



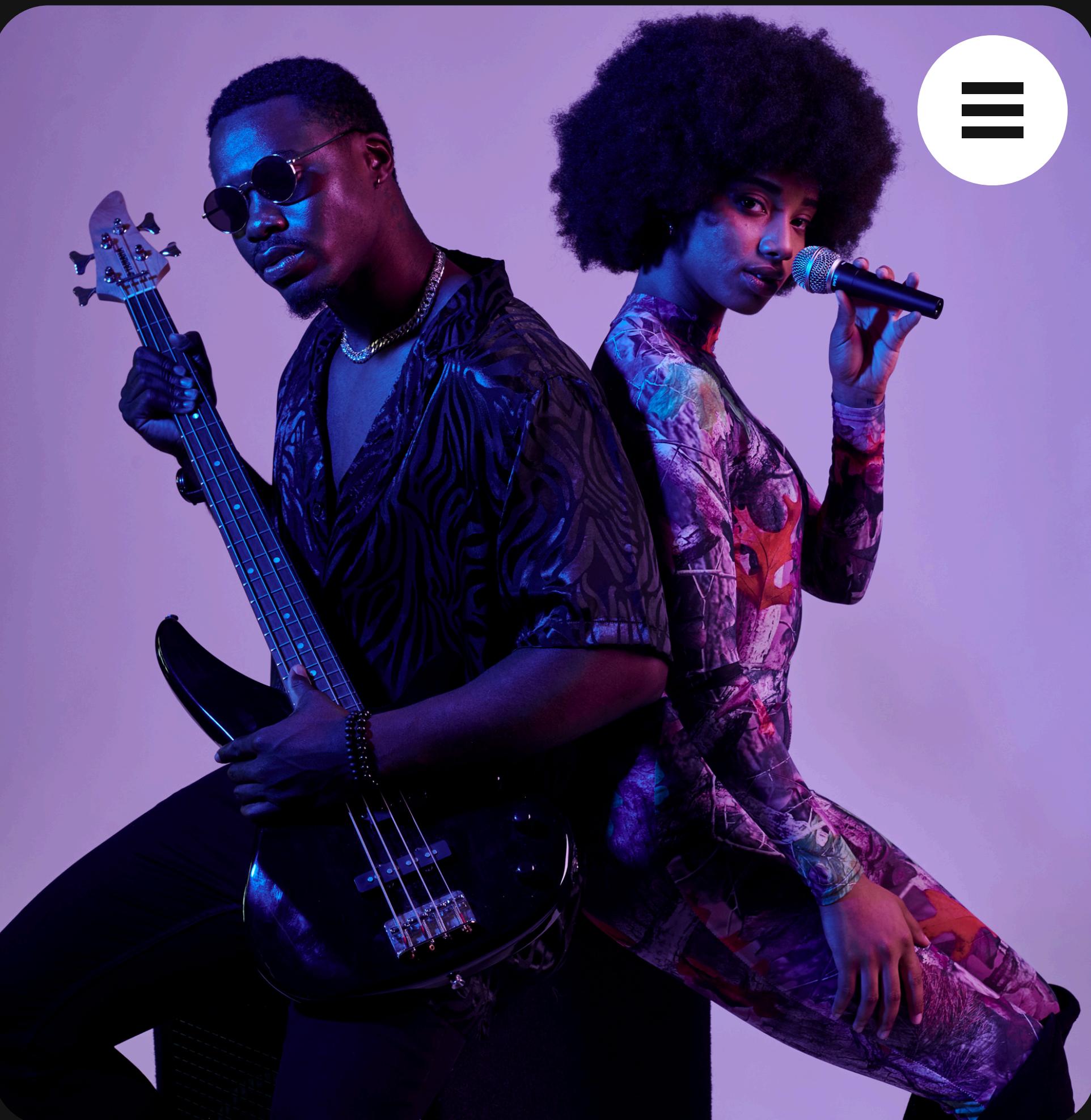
↗

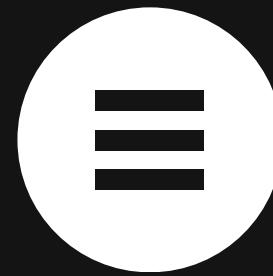
CAPSTONE PROJECT

GENRE CLASSIFICATION USING AUDIO DATA

An exploration into differences between genres, and the use of Convolutional Neural Networks (CNN)

NEXT

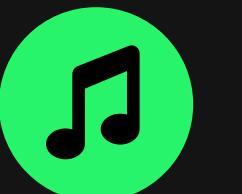




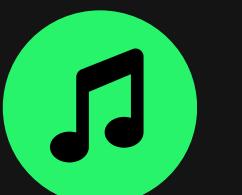
THE INSPIRATION



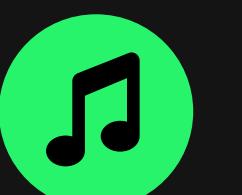
24K MAGIC
Bruno Mars



Funk?

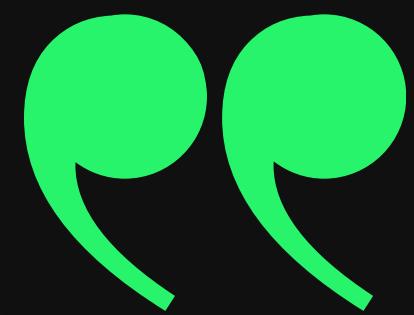


Disco?



Contemporary
R&B?

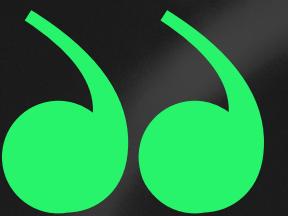


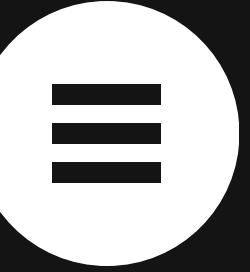


THE TASK

To classify a song's genre
using its audio data.

Accuracy > 10%

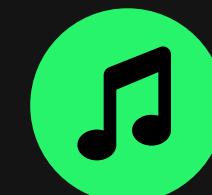




SOURCING THE DATA



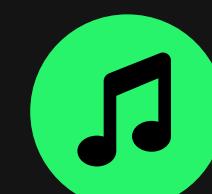
Kaggle
GTZAN Dataset -
Music Genre
Classification



10 genres

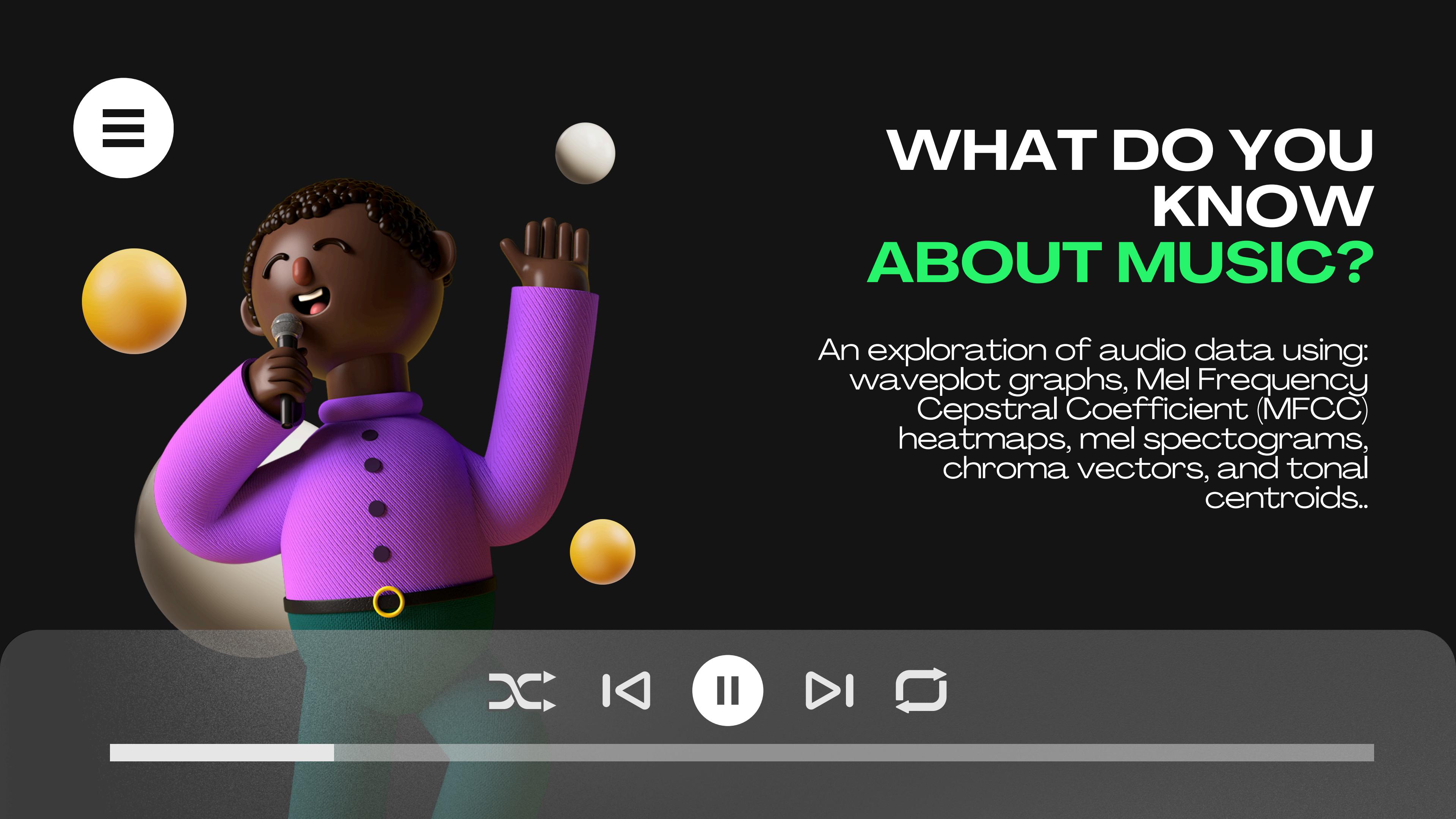


100 audio files each



30 second audio files





WHAT DO YOU KNOW ABOUT MUSIC?

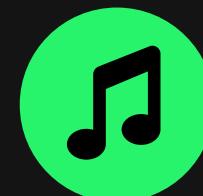
An exploration of audio data using:
waveplot graphs, Mel Frequency
Cepstral Coefficient (MFCC)
heatmaps, mel spectrograms,
chroma vectors, and tonal
centroids..



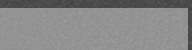
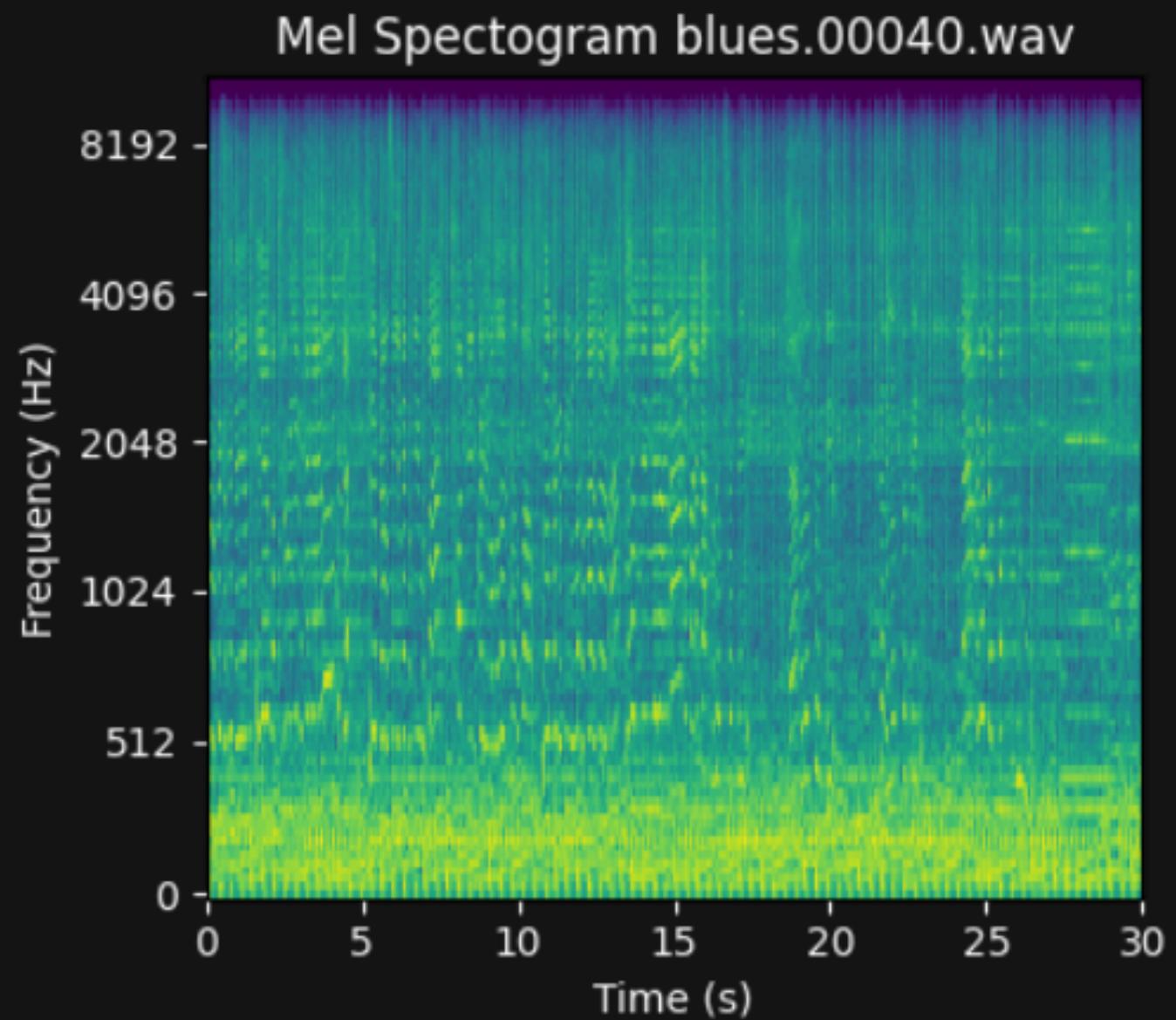
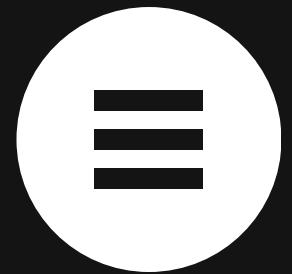
MEL SPECTROGRAMS

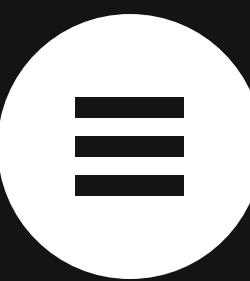
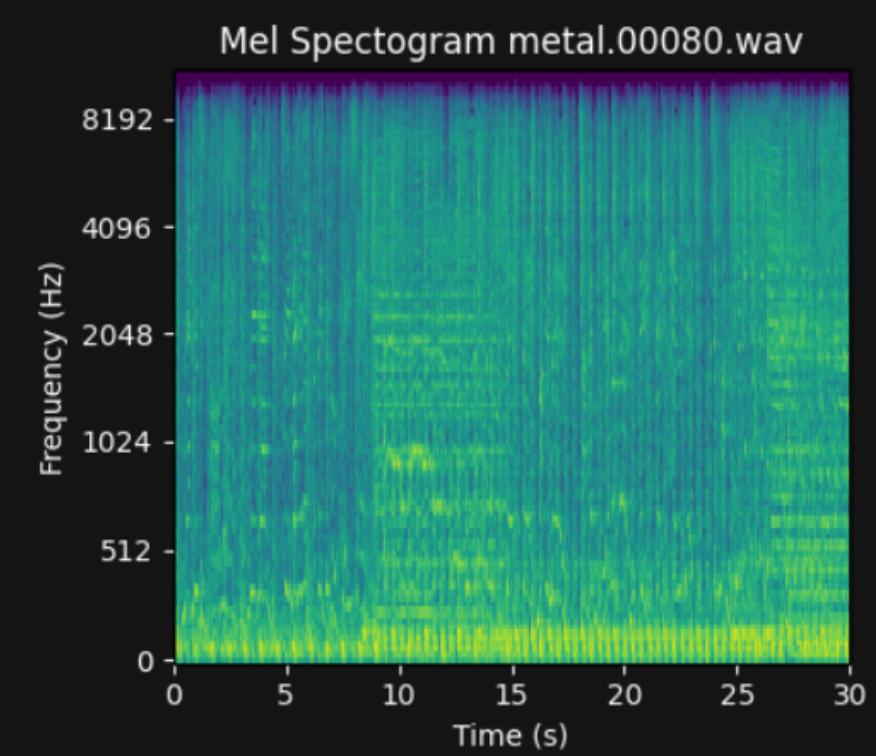
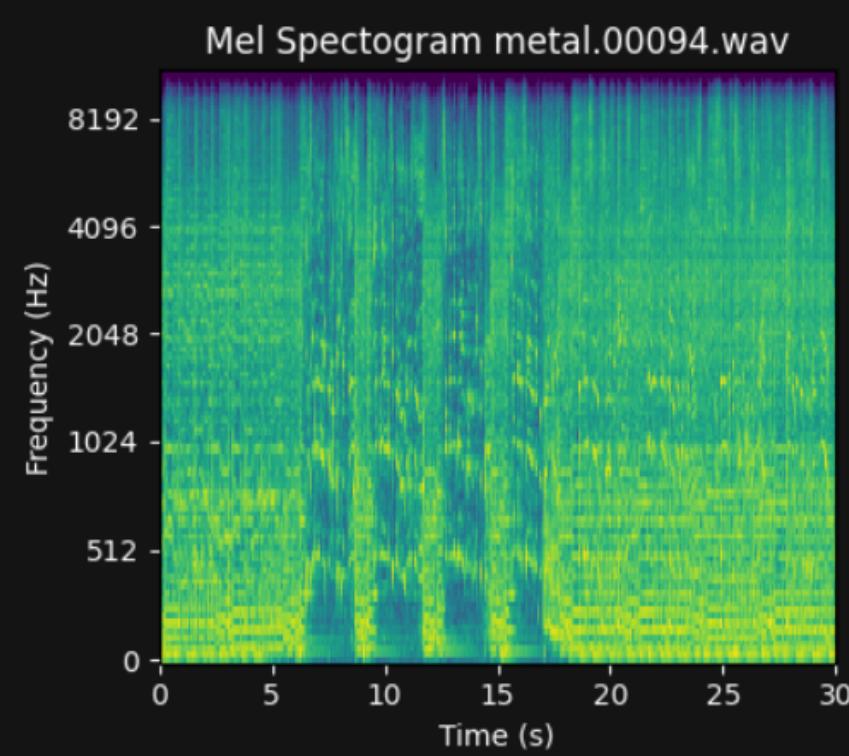
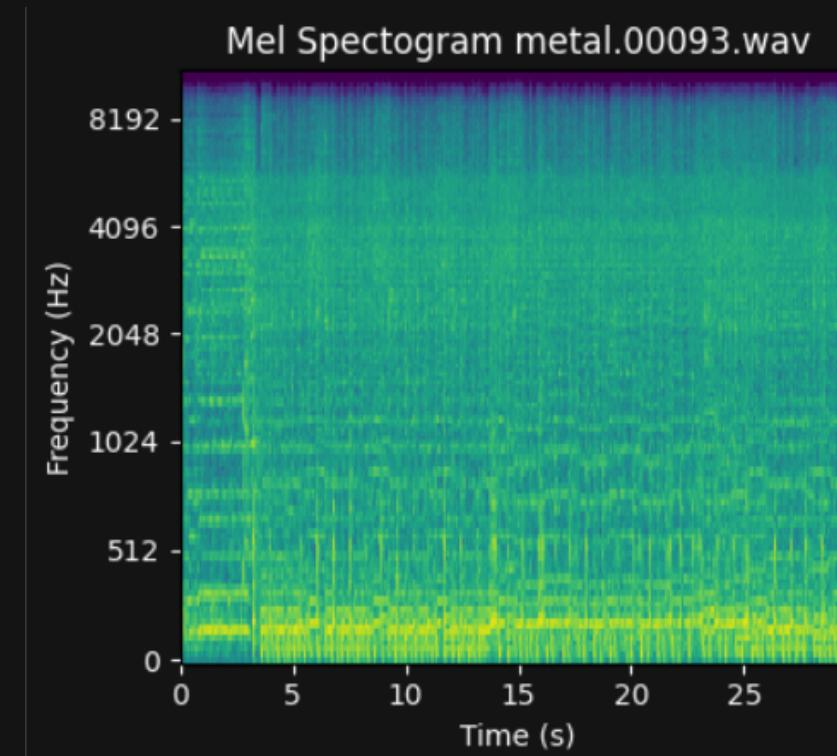
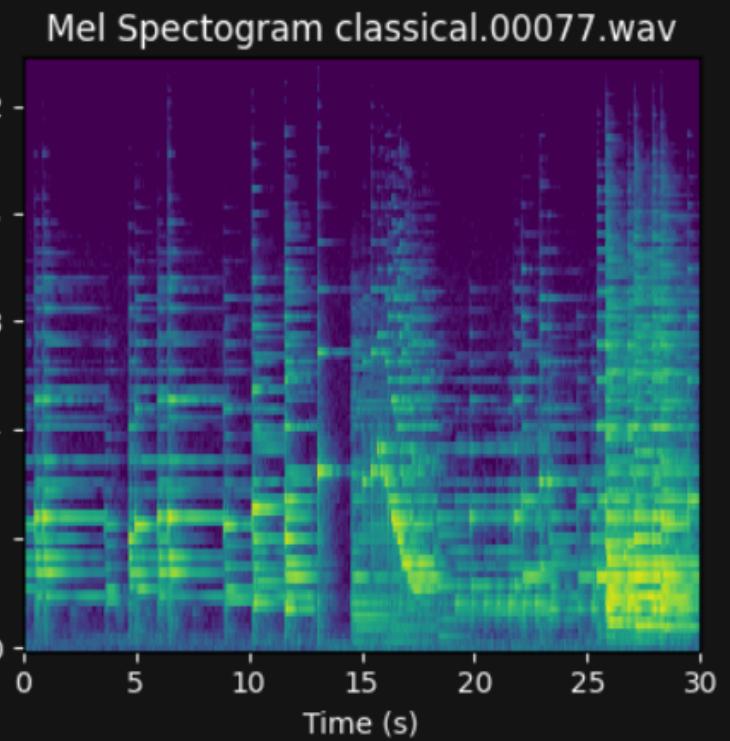
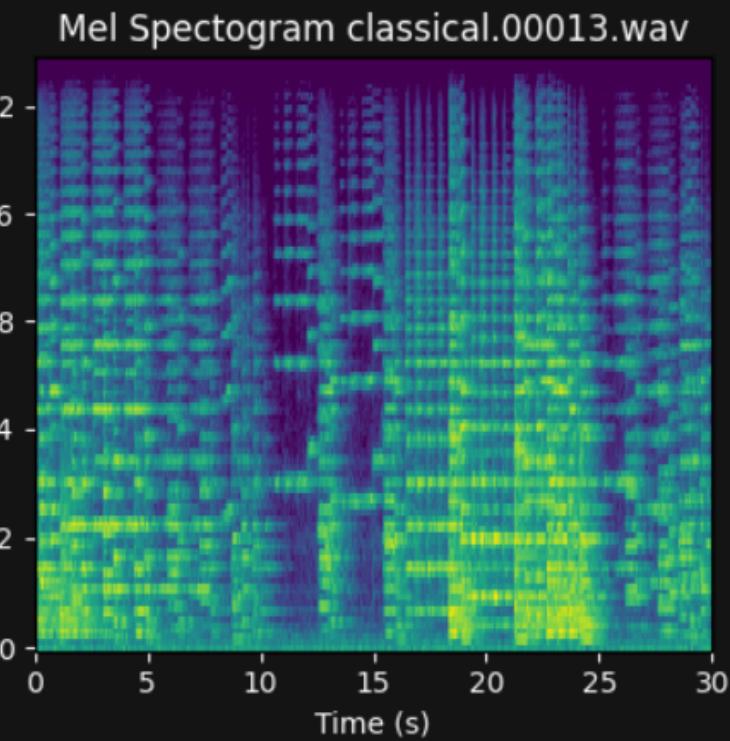
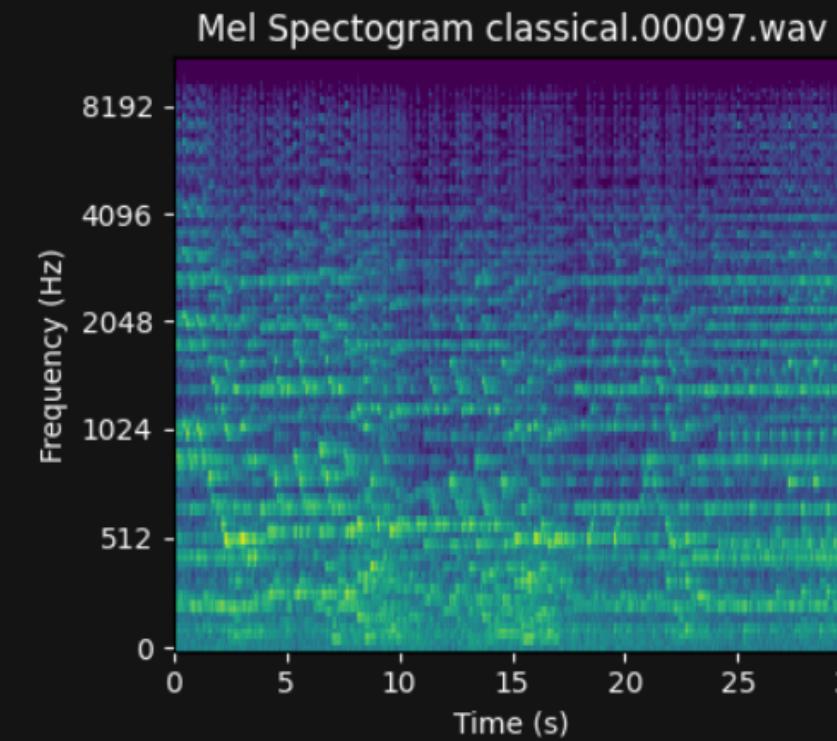


Frequency vs time

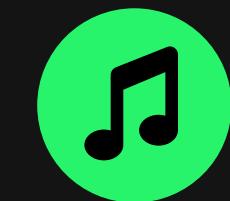
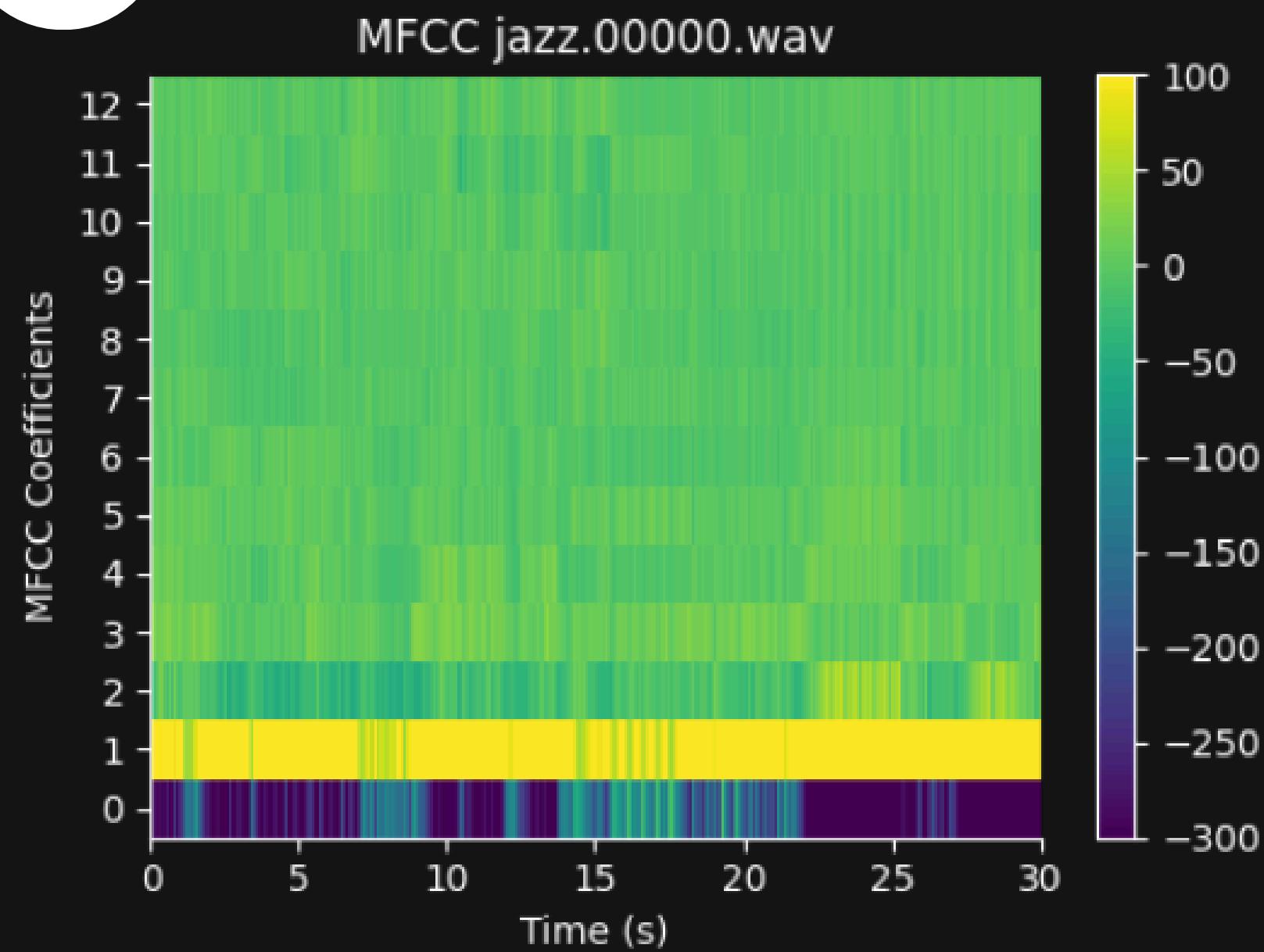
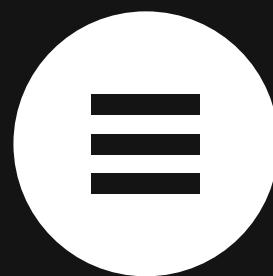


Mel scale





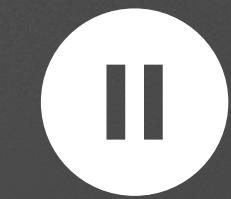
MFCC HEATMAPS

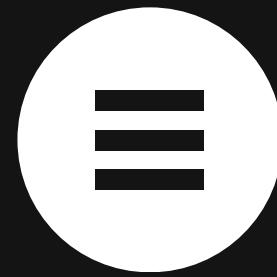


Lower coefficients: broad spectral shape of the signal. Most related to the timbre of the sound.

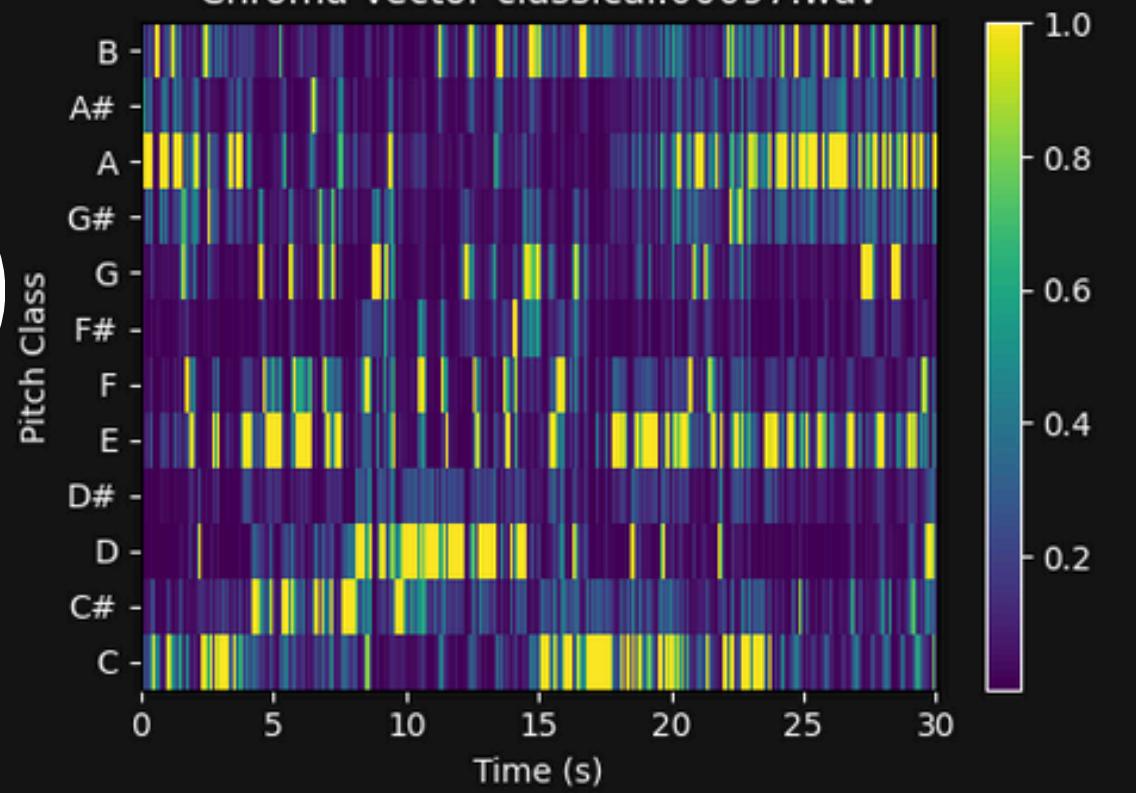


Higher coefficients: finer detail of the spectral envelope. Capture harmonics and melodies.

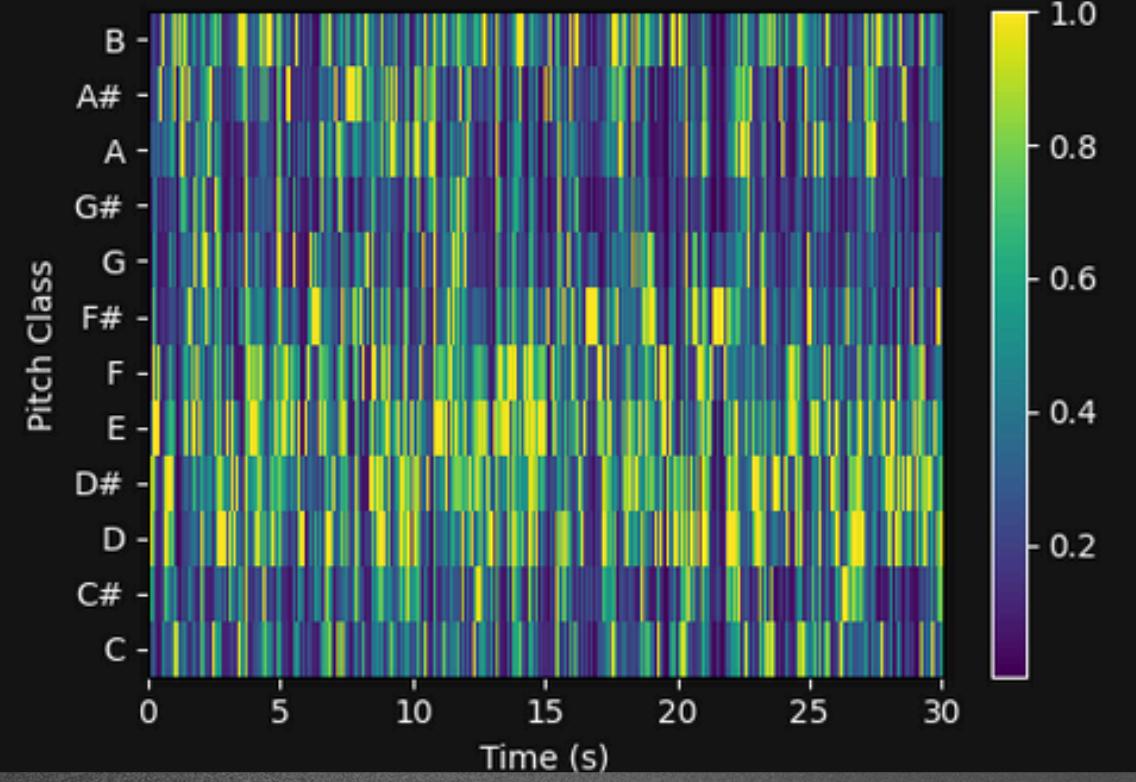




Chroma Vector classical.00097.wav



Chroma Vector disco.00048.wav



CHROMA VECTORS



Represents pitches
in the music



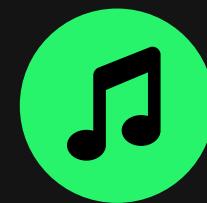
Pitch - how high or
low a sound is



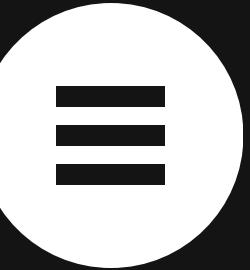
TONAL CENTROIDS



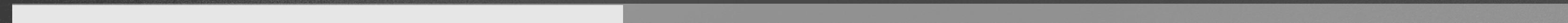
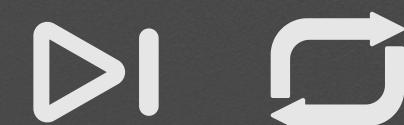
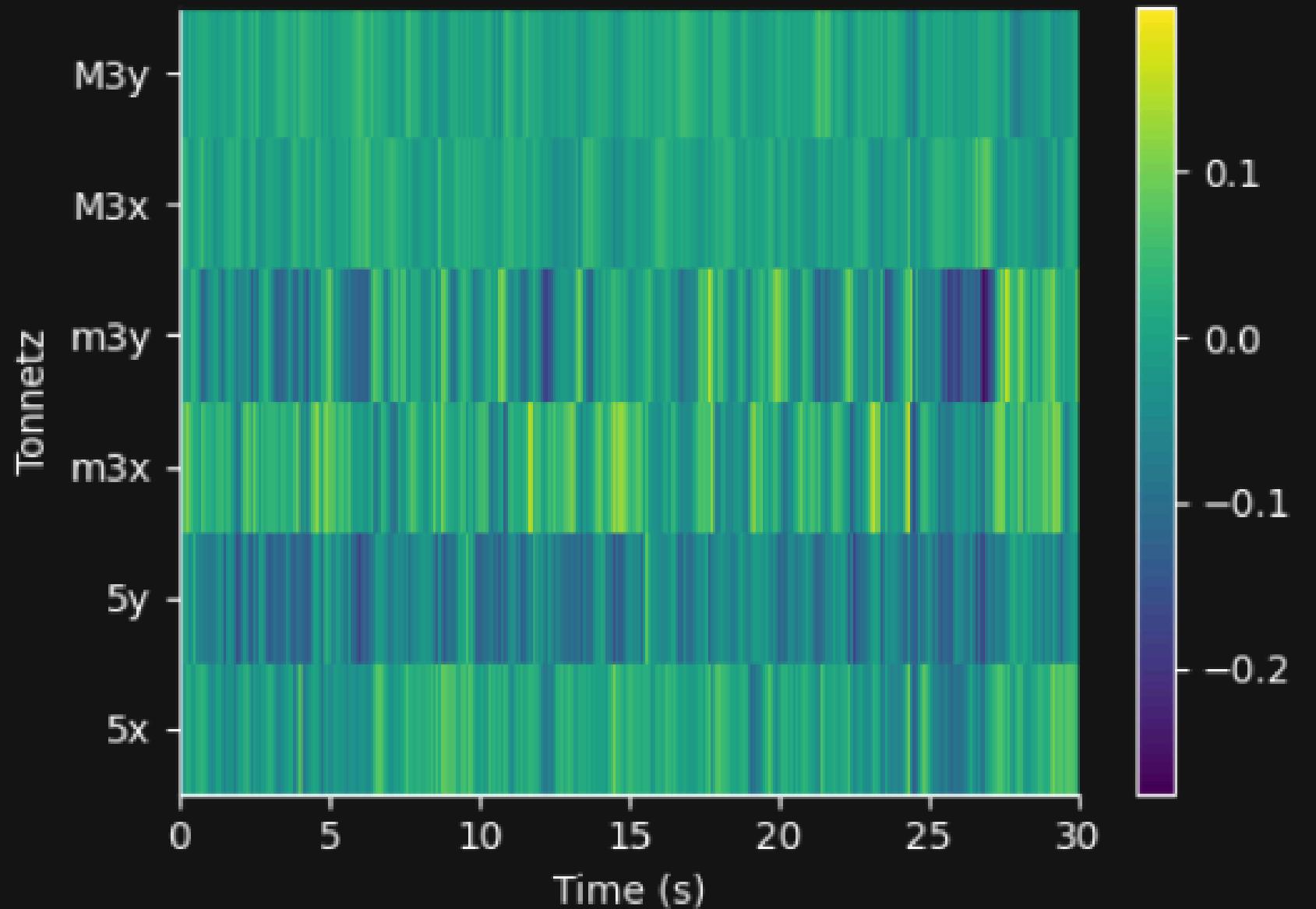
6 pitch classes



Represent perfect
fifth, minor third, and
major third



Tonnetz pop.00009.wav

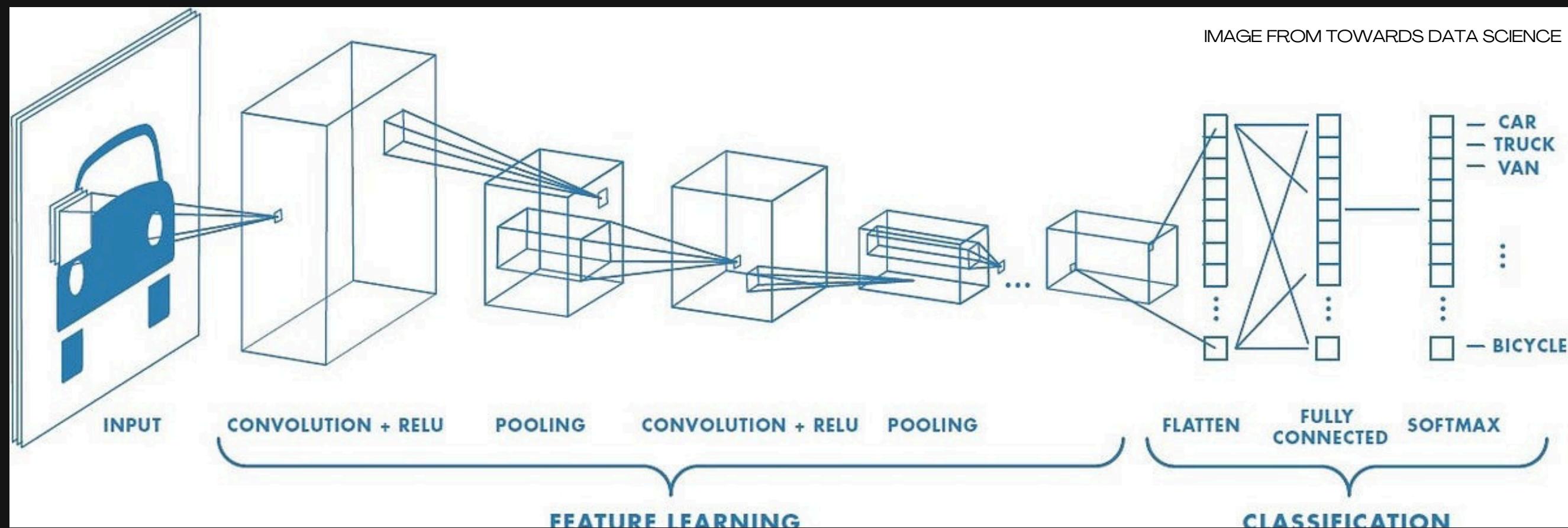
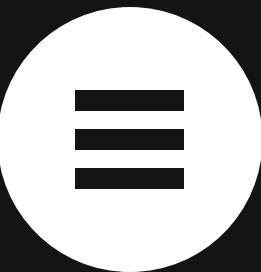




MODELLING



CONVOLUTIONAL NEURAL NETWORKS



Three main types of layers:

- Convolutional
- Pooling
- Fully connected



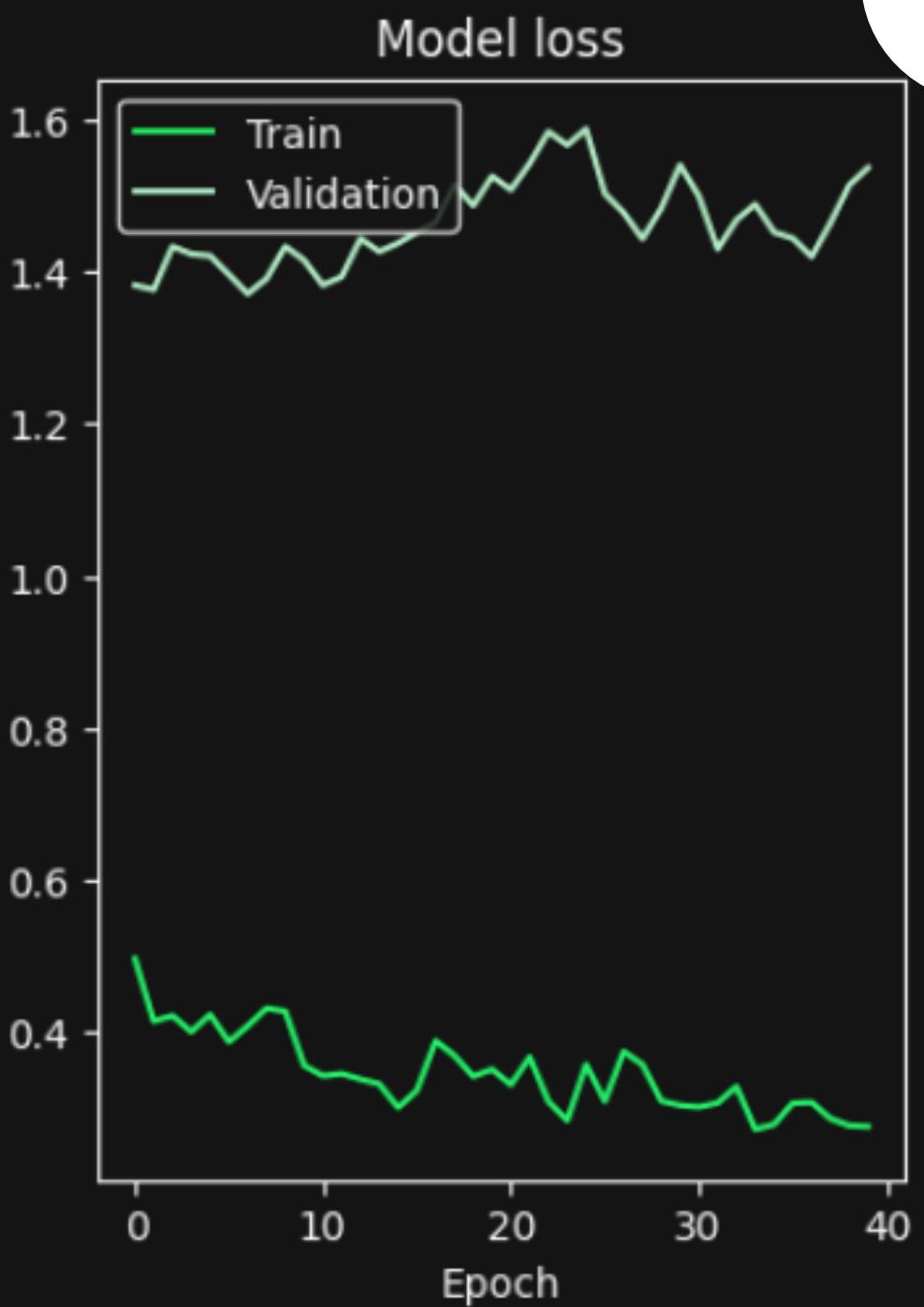
Model: "sequential"

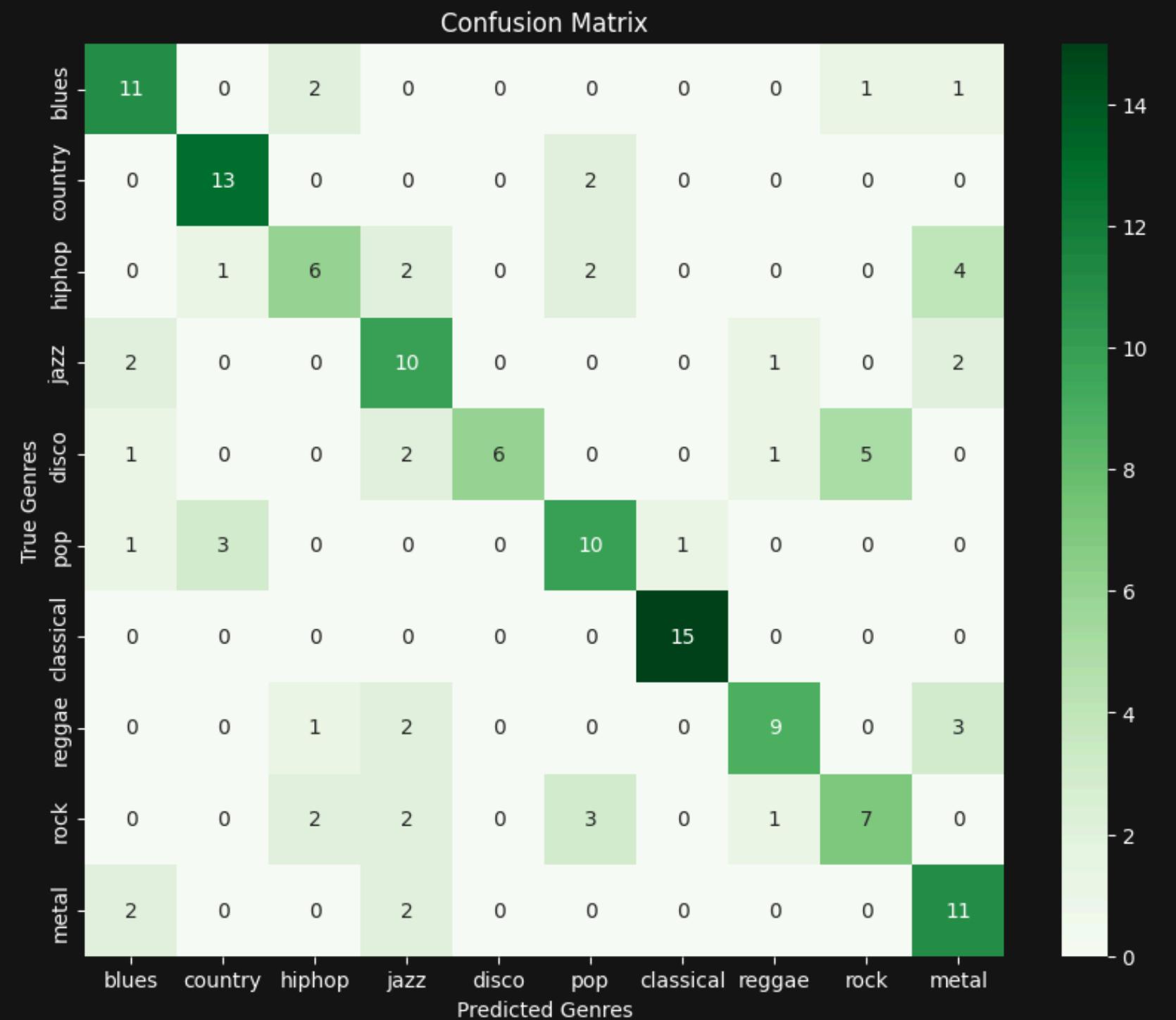
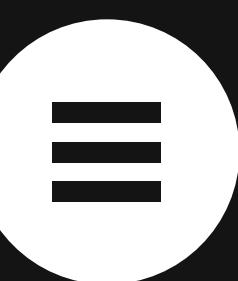
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 558, 1, 32)	320
max_pooling2d (MaxPooling2D)	(None, 279, 1, 32)	0
dropout (Dropout)	(None, 279, 1, 32)	0
conv2d_1 (Conv2D)	(None, 279, 1, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 139, 1, 64)	0
dropout_1 (Dropout)	(None, 139, 1, 64)	0
conv2d_2 (Conv2D)	(None, 139, 1, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 69, 1, 128)	0
dropout_2 (Dropout)	(None, 69, 1, 128)	0
flatten (Flatten)	(None, 8832)	0
dense (Dense)	(None, 512)	4522496
dropout_3 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 10)	5130

Total params: 4620298 (17.63 MB)

Trainable params: 4620298 (17.63 MB)

Non-trainable params: 0 (0.00 Byte)





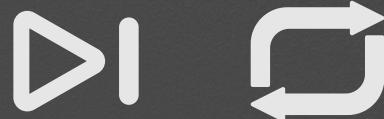
Good at predicting classical and country

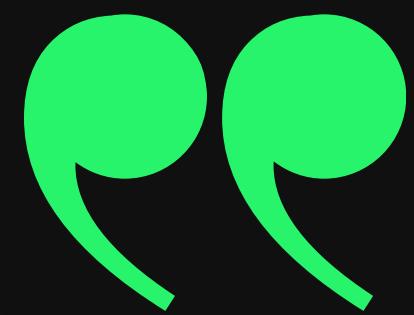


Not so good at predicting hiphop and disco



Could combine some genres to raise accuracy score

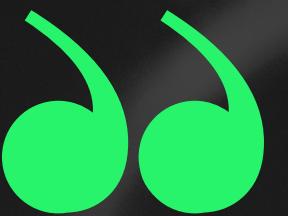




THE TASK

To classify a song's genre
using its audio data.

Accu. 65% 10%



TIME FOR THE **DEMO**





IF I WERE TO DO IT AGAIN

- Train on more data: Million Song Dataset, Indian Music Genre Dataset
- Play with more CNN layers
- Input the full MFCC heatmaps, mel spectograms, etc.
- Try different models, e.g. KNN or LSTM models



THANKS FOR LISTENING

ANY QUESTIONS?

