Assignment 2

Spectral Analysis on human emotion of happy and sad

Data:

https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio

Select 500 audios files (~approx ratio of 250:250) with below Filename identifiers

- Modality (03 = audio-only).
- Vocal channel (01 = speech).
- Emotion (03 = happy, 04 = sad).
- 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).

For more details on dataset please refer the About Dataset section of Kaggle link

Task:

1. Select 3 or 4 audio files of happy and sad, generate following plots and explain the inferences in 1-2 lines,
   a. Waveform
   b. Spectrogram on log
   c. Zero Crossing rates
   d. Spectral Centroids
   e. Spectral Bandwidth
   f. MFCC
2. Extract the MFCC features with 40 components from those 500 audios and build simple logistic regression. Comment on the model output performance. (Try to adjust MFCC components).
3. Extract the mel spectrogram features with 128 mel bands from those 500 audios and build simple logistic regression. Comment on the model output performance.
4. Compare model results of MFCC and Mel Spectrogram