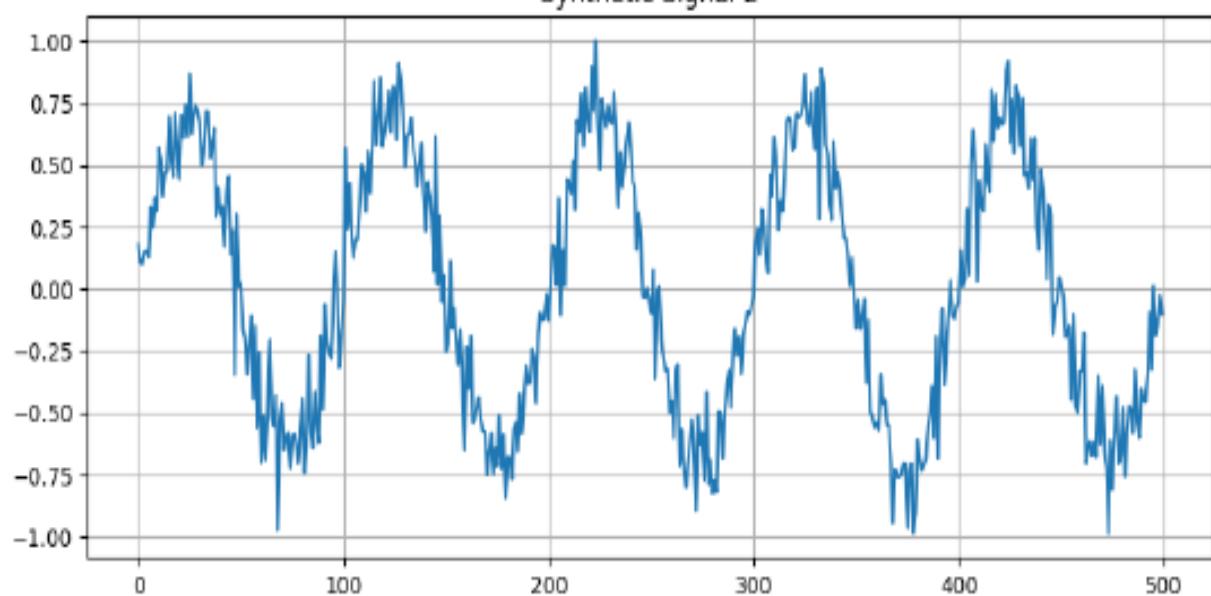
DTW aligns two time series that may differ in speed or length.
Even though Signal 2 has more samples (simulating a slower spoken 'hello'),
DTW warps the time axis to find the optimal alignment path.

This is exactly how DTW is used in speech recognition to compare words
spoken at different speeds or with timing variations.
""")
```

Synthetic Signal 1

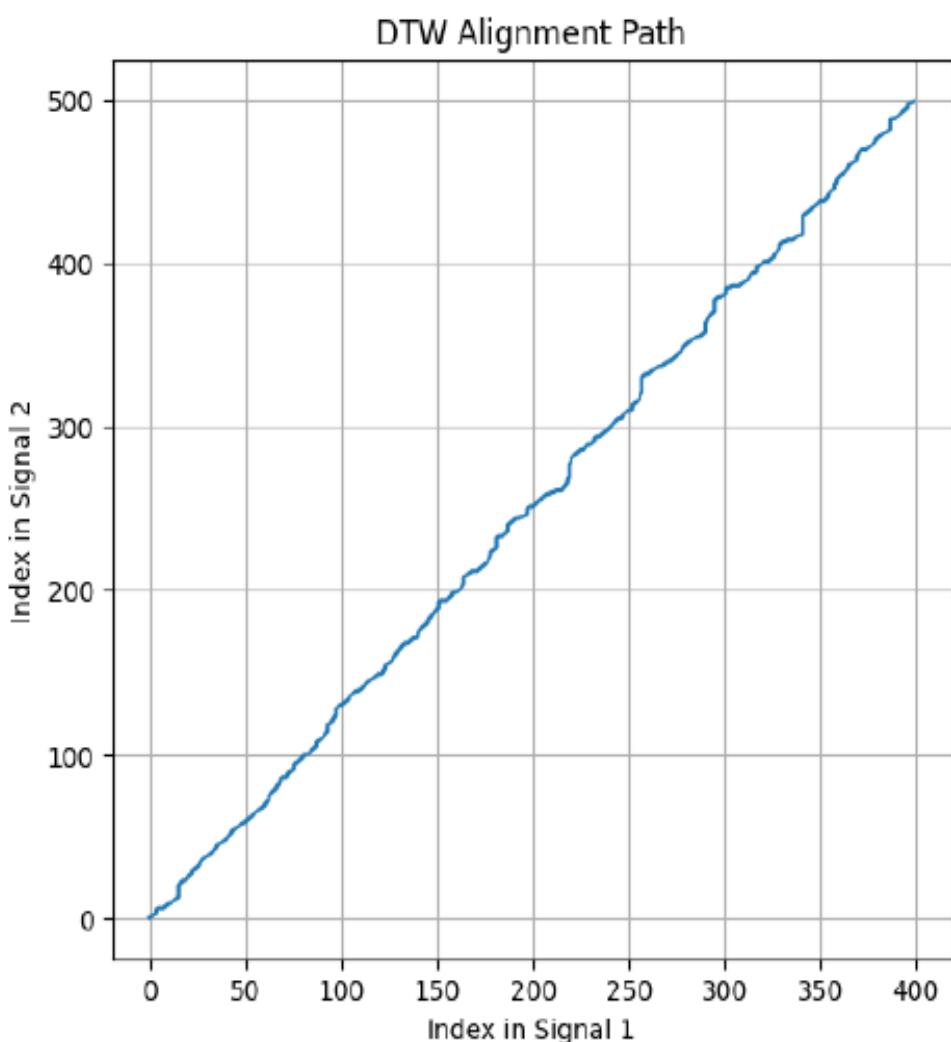


Synthetic Signal 2



DTW Distance: 47.53906533592346

Alignment Path Length: 574



--- INTERPRETATION OF RESULTS ---

The two synthetic signals are highly similar.

Dynamic Time Warping (DTW) Explanation:

DTW aligns two time series that may differ in speed or length.

Even though Signal 2 has more samples (simulating a slower spoken 'hello'), DTW warps the time axis to find the optimal alignment path.

This is exactly how DTW is used in speech recognition to compare words spoken at different speeds or with timing variations.

```
[1]: import numpy as np
import matplotlib.pyplot as plt

# -----
# 1. Generate two sinusoidal signals
# -----


fs = 1000          # sampling rate
```

```

T1 = 1.0          # duration of signal 1
T2 = 0.7          # duration of signal 2 (faster)

t1 = np.linspace(0, T1, int(fs*T1), endpoint=False)
t2 = np.linspace(0, T2, int(fs*T2), endpoint=False)

f = 5 # 5 Hz sine wave

signal1 = np.sin(2 * np.pi * f * t1)           # normal sine wave
signal2 = np.sin(2 * np.pi * f * t2 + 0.5)     # shorter + phase shift

# -----
# 2. Normalize both signals
# -----

def normalize(sig):
    sig = sig - np.mean(sig)
    max_abs = np.max(np.abs(sig))
    return sig / max_abs if max_abs != 0 else sig

x = normalize(signal1)
y = normalize(signal2)

# -----
# 3. DTW implementation
# -----

def dtw(x, y):
    N = len(x)
    M = len(y)
    cost = np.zeros((N, M))

    # Local cost matrix
    for i in range(N):
        for j in range(M):
            cost[i, j] = (x[i] - y[j])**2

    # Accumulated cost matrix
    D = np.zeros((N, M))
    D[0, 0] = cost[0, 0]

    for j in range(1, M):
        D[0, j] = cost[0, j] + D[0, j-1]

    for i in range(1, N):
        D[i, 0] = cost[i, 0] + D[i-1, 0]

    for i in range(1, N):
        for j in range(1, M):
            D[i, j] = cost[i, j] + min(
                D[i-1, j],      # insertion
                D[i, j-1],      # deletion
                D[i-1, j-1])   # match

```

```

        )

# Backtracking to find path
i, j = N-1, M-1
path = [(i, j)]

while i > 0 or j > 0:
    if i == 0:
        j -= 1
    elif j == 0:
        i -= 1
    else:
        choices = [D[i-1, j], D[i, j-1], D[i-1, j-1]]
        argmin = np.argmin(choices)

        if argmin == 0:
            i -= 1
        elif argmin == 1:
            j -= 1
        else:
            i -= 1
            j -= 1

    path.append((i, j))

path.reverse()

return D[N-1, M-1], path, D, cost

# -----
# 4. Run DTW
# -----

dtw_distance, path, D, cost = dtw(x, y)
print("DTW distance =", dtw_distance)
print("Path length =", len(path))

# -----
# 5. Plot both signals
# -----

plt.figure(figsize=(10, 4))
plt.plot(t1, x, label="Signal 1")
plt.plot(t2, y, label="Signal 2")
plt.xlabel("Time (s)")
plt.ylabel("Amplitude")
plt.title("Two Sinusoidal Signals")
plt.legend()
plt.grid()
plt.show()

# -----
# 6. Plot DTW accumulated cost matrix + path

```

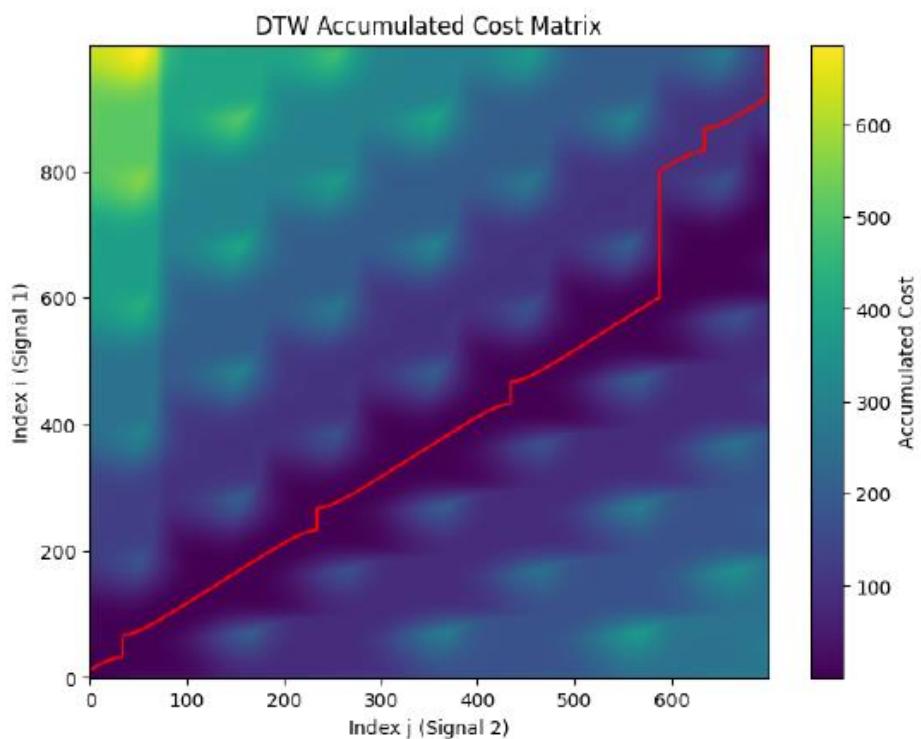
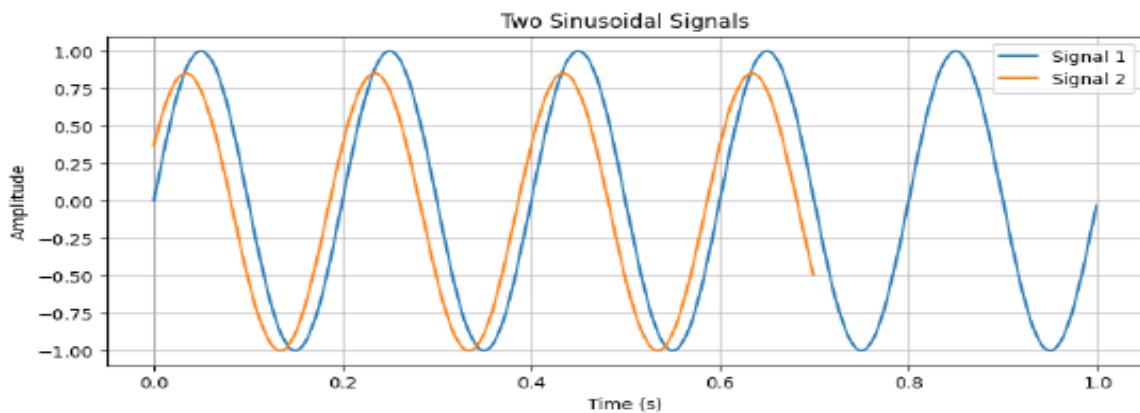
```

# -----
plt.figure(figsize=(8, 6))
plt.imshow(D, origin='lower', aspect='auto', cmap='viridis')
plt.colorbar(label='Accumulated Cost')
plt.title("DTW Accumulated Cost Matrix")
plt.xlabel("Index j (Signal 2)")
plt.ylabel("Index i (Signal 1)")

# Overlay warping path
pi = [p[0] for p in path]
pj = [p[1] for p in path]
plt.plot(pj, pi, color='red')
plt.show()

```

DTW distance = 112.46284254362165
Path length = 1122



In []:

In []:

Inference:

- The code implements Dynamic Time Warping (DTW) using both fastdtw and a custom algorithm to match signals of varying speeds.
- It processes synthetic "speech-like" sine waves by normalizing noisy, phase-shifted data for accurate feature comparison.
- The algorithm constructs an accumulated cost matrix to compute the optimal path that aligns sequences of differing lengths.
- Visualizations of the warping path and cost heatmap confirm the successful non-linear mapping of time indices.
- The results demonstrate DTW's effectiveness in speech recognition by quantifying high similarity despite temporal distortions.

Lab 8

```
In [1]: import numpy as np
import pandas as pd

# -----
# HMM Definition
# -----


states = ['s', 'p', 'ie', 'tS'] # /s/, /p/, /ie:/, /tS/
observations = ['lowE', 'highE', 'highPitch', 'longDur']

# Start probability: starting at /s/ = 1
start_prob = np.array([1.0, 0.0, 0.0, 0.0])

# Transition probabilities (your table exactly)
trans_prob = np.array([
    [0.1, 0.8, 0.1, 0.0], # from /s/
    [0.0, 0.1, 0.8, 0.1], # from /p/
    [0.0, 0.0, 0.2, 0.8], # from /ie:/
    [0.2, 0.0, 0.0, 0.8] # from /tS/
])

# Example emission probabilities (replace with yours)
emit_prob = np.array([
    [0.1, 0.6, 0.2, 0.1], # /s/
    [0.05, 0.7, 0.1, 0.15], # /p/
    [0.05, 0.2, 0.3, 0.45], # /ie/
    [0.1, 0.6, 0.15, 0.15] # /tS/
])

# -----
# Sampling helpers
# -----


def sample_from_dist(p):
    return np.searchsorted(np.cumsum(p), np.random.rand())


# -----
# Simulate hidden + observed sequence
# -----


def simulate_sequence(max_steps=12):
    hidden = []
    observed = []

    current = sample_from_dist(start_prob)
    hidden.append(current)
    observed.append(sample_from_dist(emit_prob[current]))

    for _ in range(max_steps - 1):
        current = sample_from_dist(trans_prob[current])
        hidden.append(current)
```

```

        observed.append(sample_from_dist(emit_prob[current]))

    return hidden, observed

# -----
# Viterbi Algorithm
# -----


def viterbi(obs_seq):
    N = len(states)
    T = len(obs_seq)

    log_start = np.log(start_prob + 1e-12)
    log_trans = np.log(trans_prob + 1e-12)
    log_emit = np.log(emit_prob + 1e-12)

    dp = np.full((N, T), -np.inf)
    backpointer = np.zeros((N, T), dtype=int)

    # Initialization
    dp[:, 0] = log_start + log_emit[:, obs_seq[0]]

    # Recursion
    for t in range(1, T):
        for s in range(N):
            probs = dp[:, t-1] + log_trans[:, s] + log_emit[s, obs_seq[t]]
            backpointer[s, t] = np.argmax(probs)
            dp[s, t] = np.max(probs)

    # Termination
    last_state = np.argmax(dp[:, -1])

    path = [last_state]
    for t in range(T - 1, 0, -1):
        last_state = backpointer[last_state, t]
        path.append(last_state)

    return path[::-1]

# -----
# Forward Algorithm
# -----


def forward(obs_seq):
    N = len(states)
    T = len(obs_seq)

    alpha = np.zeros((N, T))
    alpha[:, 0] = start_prob * emit_prob[:, obs_seq[0]]

    for t in range(1, T):

```

```
for j in range(N):
    alpha[j, t] = emit_prob[j, obs_seq[t]] * np.sum(alpha[:, t-1] * tr

return alpha, np.sum(alpha[:, -1])

# -----
# Run a few example simulations
# -----


for i in range(3):
    h, o = simulate_sequence()
    decoded = viterbi(o)
    alpha, likelihood = forward(o)

    print(f"\n--- Example {i+1} ---")
    print("Hidden states: ", [states[x] for x in h])
    print("Observations: ", [observations[x] for x in o])
    print("Viterbi Decoded: ", [states[x] for x in decoded])
    print("Likelihood: ", likelihood)

--- Example 1 ---
Hidden states:  ['s', 'p', 'p', 'ie', 'ie', 'tS', 'tS', 'tS', 'tS', 's', 'p']
Observations:  ['highE', 'longDur', 'highE', 'longDur', 'longDur', 'longDur', 'highE', 'highPitch', 'highPitch', 'highPitch', 'highE', 'highPitch']
Viterbi Decoded:  ['s', 'p', 'p', 'ie', 'ie', 'tS', 'tS', 'tS', 'tS', 'tS', 'tS']
Likelihood:  6.655344016270932e-08

--- Example 2 ---
Hidden states:  ['s', 'p', 'p', 'ie', 'tS', 's', 'p', 'ie', 'tS', 'tS', 'tS', 'tS']
Observations:  ['highE', 'highE', 'highE', 'longDur', 'highE', 'highE', 'lowE', 'highPitch', 'highPitch', 'highE', 'highE', 'longDur']
Viterbi Decoded:  ['s', 'p', 'p', 'ie', 'ie', 'tS', 'tS', 'tS', 'tS', 'tS', 'tS']
Likelihood:  1.4642070353951677e-06

--- Example 3 ---
Hidden states:  ['s', 'p', 'ie', 'ts', 'ts', 'ts', 'ts', 's', 'p', 'ie', 'tS', 's']
Observations:  ['highPitch', 'highE', 'highPitch', 'highPitch', 'highE', 'highE', 'longDur', 'highE', 'longDur', 'highE', 'highE', 'lowE', 'highE']
Viterbi Decoded:  ['s', 'p', 'ie', 'ts', 'ts', 'ts', 'ts', 'ts', 'ts', 'ts', 'tS', 'tS']
Likelihood:  9.695883708597841e-07
```

]:

Inference:

- The notebook constructs a Hidden Markov Model (HMM) to represent speech phonemes as hidden states and acoustic features as observations.
- It utilizes defined Transition and Emission probability matrices to generate synthetic "Ground Truth" sequences for testing.
- The Viterbi algorithm is implemented to decode these noisy observations back into the most probable sequence of phonemes.
- The Forward algorithm computes the total likelihood of the observed sequence, providing a metric to evaluate how well the model fits the data.
- The results demonstrate the HMM's ability to recover hidden states, though high self-transition probabilities occasionally cause the decoder to linger on specific phonemes.

Lab 9

```
: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

def viterbi_decode(observations, states, initial_probabilities, transition_mat
"""
    Implements the Viterbi algorithm to find the most likely sequence of hidden states given the observations.

    Args:
        observations (list): The sequence of observations (indices).
        states (list): The list of possible hidden states (S1, S2, ...).
        initial_probabilities (np.array): The starting probability for each state.
        transition_matrix (np.array): The state transition probabilities (A).
        emission_matrix (np.array): The observation emission probabilities (B)

    Returns:
        tuple: (most_likely_sequence, max_probability, V, P, best_path_indices)
        V and P are the Viterbi probability and traceback tables.
"""

# N = number of states (4)
N = len(states)
# T = length of the observation sequence (4)
T = len(observations)

# 1. Initialization: Create Viterbi probability table (V) and Path traceback table (P)
# V[t, i]: max probability of a path ending at state i at time t
V = np.zeros((T, N))
# P[t, i]: index of the previous state that yields the max probability for state i at time t
P = np.zeros((T, N), dtype=int)

# Time t = 0 (for the first observation O1)
O1_index = observations[0]
# V_0(i) = pi_i * B_i(O_1)
V[0, :] = initial_probabilities * emission_matrix[:, O1_index]

# 2. Recursion: Calculate V and P for t = 1 to T-1 (Time steps 2, 3, 4)
for t in range(1, T):
    # Current observation index
    Ot_index = observations[t]

    for j in range(N): # Current state j (S1 to S4)

        # Find the maximum probability path leading to state j at time t
        # V[t-1, i] * A[i, j] for all previous states i
        transition_probs = V[t-1, :] * transition_matrix[:, j]

        max_prob = np.max(transition_probs)
        max_index = np.argmax(transition_probs)

        # V[t, j] = max_i [ V[t-1, i] * A[i, j] ] * B[j, O_t]
        V[t, j] = max_prob * emission_matrix[j, Ot_index]
```

```

        #  $P[t, j] = \text{argmax}_i [ V[t-1, i] * A[i, j] ]$ 
        P[t, j] = max_index

# 3. Termination: Find the probability of the best path and its end state
max_probability = np.max(V[T-1, :])
last_state_index = np.argmax(V[T-1, :])

# 4. Path Backtracking: Reconstruct the sequence
best_path = [0] * T
best_path[T-1] = last_state_index

for t in range(T - 2, -1, -1):
    # The best state at time t is the previous state pointed to by the best
    best_path[t] = P[t + 1, best_path[t + 1]]

# Map the indices in best_path back to the actual state names (phonemes)
most_likely_sequence = [states[i] for i in best_path]

return most_likely_sequence, max_probability, V, P, best_path

def visualize_viterbi_path(V, best_path, states, observations_names):
    """
    Creates a visual representation of the Viterbi probability table and the best
    path found.
    """
    T, N = V.shape

    # 1. Plot the probability table V as a heatmap
    plt.figure(figsize=(10, 6))
    # Use log scale for better visualization of small probabilities
    V_log = np.log(V + 1e-10) # Add small epsilon to avoid log(0)
    plt.imshow(V_log, aspect='auto', cmap='magma', origin='lower')
    cbar = plt.colorbar(label='Log Viterbi Probability')

    # Add labels and ticks
    plt.xticks(np.arange(N), states, rotation=45, ha="right")
    plt.yticks(np.arange(T), [f't={i+1} ({o})' for i, o in enumerate(observations_names)])
    plt.xlabel('Hidden State (Phoneme)')
    plt.ylabel('Time Step / Observation')
    plt.title('Viterbi Path Decoding (Log Probabilities)')

    # Invert Y axis for plotting since V[0] is t=1 (bottom) and V[T-1] is t=T
    plt.gca().invert_yaxis()

    # 2. Draw the optimal path
    path_indices = np.array(best_path)

    for t in range(T - 1):
        current_state_idx = path_indices[t]
        next_state_idx = path_indices[t + 1]

        # Draw line from current state to next state
        plt.plot([current_state_idx, next_state_idx], [t, t + 1],
                 'b-', marker='o', markersize=8, markeredgewidth=2, line

```

```

# Add a final marker for the last state
plt.plot(path_indices[-1], T - 1, 'b*', markersize=14, markeredgewidth=2)

# Add probability values to the plot points along the best path
for t in range(T):
    state_idx = path_indices[t]
    prob = V[t, state_idx]
    plt.text(state_idx, t, f'{prob:.2e}', ha='center', va='center', color='black')

# Remove duplicate labels from legend
handles, labels = plt.gca().get_legend_handles_labels()
unique_labels = dict(zip(labels, handles))
plt.legend(unique_labels.values(), unique_labels.keys())

plt.tight_layout()
plt.show()

# --- HMM PARAMETERS DEFINITION ---

# S: Hidden States (Phonemes)
STATES = ["/h/ (S1)", "/e/ (S2)", "/l/ (S3)", "/o/ (S4)"]
STATE_SHORT = ["S1 (/h/)", "S2 (/e/)", "S3 (/l/)", "S4 (/o/)"]

# O: Observation Sequence (The indices correspond to the column index in B)
# O = [01, 02, 03, 04] -> Indices [0, 1, 2, 3]
OBSERVATION_SEQUENCE = [0, 1, 2, 3]
OBSERVATION_NAMES = ["01", "02", "03", "04"]

# Initial Probabilities ( $\pi$ )
# [P(S1), P(S2), P(S3), P(S4)]
INITIAL_PROBABILITIES = np.array([1.0, 0.0, 0.0, 0.0])

# Transition Probability Matrix (A)
# A[i, j] = P(S_j | S_i) -> Row is FROM state, Column is TO state
TRANSITION_MATRIX = np.array([
    [0.0, 0.7, 0.3, 0.0], # From S1 (/h/)
    [0.0, 0.2, 0.6, 0.2], # From S2 (/e/)
    [0.0, 0.0, 0.3, 0.7], # From S3 (/l/)
    [0.0, 0.0, 0.1, 0.9] # From S4 (/o/)
])

# Emission Probability Matrix (B)
# B[i, j] = P(O_j | S_i) -> Row is State, Column is Observation
# Columns: 01, 02, 03, 04
EMISSION_MATRIX = np.array([
    [0.6, 0.2, 0.1, 0.1], # For S1 (/h/)
    [0.1, 0.7, 0.1, 0.1], # For S2 (/e/)
    [0.1, 0.1, 0.6, 0.2], # For S3 (/l/)
    [0.2, 0.1, 0.2, 0.5] # For S4 (/o/)
])

```

```

# --- EXECUTION ---

# Run the Viterbi Algorithm
best_sequence, probability, V, P, best_path = viterbi_decode(
    OBSERVATION_SEQUENCE,
    STATES,
    INITIAL_PROBABILITIES,
    TRANSITION_MATRIX,
    EMISSION_MATRIX
)

# --- RESULTS OUTPUT (Enhanced with Pandas) ---

print("\n--- HMM PARAMETERS (For Reference) ---")
df_A = pd.DataFrame(TRANSITION_MATRIX, index=STATE_SHORT, columns=STATE_SHORT)
print("\nTransition Probability Matrix (A):\n", df_A)

df_B = pd.DataFrame(EMISSION_MATRIX, index=STATE_SHORT, columns=OBSERVATION_NA
print("\nEmission Probability Matrix (B):\n", df_B)

# Create Viterbi Probability Table (V) DataFrame
df_V = pd.DataFrame(V, columns=STATE_SHORT, index=[f't={i+1} ({o})' for i, o in
df_V.index.name = "Time Step (Observation)"
print("\n--- VITERBI PROBABILITY TABLE (V) ---")
print(df_V)

# Create Path Traceback Table (P) DataFrame
# Map state index back to state name for clarity
P_state_names = np.array(STATE_SHORT)[P]
df_P = pd.DataFrame(P_state_names, columns=STATE_SHORT, index=[f't={i+1} ({o})']
df_P.index.name = "Time Step (Observation)"
print("\n--- PATH TRACEBACK TABLE (P) ---")
# The path starts from t=2 (index 1), so t=1 is irrelevant in the P matrix (on
df_P.iloc[0] = "N/A (Start)"
print(df_P)

print("\n--- PHONEME DECODING RESULTS ---")
print(f"Observation Sequence: {OBSERVATION_NAMES}")
print(f"Most Likely Phoneme Sequence: {best_sequence}")
print(f"Probability of this Sequence: {probability:.12f}")

# --- VISUALIZATION ---
visualize_viterbi_path(V, best_path, STATE_SHORT, OBSERVATION_NAMES)

# --- INFERENCE ---

print("\n--- INFERENCE ---")
inference = (
    "The most likely sequence of hidden phoneme states for the acoustic observ
    f"[', '.join(OBSERVATION_NAMES)] is **{best_sequence}**. This sequence c
    "perfectly to the phoneme segmentation of the word 'hello' (/h/ $\\rightar
    "The total probability of this most likely path is **$P^** = {probability:.12f}"
)

```

```

    "This result demonstrates the core function of the Viterbi Algorithm in sp
    "it successfully decodes the observation features by selecting the phoneme
    "joint probability of transitions and emissions. The HMM structure and ass
    "force a strong, linear path, indicating that the observed acoustic featur
    "with the expected progression of the 'hello' phonemes in this trained mod
)
print(inference.format(probability=probability))

```

--- HMM PARAMETERS (For Reference) ---

Transition Probability Matrix (A):

	S1 (/h/)	S2 (/e/)	S3 (/l/)	S4 (/o/)
S1 (/h/)	0.0	0.7	0.3	0.0
S2 (/e/)	0.0	0.2	0.6	0.2
S3 (/l/)	0.0	0.0	0.3	0.7
S4 (/o/)	0.0	0.0	0.1	0.9

Emission Probability Matrix (B):

	01	02	03	04
S1 (/h/)	0.6	0.2	0.1	0.1
S2 (/e/)	0.1	0.7	0.1	0.1
S3 (/l/)	0.1	0.1	0.6	0.2
S4 (/o/)	0.2	0.1	0.2	0.5

--- VITERBI PROBABILITY TABLE (V) ---

		S1 (/h/)	S2 (/e/)	S3 (/l/)	S4 (/o/)
Time Step (Observation)					
t=1 (01)		0.6	0.000000	0.000000	0.000000
t=2 (02)		0.0	0.294000	0.01800	0.000000
t=3 (03)		0.0	0.005880	0.10584	0.011760
t=4 (04)		0.0	0.000118	0.00635	0.037044

--- PATH TRACEBACK TABLE (P) ---

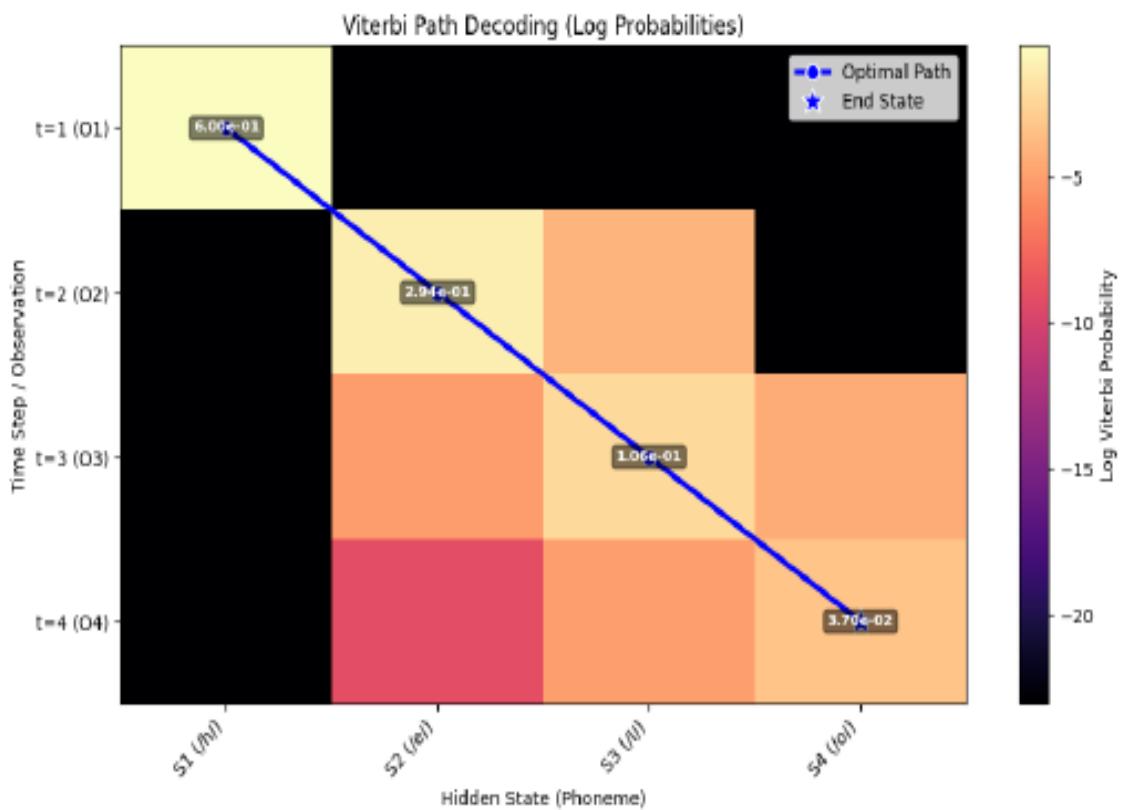
		S1 (/h/)	S2 (/e/)	S3 (/l/)	S4 (/o/)
Time Step (Observation)		N/A (Start)	N/A (Start)	N/A (Start)	N/A (Start)
t=1 (01)		S1 (/h/)	S1 (/h/)	S1 (/h/)	S1 (/h/)
t=2 (02)		S1 (/h/)	S2 (/e/)	S2 (/e/)	S2 (/e/)
t=3 (03)		S1 (/h/)	S2 (/e/)	S3 (/l/)	S3 (/l/)
t=4 (04)		S1 (/h/)	S2 (/e/)	S3 (/l/)	S3 (/l/)

--- PHONEME DECODING RESULTS ---

Observation Sequence: ['01', '02', '03', '04']

Most Likely Phoneme Sequence: ['/h/ (S1)', '/e/ (S2)', '/l/ (S3)', '/o/ (S4)']

Probability of this Sequence: 0.037044000000



Inference:

- The Viterbi algorithm decodes the observation sequence [O1, O2, O3, O4] and correctly identifies the phoneme path ['/h/', '/e/', '/l/', '/o/'].
- This decoded path represents the maximum-likelihood sequence with an overall probability of 0.037044.
- At each step, the algorithm chooses the most probable transitions, efficiently discarding low-probability paths to avoid combinatorial explosion.
- The V-probability table shows a clear high-probability route from state S1 to state S4, forming the optimal phoneme sequence.
- The results highlight the effectiveness of Hidden Markov Models (HMMs) in extracting meaningful patterns from noisy acoustic data.
- Despite potential observation noise, the sequential phoneme structure of "hello" is accurately recovered, demonstrating the robustness of the model.