Slotify

An ensemble model for music genre classification

FMA Dataset

106,574

Tracks in entire dataset

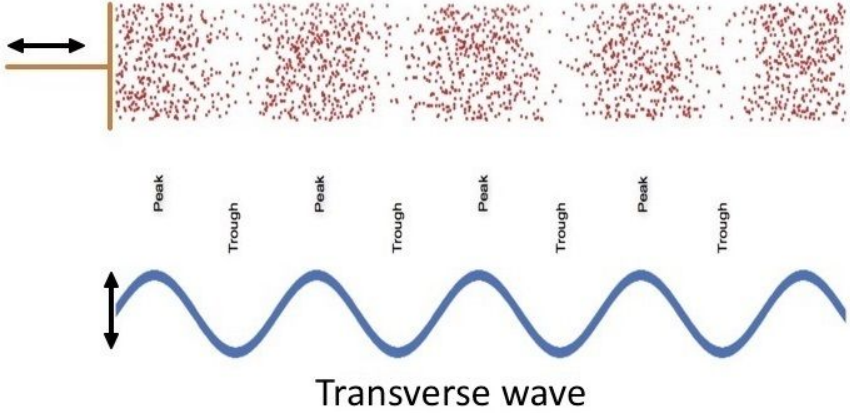
8,000

Tracks in small subset summing

7.2 GB

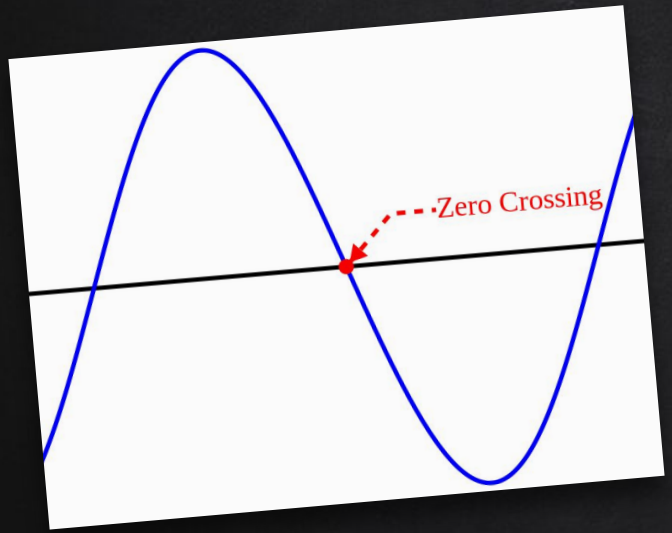
Hip-Hop
Pop
Folk
Experimental

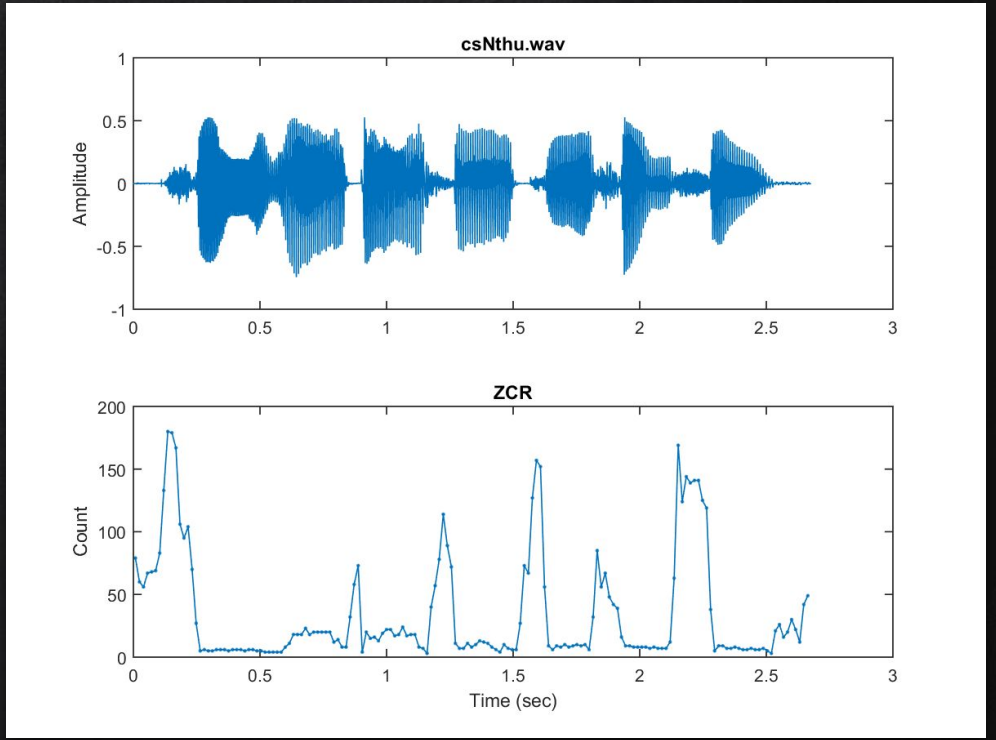
Rock
International
Electronic
Instrumental



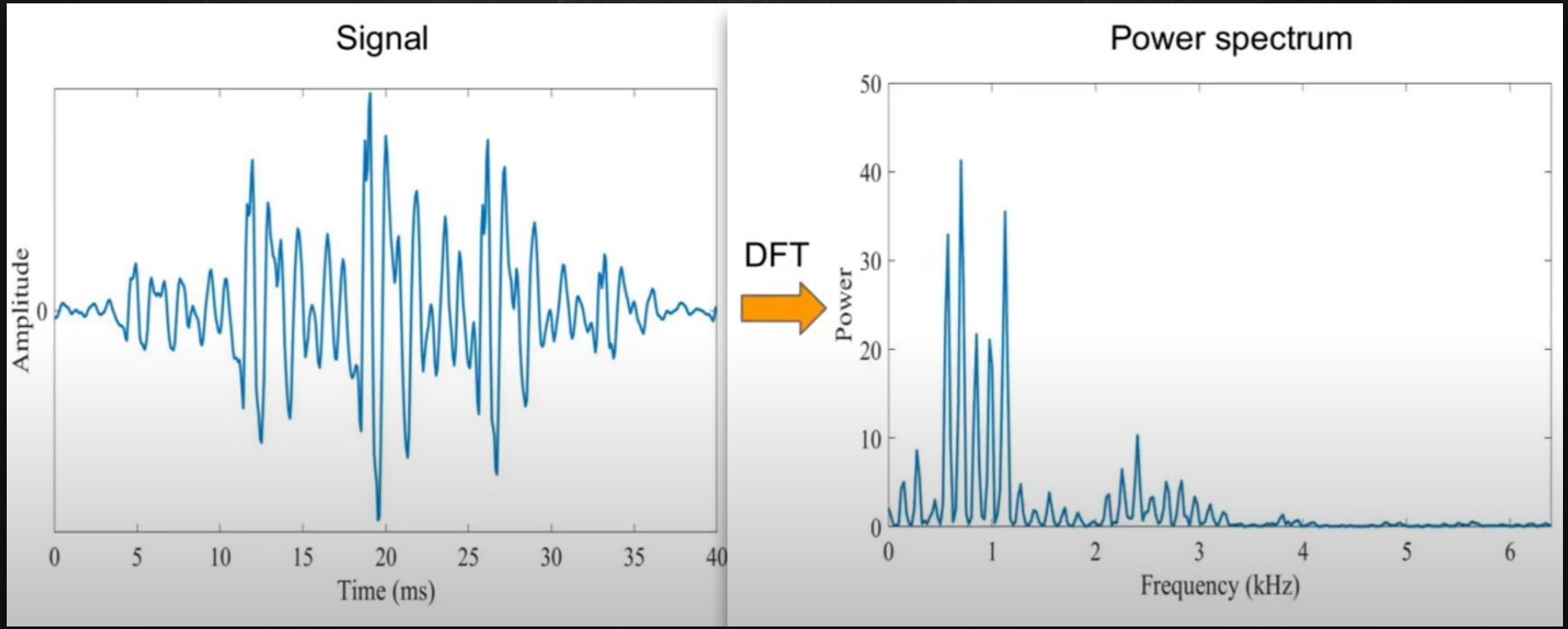
© The University of Waikato Te Whare Wānanga o Waikato I www.sciencelearn.org.nz

Feature: Zero Crossing Rate



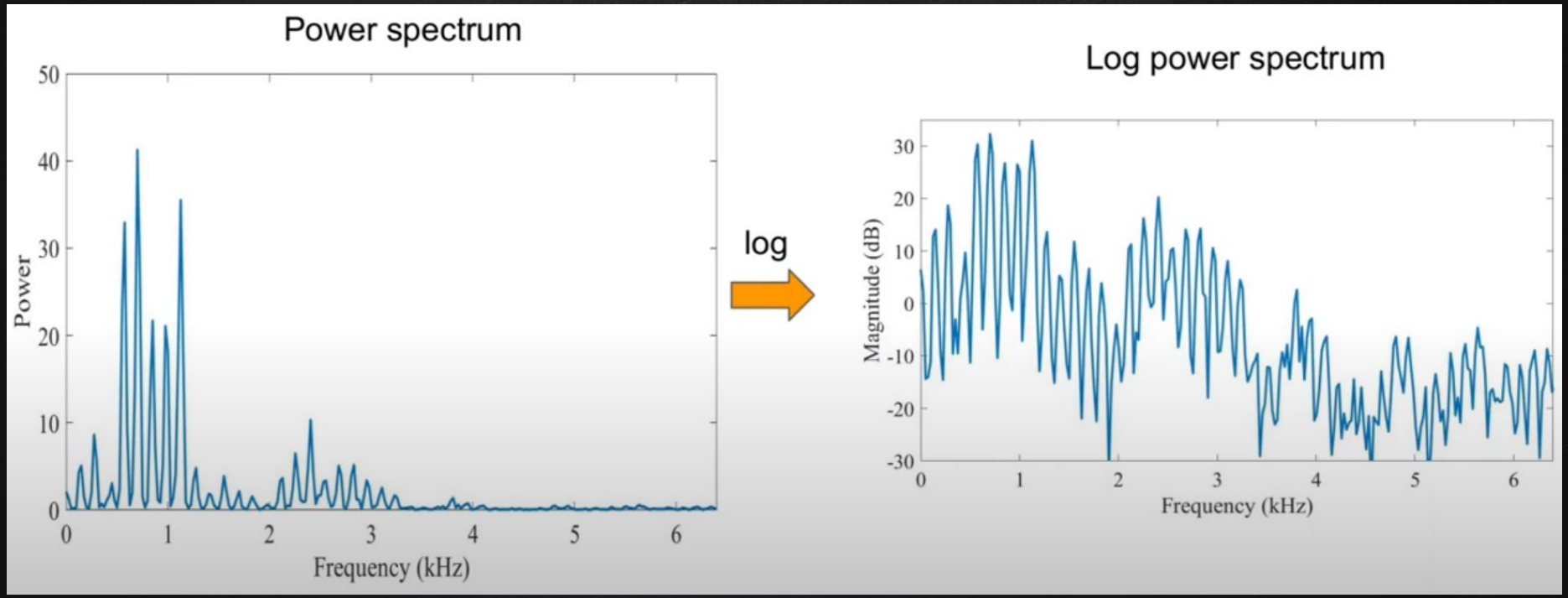


Feature: Mel-Frequency Cepstrum Coefficients



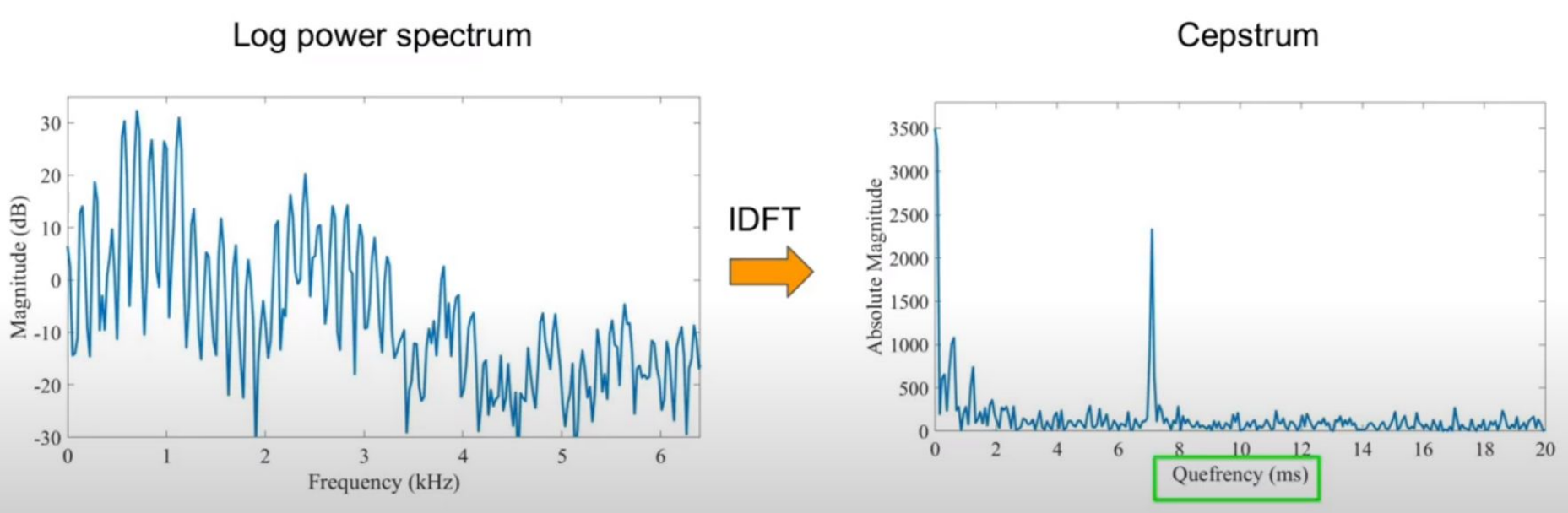
Source: "Mel-Frequency Cepstral Coefficients Explained Easily" Valerio Velardo
The Sound of AI [https://www.youtube.com/watch?v=4_SH2nfbQZ8]

Feature: Mel-Frequency Cepstrum Coefficients



Source: "Mel-Frequency Cepstral Coefficients Explained Easily" Valerio Velardo
The Sound of AI [https://www.youtube.com/watch?v=4_SH2nfbQZ8]

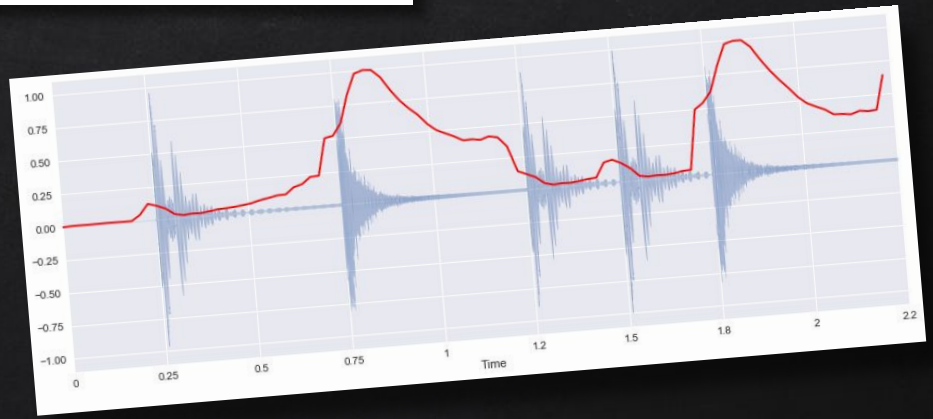
Feature: Mel-Frequency Cepstrum Coefficients



Source: "Mel-Frequency Cepstral Coefficients Explained Easily" Valerio Velardo
The Sound of AI [https://www.youtube.com/watch?v=4_SH2nfbQZ8]

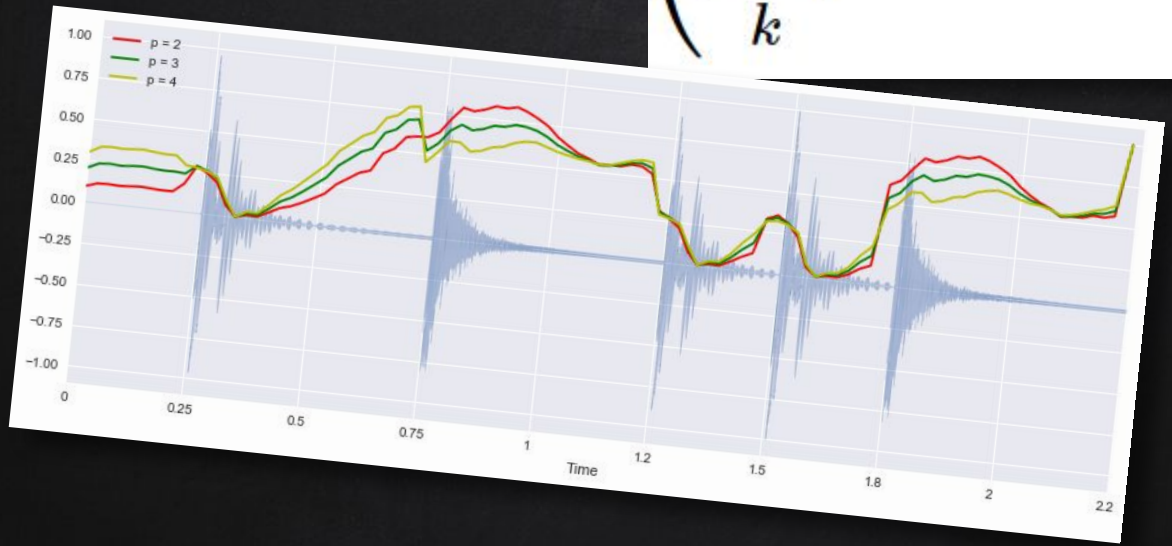
Spectral Centroid

$$f_c = \frac{\sum_k S(k) f(k)}{\sum_k S(k)}$$



Spectral Bandwidth

$$\left(\sum_k S(k)(f(k) - f_c)^p \right)^{\frac{1}{p}}$$



Spectral Contrast

$$Peak_k = \log\{\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x'_{k,i}\} \quad (1)$$

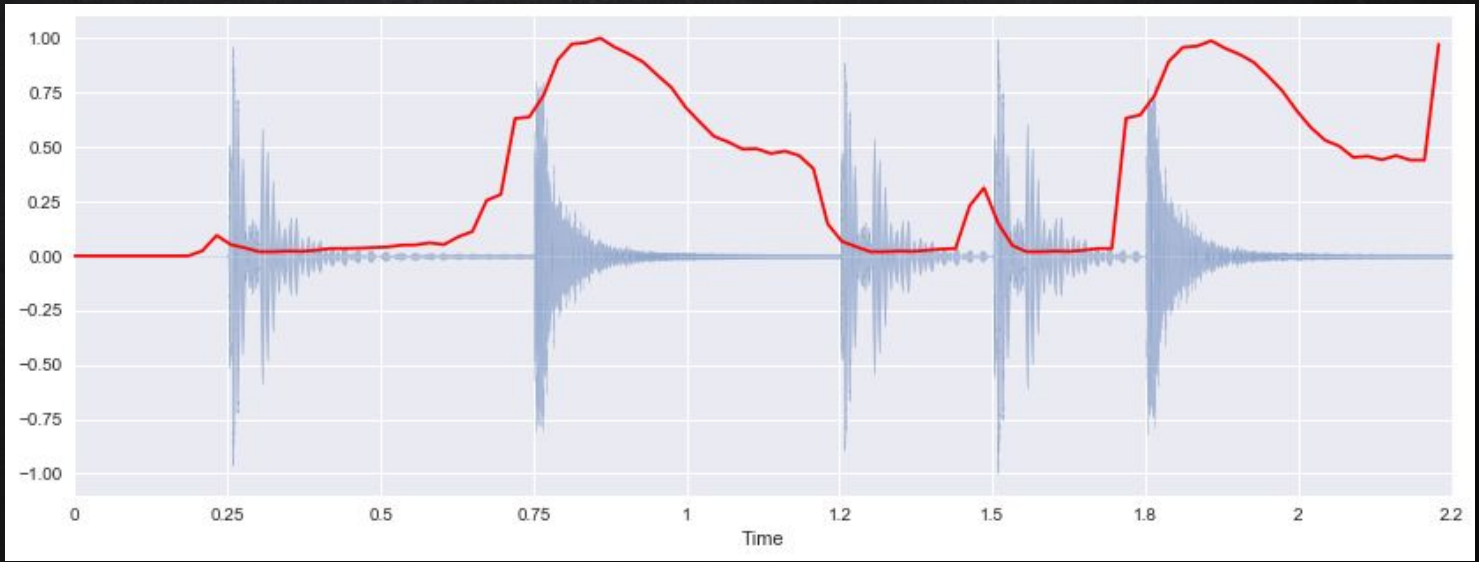
$$Valley_k = \log\{\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x'_{k,N-i+1}\} \quad (2)$$

And their difference is:

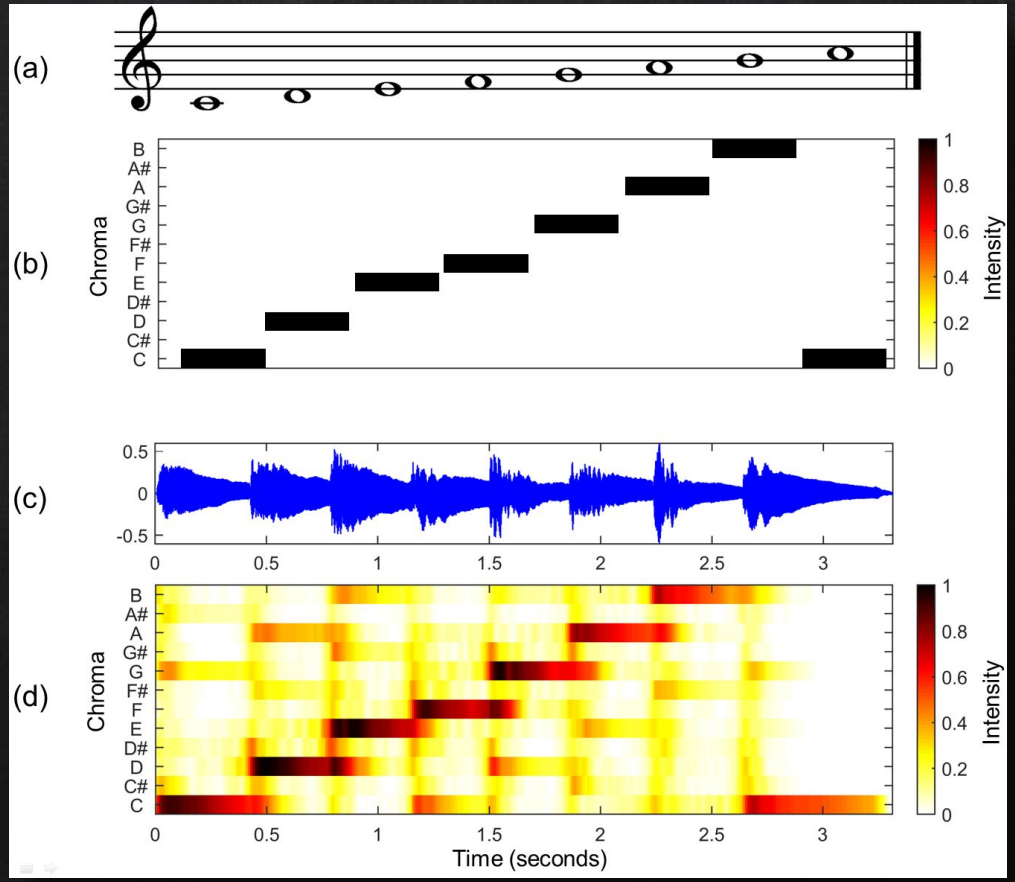
$$SC_k = Peak_k - Valley_k \quad (3)$$

where N is total number in k-th sub-band, $k \in [1,6]$.

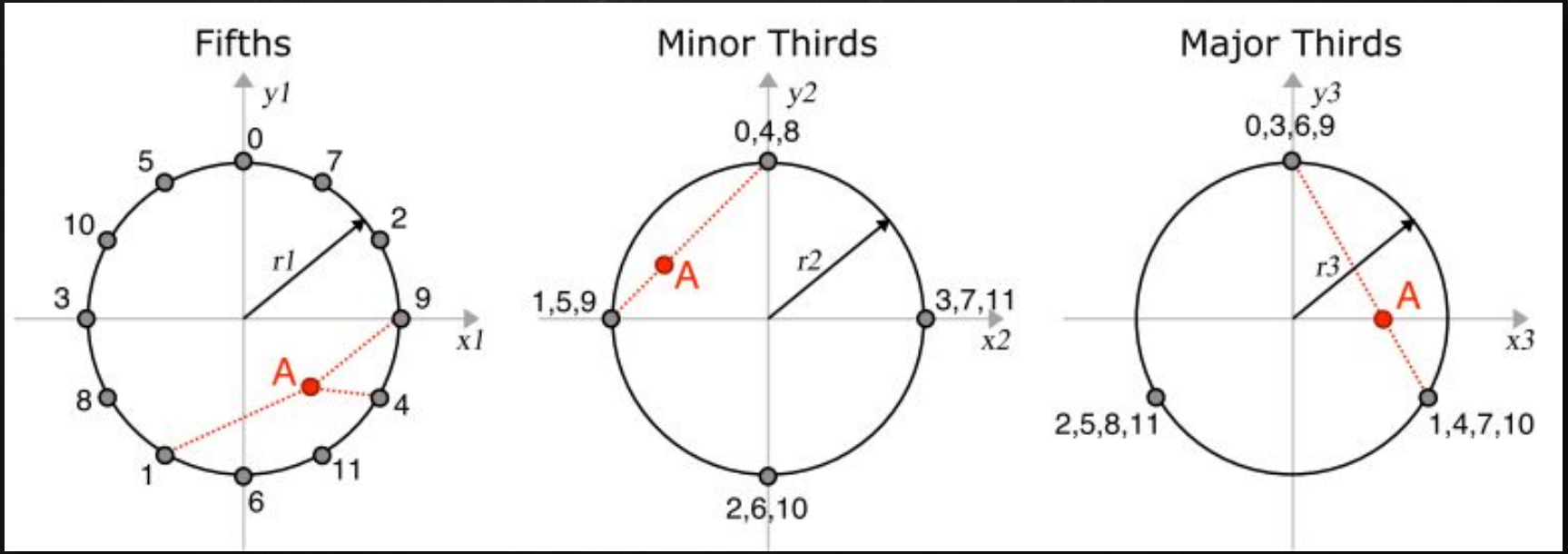
Spectral Roll-Off



Chroma Features



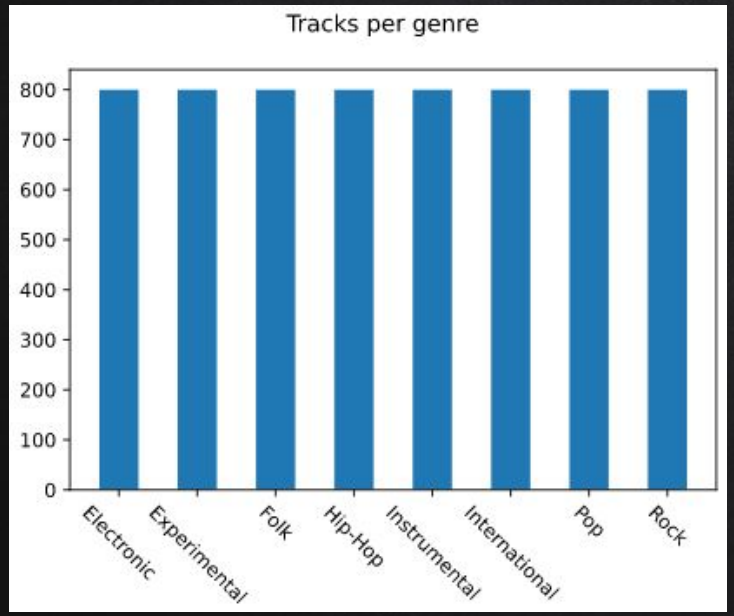
Tonal Centroid Features (Tonnetz)

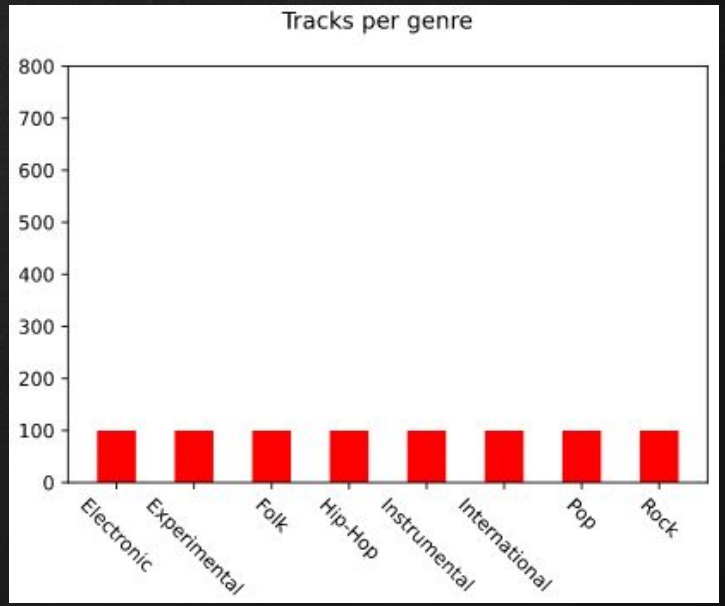


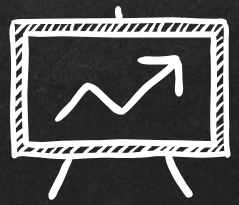
Our Method

Dataset prep

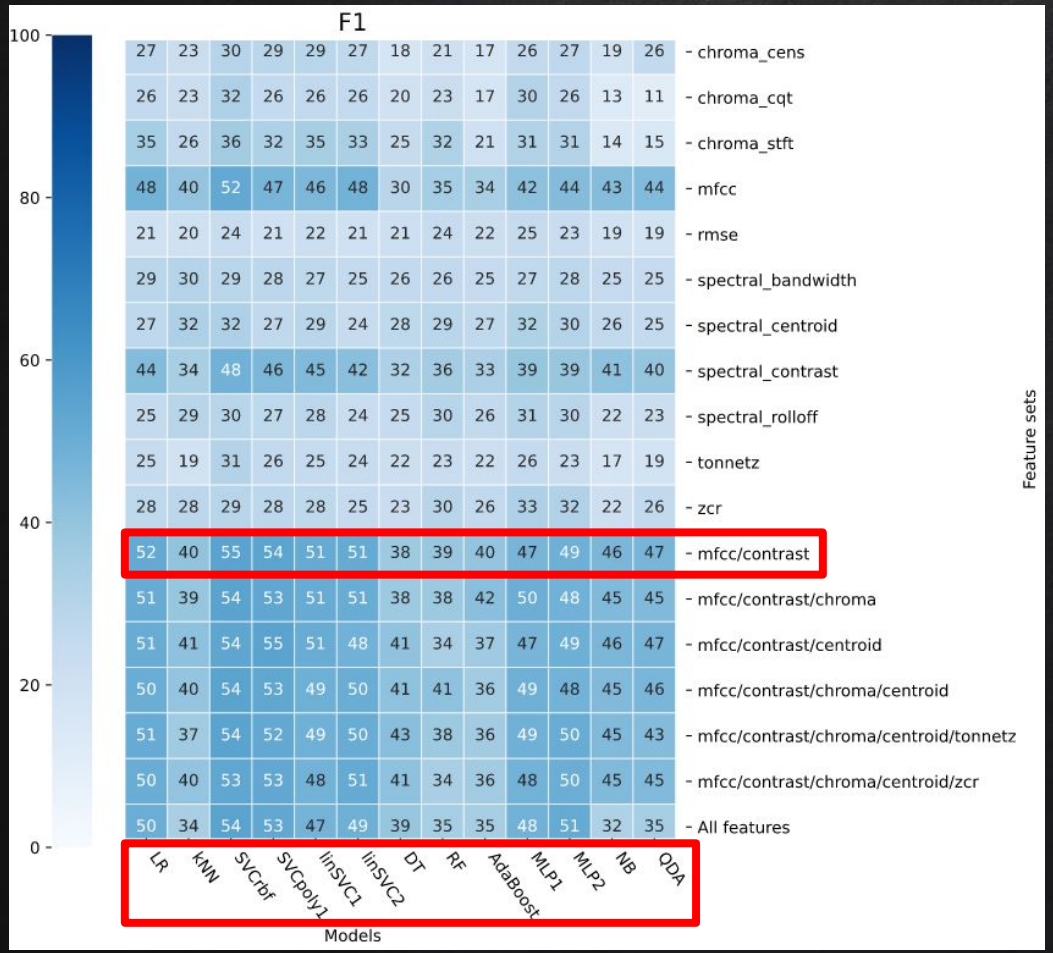
Train : Validation : Test = 8:1:1
Equally Distributed for each genre!







Feature and Model Selection



Ensemble One:

LR
linSVC1
linSVC2
polynomial SVC
RBF SVC

Ensemble Two:

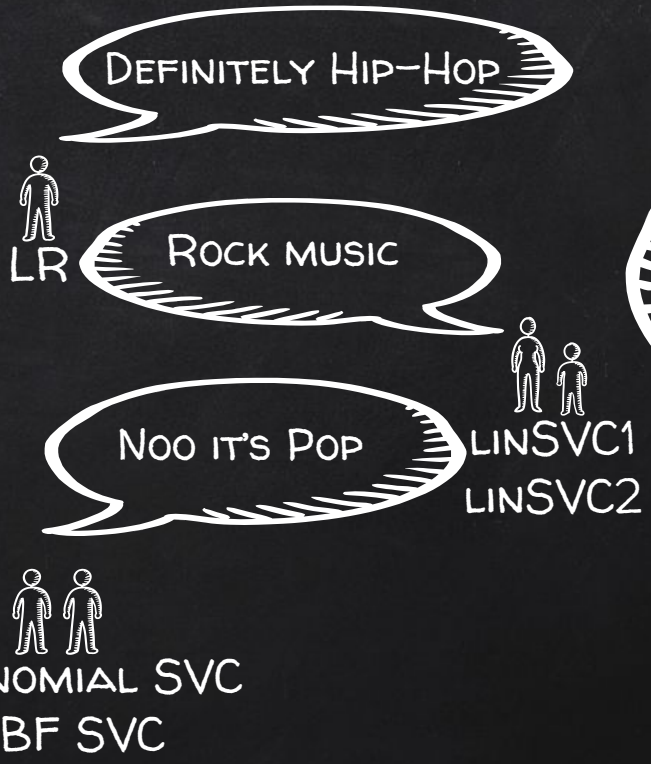
MLP2
LR
RBF SVC

Ensemble Voting Mechanism

Let's vote!

<u>Ensemble One:</u>

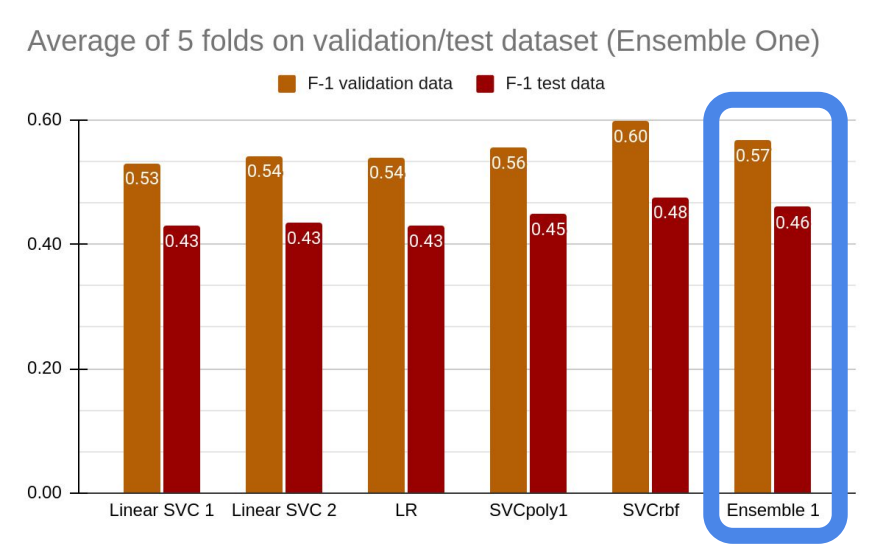
LR
linSVC1
linSVC2
polynomial SVC
RBF SVC

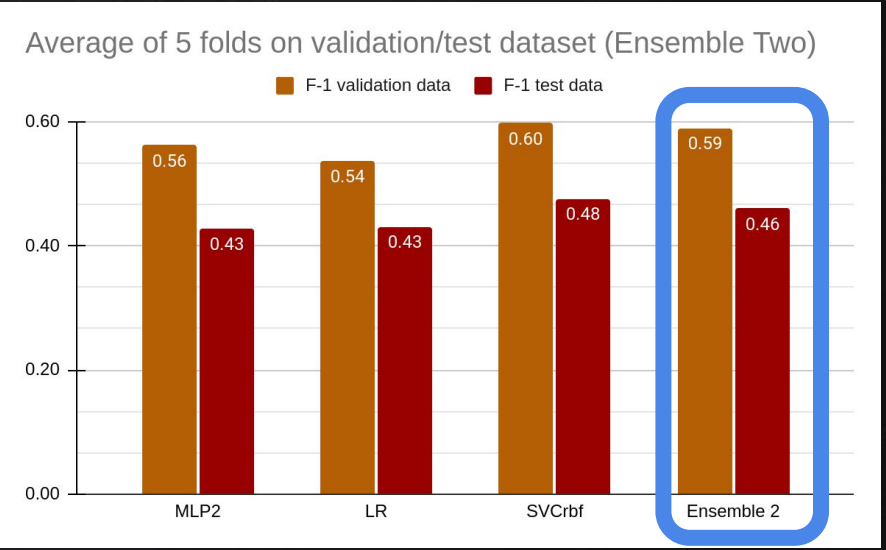


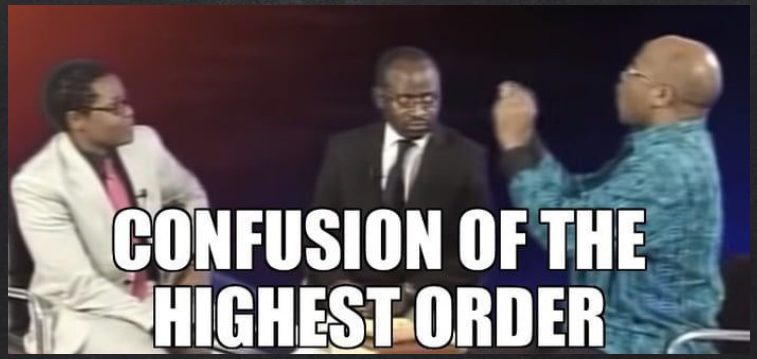
2:2:1 votes!?

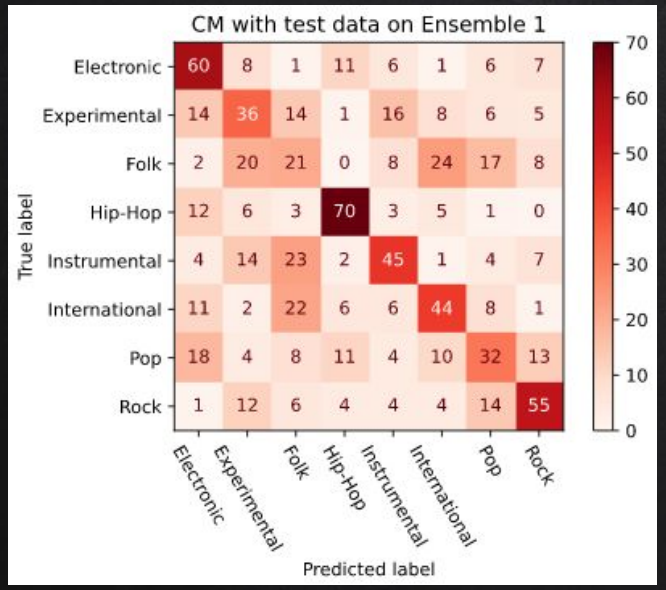
Evaluation of Ensemble One / Two

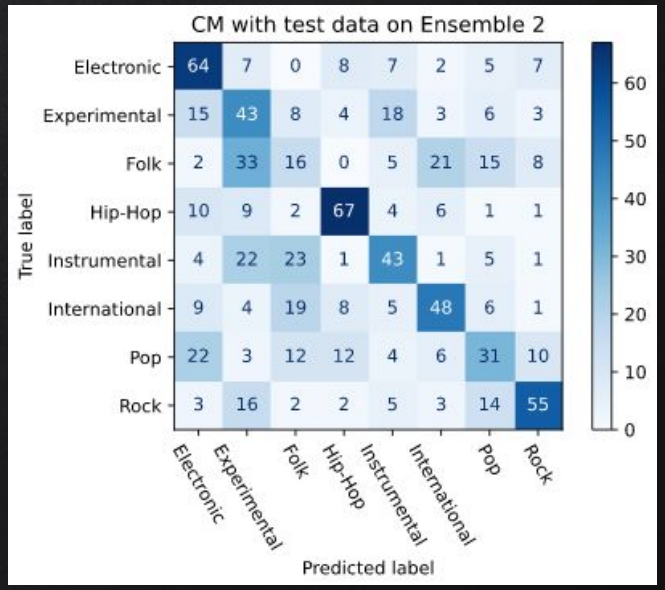
F1 Performance






CONFUSION OF THE
HIGHEST ORDER


CM with test data on Ensemble 1
True label
Predicted label
Electronic
Experimental
Folk
Hip-Hop
Instrumental
International
Pop
Rock
60 8 1 11 6 1 6 7
14 36 14 1 16 8 6 5
2 20 21 0 8 24 17 8
12 6 3 70 3 5 1 0
4 14 23 2 45 1 4 7
11 2 22 6 6 44 8 1
18 4 8 11 4 10 32 13
1 12 6 4 4 4 14 55
70
60
50
40
30
20
10
0


CM with test data on Ensemble 2
True label
Predicted label
Electronic
Experimental
Folk
Hip-Hop
Instrumental
International
Pop
Rock
64 7 0 8 7 2 5 7
15 43 8 4 18 3 6 3
2 33 16 0 5 21 15 8
10 9 2 67 4 6 1 1
4 22 23 1 43 1 5 1
9 4 19 8 5 48 6 1
22 3 12 12 4 6 31 10
3 16 2 2 5 3 14 55
60
50
40
30
20
10
0

Conclusion

We don't need an ensemble model

Conclusion

We don't need an ensemble model

Or do we?

An ensemble model might still

generalize better and outperform RBF SVC on

a larger or different dataset