



SPEECH EMOTION RECOGNITION

GOURAV VERMA

DSC-680 T302, SPRING-21

APPLIED DATA SCIENCE

PROF. FADI ALSALEEM

BELLEVUE UNIVERCITY



TABLE OF CONTENTS



Data



Audio Signal



Audio Features



Data
Preparation



Methodology



Results



DATA

- Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)
 - **1440** Audio Files
 - **24** North American professional actors (12M/ 12F)
 - **8 Emotions:** Neutral, Calm, Happy, Sad, Angry, Fearful, Surprise, & Disgusted
- **Filename example: 03-01-06-01-02-01-12.wav**
 1. Audio-only (03)
 2. Speech (01)
 3. Fearful (06)
 4. Normal intensity (01)
 5. Statement "dogs" (02)
 6. 1st Repetition (01)
 7. 12th Actor (12)

Female, as the actor ID number is even.

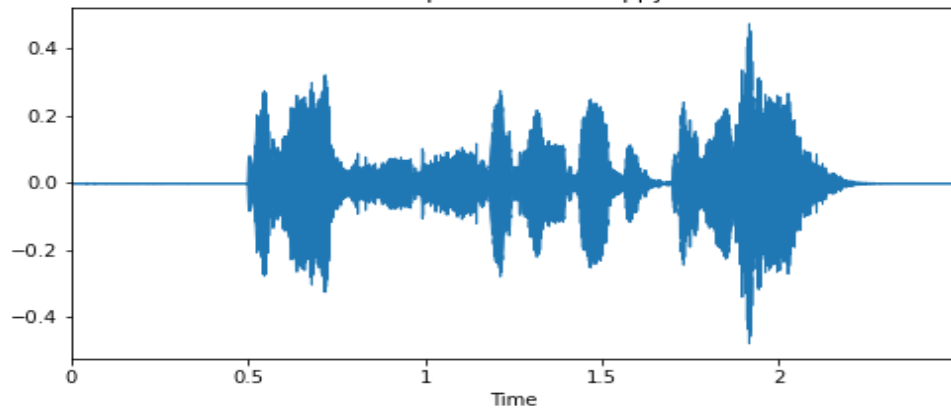


AUDIO SIGNAL

Female Happy with strong emotion



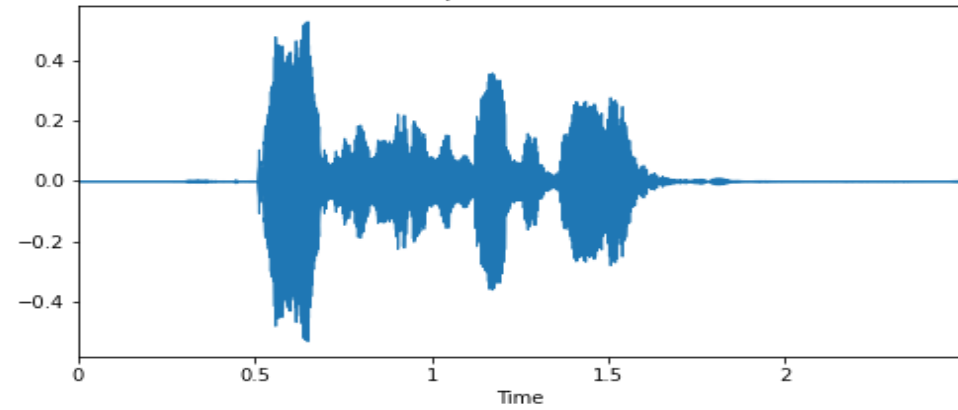
Waveplot - Female Happy



Male Fearful with strong emotion

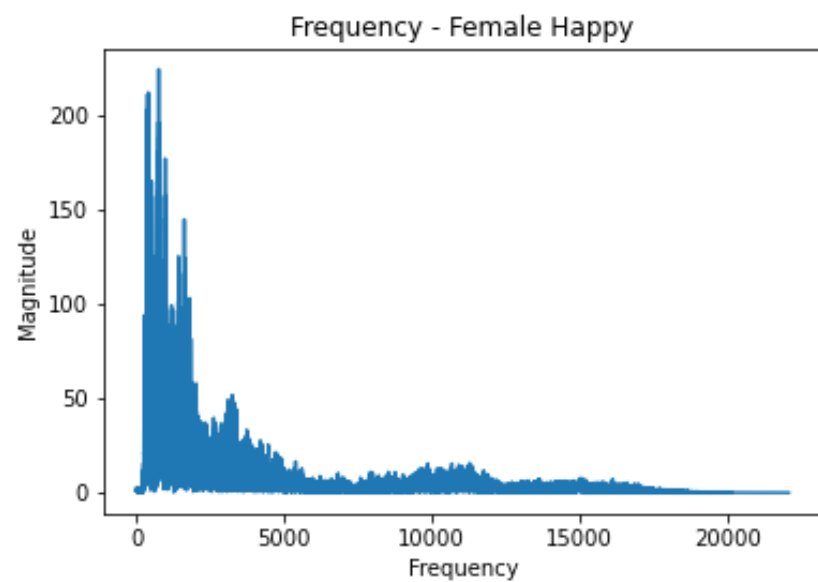


Waveplot - Male Fearful

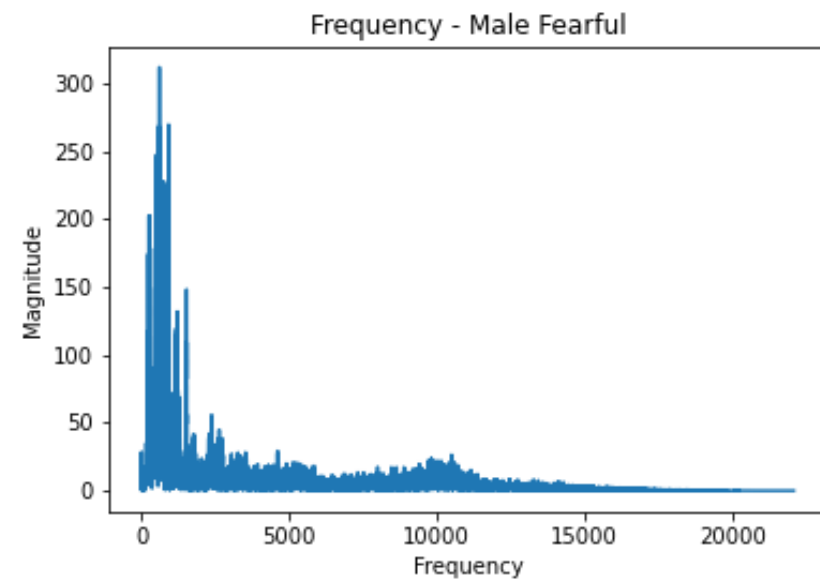


AUDIO SIGNAL

Female Happy with strong emotion

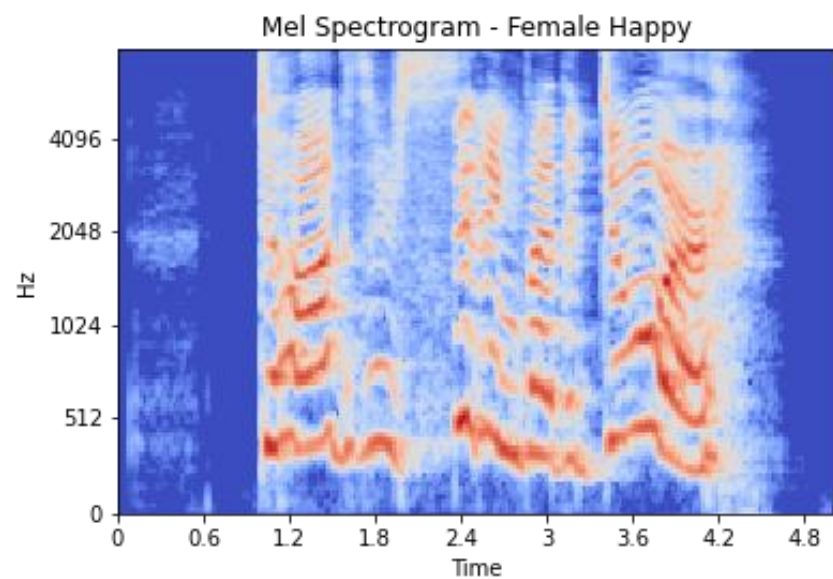


Male Fearful with strong emotion

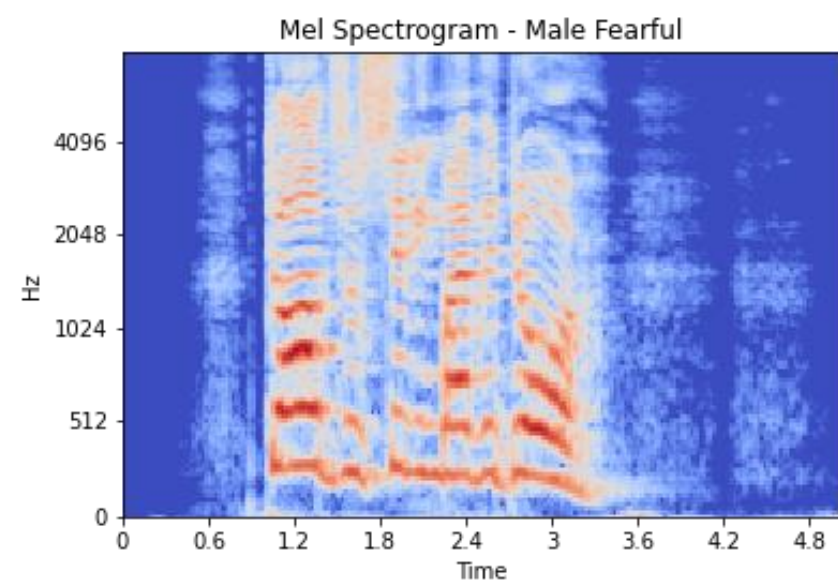


AUDIO SIGNAL FEATURES: *MEL SPECTROGRAM*

Female Happy with strong emotion

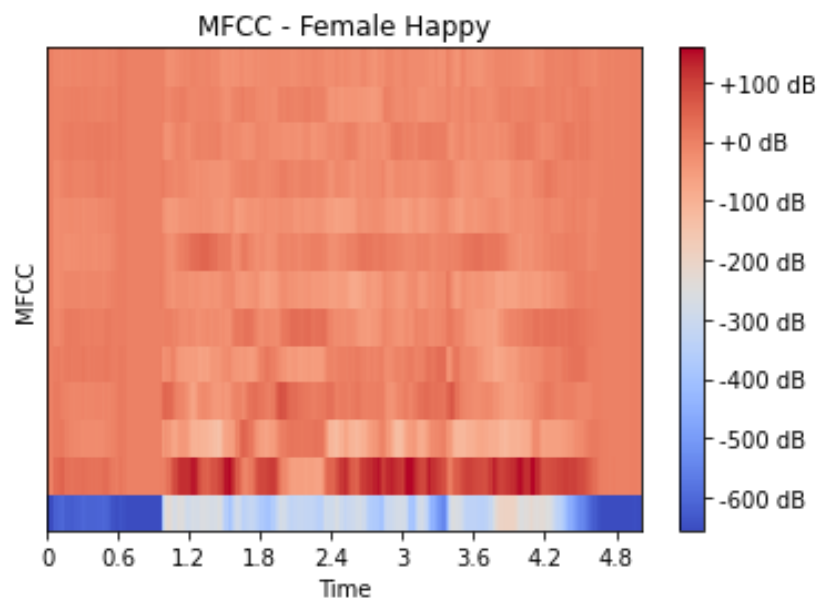


Male Fearful with strong emotion

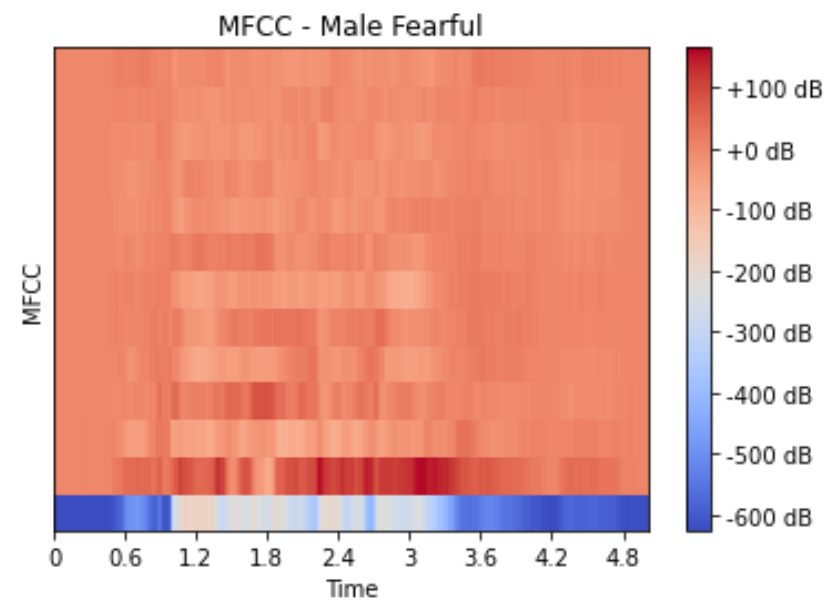


AUDIO SIGNAL FEATURES: *MEL-FREQUENCY CEPSTRAL COEFFICIENTS*

Female Happy with strong emotion



Male Fearful with strong emotion

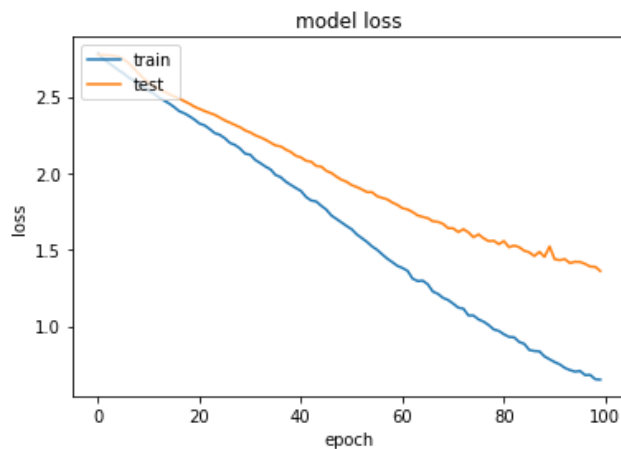
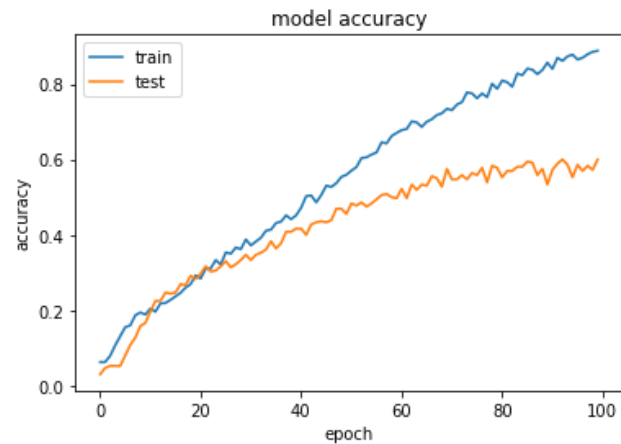


DATA PREPARATION

- Loaded all audio files into dataframe
- Created/Merged Gender and Emotion column
 - 16 gender-emotion labels (2 x 8)
- Extracted 40 MFCCs
- Added to dataframe as separate columns.

	label	MFCC1	MFCC2	MFCC3	MFCC4	MFCC5	MFCC6	MFCC7	MFCC8	MFCC9	...	MFCC31	MFCC32	N
0	male_neural	-677.117188	87.630119	3.973252	15.096436	8.129905	16.422251	-5.018030	4.110128	-5.980066	...	-0.382024	-4.446791	-1.
1	male_neural	-664.351196	89.269028	0.765047	16.066084	10.971747	16.490419	-6.770354	5.740640	-7.474891	...	-1.732326	-5.006553	-0.



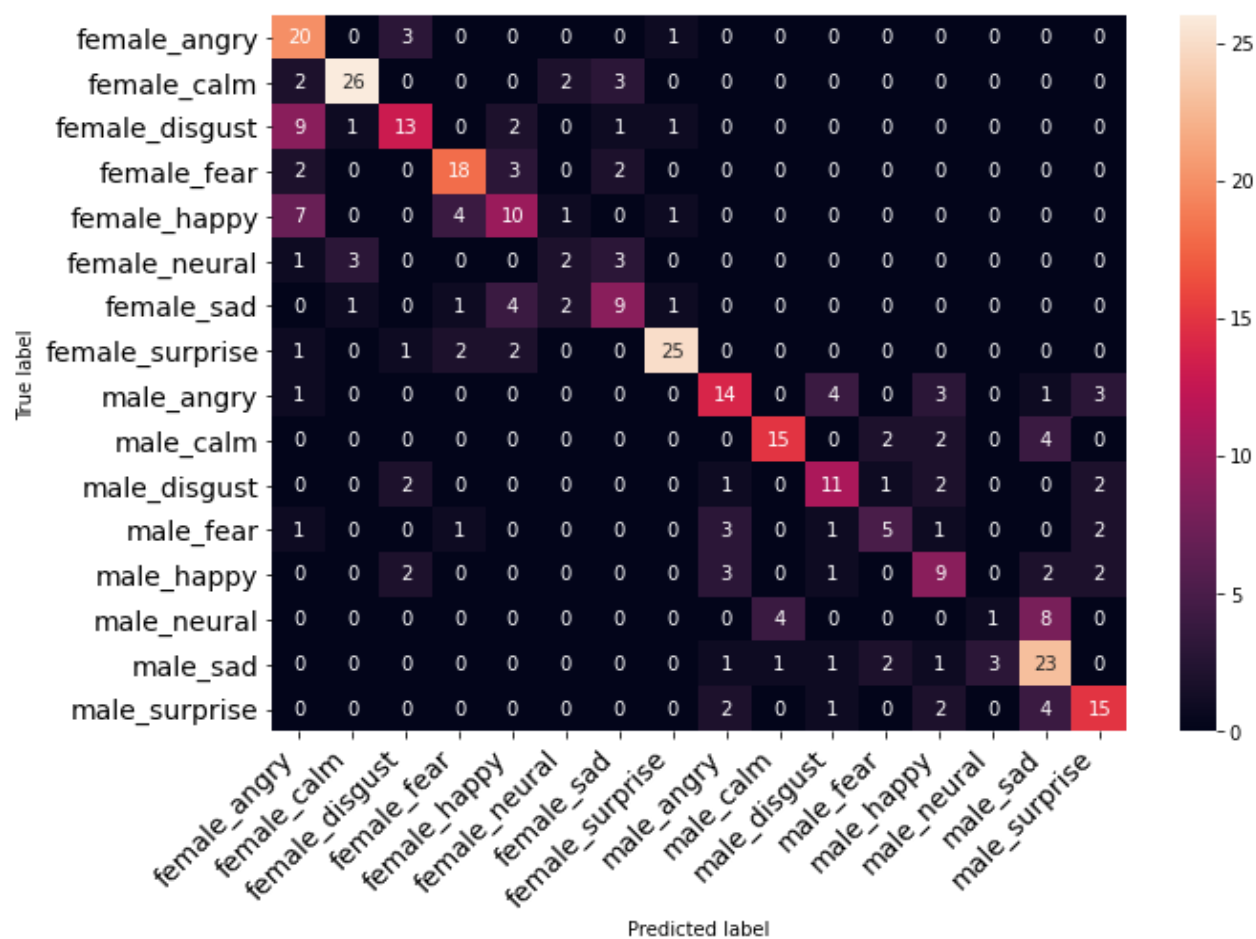


METHODOLOGY

- Librosa package for EDA and Feature Extraction
- Train-Test Split: 75%-25%
- 8-Layer 1D CNN
 - Trained on 100 epochs, batch size 16
 - Adam Optimizer
 - Accuracy 60%
- Future
 - Other Engineered Audio Features
 - 2D CNN, HMM



OUTCOME & PREDICTION



	actualvalues	predictedvalues
250	male_angry	male_surprise
251	male_sad	male_sad
252	male_angry	male_angry
253	male_calm	male_calm
254	male_angry	male_angry
255	female_disgust	female_disgust
256	male_sad	male_happy
257	female_surprise	female_surprise
258	female_calm	female_calm
259	male_calm	male_calm





THANK YOU

