Speech Emotion Recognition - Week 1
Signal Processing Fundamentals

Krishna Kukreja & Garvit Meena

November 2025

Week 1: Resources & Guide

1. Setup

1. Install Python 3.8+

2. Create virtual environment (optional):
   - Create: `python3 -m venv ser_env`
   - Activate (Linux/Mac): `source ser_env/bin/activate`
   - Activate (Windows): `ser_env\Scripts\activate`

3. Install required libraries:
   `pip install librosa scipy numpy pandas matplotlib seaborn jupyter`

2. Reading Materials

- **Sampling Theorem**: GeeksforGeeks - Sampling and Aliasing

- **FFT Explained**: 3Blue1Brown - But what is the Fourier Transform?

- **MFCC Theory**: GeeksforGeeks - MFCC for Speech Recognition

- **Interactive FFT Tutorial**: Better Explained - Interactive Guide to FFT

3. Video Tutorials

- MFCC Calculation with Librosa Python

- Valerio Velardo - Audio Processing Basics Playlist

- Extract Features from Audio File Tutorial

4. Dataset

- **RAVDESS Audio Dataset**: Download from Kaggle (approx. 450MB zip)

- **Note**: Do not use the Zenodo version as the directory structure differs.

- **Action**: Download and extract to: `data/raw/RAVDESS/`

5. Python Code Templates

Load Audio

```python
import librosa
import numpy as np

# librosa converts audio into a numpy array
# sr=22050 resamples the audio to 22,050 Hz to ensure consistency
y, sr = librosa.load('audio.wav', sr=22050)

print(f"Duration: {len(y)/sr:.2f}s")
print(f"Sampling rate: {sr} Hz")
# Checks for normalization/too quiet audio
print(f"Max amplitude: {np.max(np.abs(y)):.4f}")
```

Extract MFCC

```python
import librosa

mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=13)
delta = librosa.feature.delta(mfcc)
delta_delta = librosa.feature.delta(delta)

print(f"MFCC shape: {mfcc.shape}")
print(f"Delta shape: {delta.shape}")
```

Generate Spectrogram

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

S = librosa.feature.melspectrogram(y=y, sr=sr)
S_db = librosa.power_to_db(S, ref=np.max)

# X-axis: Time (seconds)
# Y-axis: Frequency (Mel scale)
# Color: Loudness (dB)
plt.figure(figsize=(10, 4))
librosa.display.specshow(S_db, sr=sr, x_axis='time', y_axis='mel')
plt.colorbar(format='%+2.0f dB')
plt.title('Mel-Spectrogram')
plt.tight_layout()
plt.show()
```

6. Assignments

1. Load 5 audio files, print their properties (sample rate, duration, max amplitude, RMS energy).

2. Extract MFCC from 3 different emotions, plot side-by-side.

3. Generate spectrograms for happy vs. sad, compare frequency patterns.

7. Expected Outputs

- Python environment working with all libraries installed.

- RAVDESS dataset downloaded and organized in `data/raw/RAVDESS/`.

- 3 Jupyter notebooks/scripts with assignments completed.

- 3+ visualization plots saved (spectrograms, MFCC plots).

- **GitHub**: Add everything to your repository under folder `Week-1` and write a README.md.

8. Quick Links

- Librosa Official Tutorial

- SciPy Signal Processing Documentation

- Matplotlib Tutorials

Appendix: Dataset Naming Convention

The following figure details the 7-part numerical identifier used for every file in the RAVDESS dataset, decoding the emotion, intensity, and actor details.

File naming convention

Each of the 7356 RAVDESS files has a unique filename. The filename consists of a 7-part numerical identifier (e.g., 02-01-06-01-02-01-12.mp4). These identifiers define the stimulus characteristics:

Filename identifiers

- Modality (01 = full-AV, 02 = video-only, 03 = audio-only).
- Vocal channel (01 = speech, 02 = song).
- Emotion (01 = neutral, 02 = calm, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = surprised).
- Emotional intensity (01 = normal, 02 = strong). NOTE: There is no strong intensity for the 'neutral' emotion.
- Statement (01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door").
- Repetition (01 = 1st repetition, 02 = 2nd repetition).
- Actor (01 to 24. Odd numbered actors are male, even numbered actors are female).

Filename example: 02-01-06-01-02-01-12.mp4

1. Video-only (02)
2. Speech (01)
3. Fearful (06)
4. Normal intensity (01)
5. Statement "dogs" (02)
6. 1st Repetition (01)
7. 12th Actor (12)
8. Female, as the actor ID number is even.

Figure 1: RAVDESS File Naming Convention identifying modality, emotion, intensity, and actor.