EE798P Assignment 2

Adit Jain, 200038

September 4^{th}, 2023

1 Plotting

1.1 Waveform

A waveform is a graphical representation of a signal that shows how the amplitude (strength or intensity) of the signal varies with time. In other words, it is a visual depiction of how a signal changes over a period.

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

audio_file = "../assets/sound.ogg"
y, sr = librosa.load(audio_file)

time = np.arange(0, len(y)) / sr

plt.figure(figsize=(10, 4))
plt.plot(time, y)
plt.title("Waveform of Audio")
plt.xlabel("Time (s)")
plt.ylabel("Amplitude")
plt.grid()
plt.show()
```

1.2 Spectrogram

A spectrogram is a 2D representation of an audio signal over time, where the x-axis represents time, the y-axis represents frequency, and the color/intensity represents the amplitude (energy) of the signal at different time-frequency bins.

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


def plot_spectrogram(y, sr, window_size, hop_length, n_fft):
    spectogram = np.abs(
        librosa.stft(y, n_fft=n_fft, hop_length=hop_length, win_length=window_size)
    )
    spectogram_db = librosa.amplitude_to_db(spectogram, ref=np.max)
```

```
    plt.figure(figsize=(10, 6))
    librosa.display.specshow(
        spectogram_db, sr=sr, hop_length=hop_length, x_axis="time", y_axis="log"
    )
    plt.colorbar(format="%+2.0f dB")
    plt.title(f"Spectrogram (window={window_size}s, hop={hop_length}s, N={n_fft})")
    plt.xlabel("Time (s)")
    plt.ylabel("Frequency (Hz)")
    plt.show()


audio_file = "../assets/aa.ogg"
y, sr = librosa.load(audio_file)

plot_spectrogram(y, sr, 1024, 512, 4096)
```

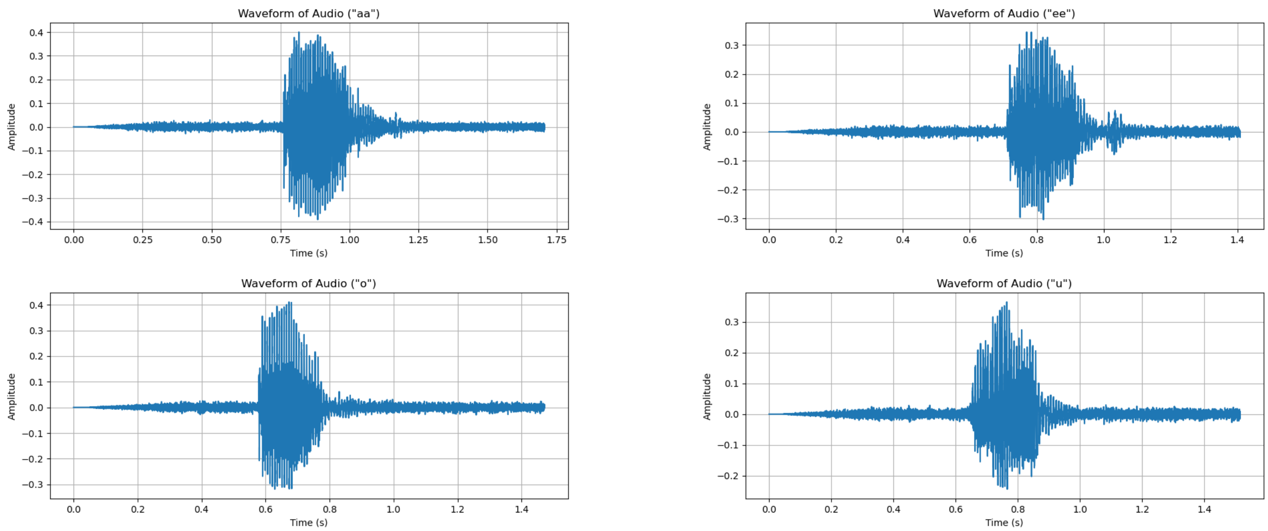
1.3 Spectrum

Spectrum refers to the distribution of energy across different frequencies in an audio signal. It provides information about the fundamental frequency (pitch) and the harmonic structure of a sound.

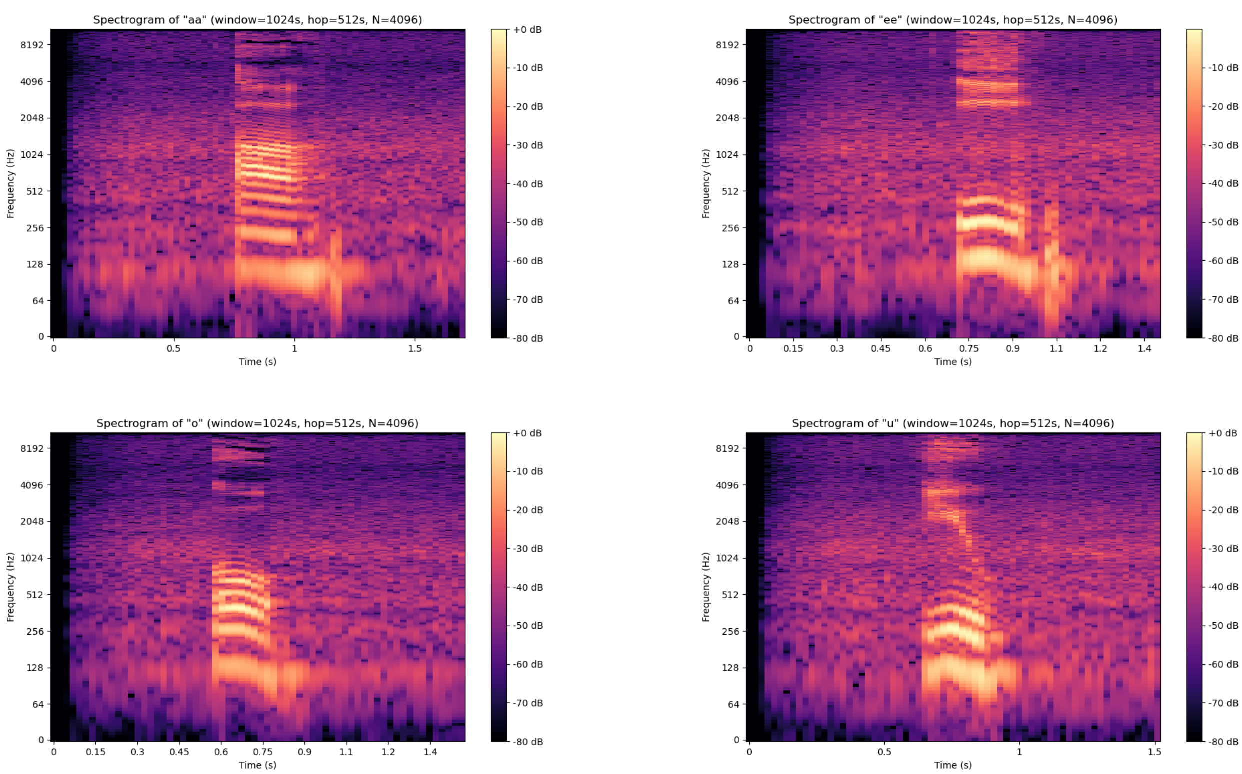
```
import librosa
import numpy as np
import matplotlib.pyplot as plt


def plot_spectrum(y, sr, start_time, n_fft):
    y_segment = y[int(start_time * sr) : int((start_time + 1) * sr)]
    X = np.abs(librosa.stft(y_segment, n_fft=n_fft))

    freqs = np.fft.fftfreq(n_fft, 1 / sr)
    positive_freqs = freqs[: n_fft // 2]
    spectrum = np.mean(X[: n_fft // 2], axis=1)

    plt.figure(figsize=(8, 4))
    plt.plot(positive_freqs, spectrum)
    plt.title(f"Spectrum (start_time={start_time}s, N={n_fft})")
    plt.xlabel("Frequency (Hz)")
    plt.ylabel("Magnitude")
    plt.grid(True)
    plt.show()


audio_file = "../assets/aa.ogg"
y, sr = librosa.load(audio_file)

plot_spectrum(y, sr, 0, 1024)
```

2 Vowels

2.1 Waveforms



2.2 Spectrograms

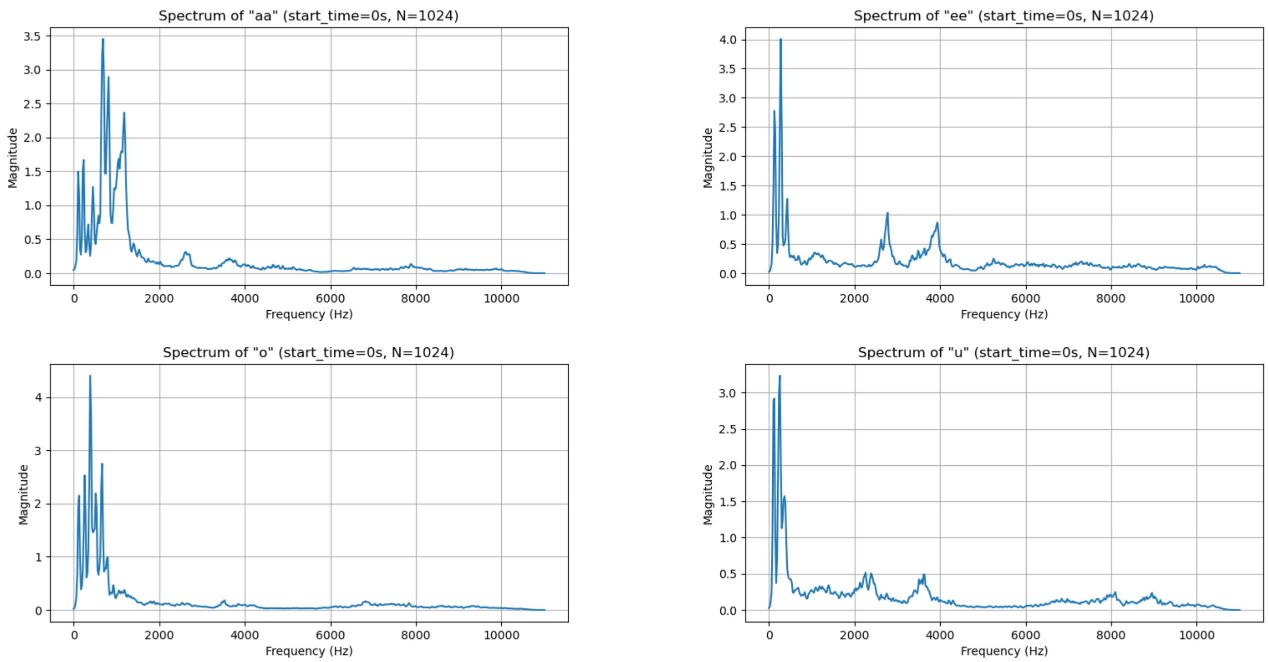


2.3 Spectrogram analysis

- In the spectrogram of "aa," you would see a darker band of energy concentrated in the lower frequency range, indicating the presence of a low-frequency fundamental and harmonics. This reflects the open mouth configuration of this vowel sound.

- The spectrogram of ”ee” would show a higher concentration of energy in the higher frequency range, indicating a higher-pitched sound with harmonics closer together, reflecting the constricted mouth configuration.
- The spectrogram of ”o” would exhibit energy distributed across a moderate frequency range, reflecting a moderately rounded mouth shape
- The spectrogram of ”u” would show a broad distribution of energy across frequencies, particularly in the mid to high-frequency range, suggesting a rounded and constricted mouth shape with a higher-pitched sound.

2.4 Spectrums

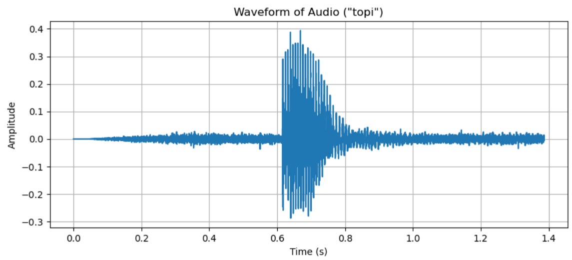


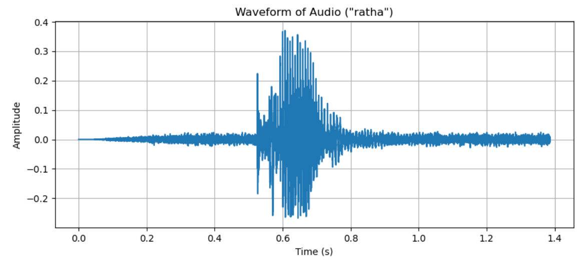
2.5 Spectrum analysis

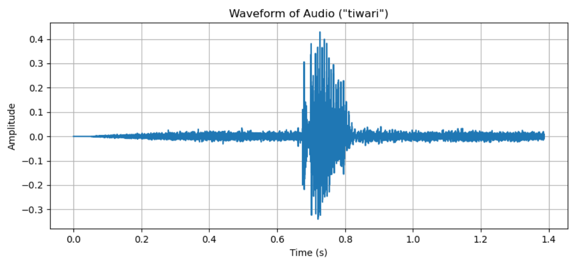
- The ”aa” vowel typically has a relatively low fundamental frequency, indicating a lower pitch. The harmonics in the spectrum are widely spaced and are stronger in the lower frequency range, suggesting a more open mouth configuration.
- The ”ee” vowel is characterized by a higher fundamental frequency, indicating a higher pitch. The harmonics in the spectrum are more closely spaced and stronger in the higher frequency range, indicating a more constricted mouth configuration.
- The ”o” vowel typically has a moderate fundamental frequency, with harmonics distributed across a moderate frequency range. It suggests a somewhat rounded mouth shape.
- The ”u” vowel has a relatively high fundamental frequency and harmonics that are evenly spaced across a wide frequency range, suggesting a rounded and constricted mouth shape.

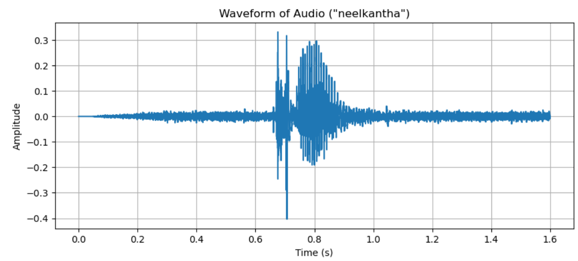
3 T-Sounds

3.1 Waveforms

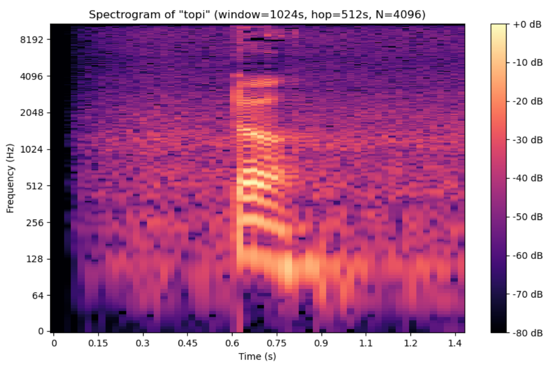


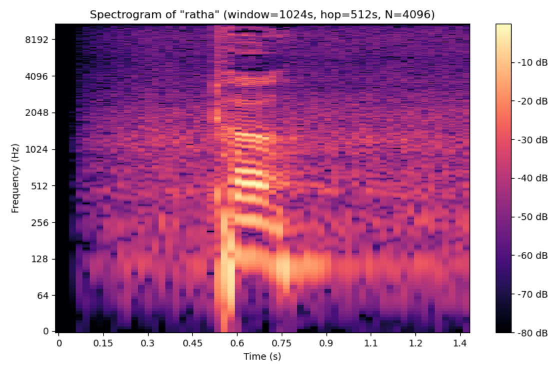


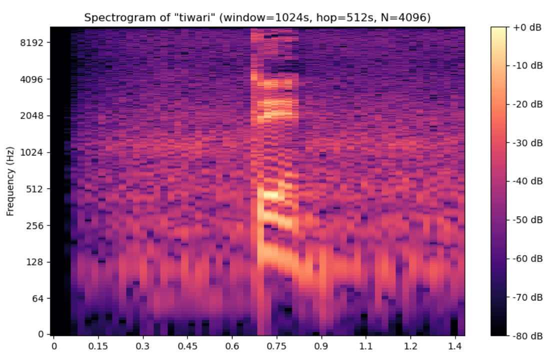


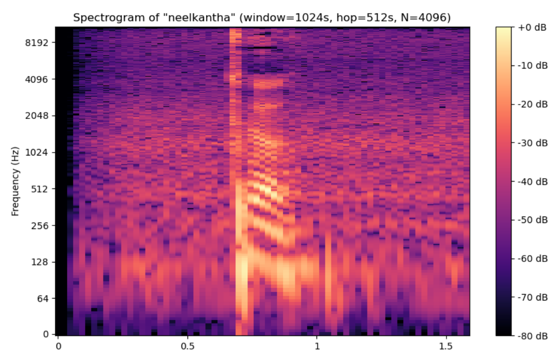


3.2 Spectrograms

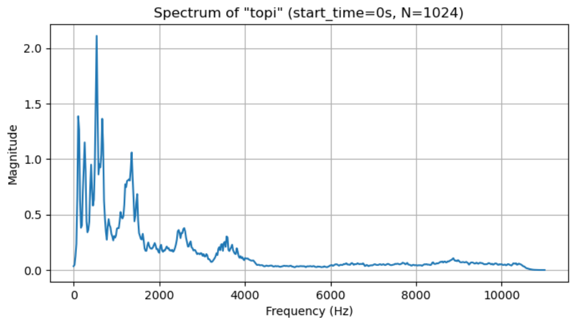


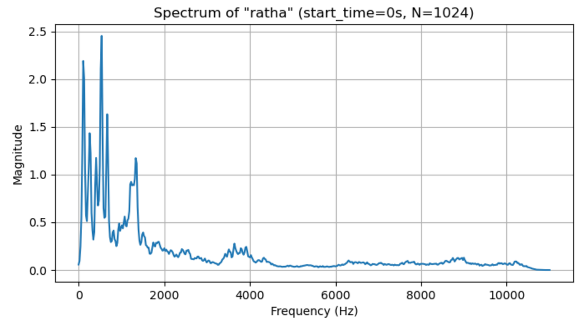


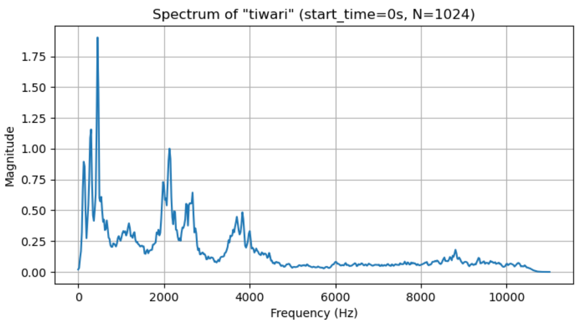


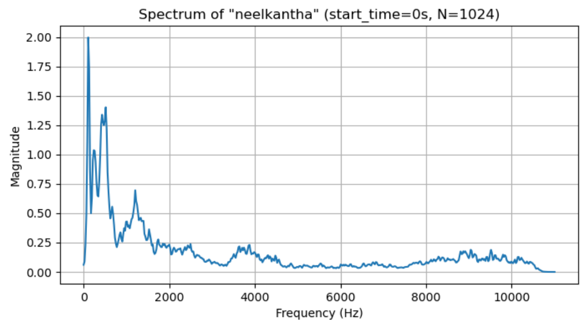


3.3 Spectrums









3.4 "T" Sound

- The "t" sound is a voiceless alveolar stop consonant. It is characterized by a brief interruption of airflow followed by a release of air. It does not have voicing, which means that the vocal cords do not vibrate during its production.
- The spectrogram of "t" would display a short burst of high-frequency energy, typically appearing as a narrow band of high-intensity activity. There would be no continuous harmonic structure, indicating a voiceless consonant.
- In the spectrum of the "t" sound, you would see a sharp peak in the high-frequency range, representing the burst of energy when the airflow is released after the closure. There is no significant voicing, so the lower-frequency harmonics are relatively weak or absent.

3.5 "Th" Sound

- The "th" sound is a voiceless dental fricative. Unlike the "t" sound, it involves a continuous airflow with some friction between the tongue and the upper front teeth, and it does not have voicing.
- The spectrogram of "th" would display a relatively continuous, high-frequency noise-like signal, indicating the ongoing friction. There would be no clear harmonics or periodicity.
- In the spectrum of the "th" sound, you would observe a continuous distribution of energy in the high-frequency range, reflecting the friction between the tongue and teeth. There would be no strong harmonics in the lower frequency range.

3.6 Word analysis

- In ”tiwari” and ”topi,” the ”t” sound is produced. In their spectra and spectrograms, you would see a sharp burst of high-frequency energy with no continuous harmonic structure.

- In ”ratha” and ”neelkanth,” the ”th” sound is produced. In their spectra and spectrograms, you would observe a continuous distribution of high-frequency energy without strong harmonics.

4 Scaling

4.1 dB Scaling

The dB scale is a logarithmic scale used to represent the magnitude or intensity of a signal. It’s often used in audio analysis to better capture the human perception of sound, as our ears are sensitive to relative changes in loudness rather than absolute power levels. When applied to spectra and spectrograms:

- When you represent the amplitude spectrum (magnitude of the frequency components) in dB, you’re essentially converting it from a linear scale to a logarithmic scale. This compresses the dynamic range, making it easier to visualize both weak and strong components. It’s particularly useful for visualizing spectral features more clearly.

- Using the dB scale in a spectrogram helps emphasize the relative loudness of different frequency components over time. It can reveal subtle details in the audio, especially in regions with low energy. For example, it can highlight quiet background noise or faint harmonics in a musical instrument.

4.2 Mel Scale

The mel scale is a perceptual scale that models the non-linear way in which humans perceive pitch and frequency. It’s particularly useful for capturing the tonal characteristics of sound. When applied to spectra and spectrograms:

- Converting the frequency scale to the mel scale helps align spectral components with human auditory perception. It redistributes the frequency bins so that they are more closely spaced in lower-frequency regions (where human perception is more sensitive) and less closely spaced in higher-frequency regions. This makes it easier to extract features related to pitch and timbre.

- A spectrogram with a mel-frequency scale (mel-spectrogram) provides a more perceptually meaningful representation of audio. It’s commonly used in speech and music analysis, as it focuses on the regions of the spectrum that are relevant to human perception. Features extracted from a mel-spectrogram, such as mel-frequency cepstral coefficients (MFCCs), are widely used in speech and audio processing tasks.

4.3 Visualization

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


def plot_mel_spectrogram(y, sr, n_fft, hop_length, n_mels):
    mel_spec = librosa.feature.melspectrogram(
        y=y, sr=sr, n_fft=n_fft, hop_length=hop_length, n_mels=n_mels
    )
    mel_spec_db = librosa.power_to_db(mel_spec, ref=np.max)

    plt.figure(figsize=(10, 6))
    librosa.display.specshow(
        mel_spec_db, sr=sr, hop_length=hop_length, x_axis="time", y_axis="mel"
    )
    plt.colorbar(format="%+2.0f dB")
    plt.title("Mel Spectrogram (dB)")
    plt.xlabel("Time (s)")
    plt.ylabel("Mel Frequency (Hz)")
    plt.show()


audio_file = "../assets/aa.ogg"
y, sr = librosa.load(audio_file)

plot_mel_spectrogram(y, sr, 2048, 512, 128)
```

The mel spectrogram of different vowels ("aa", "ee", "o" and "u") obtained using the above code are as follow: