

Source :

- <https://librosa.org/doc/latest/feature.html#feature>
- https://ravinkumar.com/GenAiGuidebook/audio/audio_feature_extraction.html

MFCC

Các bước tính

- Chia tín hiệu âm thanh thành các khung nhỏ (framing).
- Áp dụng cửa sổ (windowing) lên mỗi khung.
- Tính phổ biên độ (magnitude spectrum) bằng biến đổi Fourier nhanh (FFT).
- Áp dụng bộ lọc tam giác theo thang tần Mel (Mel filterbank).
 - o (thang mel là thang tần số của con người cảm nhận được)
- Lấy logarit của năng lượng ở mỗi đầu ra bộ lọc.
 - o (con người cảm nhận biên độ theo logarit)
- Áp dụng biến đổi cosin rời rạc (DCT) lên các log năng lượng đó.
 - o (nén dữ liệu và giảm tương quan)
- Lấy các hệ số đầu tiên (thường là 12–13) làm đặc trưng MFCC.

Lý do lựa chọn :

- FFT để xét theo miền tần số
- thang mel là thang tần số con người cảm nhận được
- deci Ben là thang biên độ con người cảm nhận được
- DCT chỉ chọn dữ liệu quan trọng, giảm tương quan

Link thư viện :

<https://librosa.org/doc/latest/generated/librosa.feature.mfcc.html#librosa.feature.mfcc>

Zero Crossing Rate

Định nghĩa :

- Tỷ lệ vượt qua 0

$$ZCR = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{|\text{sign}(x[i]) - \text{sign}(x[i-1])|}{2}$$

Lý do lựa chọn

- Zcr cao tương đương với
 - o thay đổi thường xuyên hơn, tần số cao hơn
- Zcr thấp tương đương với
 - o Tần số thấp hơn, có nhiều tông hơn

Tâm phổ

Định nghĩa :

$$\text{Spectral Centroid} = \frac{\sum(f \cdot m(f))}{\sum m(f)}$$

Lý do lựa chọn :

- Tâm phổ biểu thị tâm của tần số âm thanh

Spectral roll off

Định nghĩa :

- f = spectral roll off
- E = tổng năng lượng của mọi tần số
- f_0 = tần số nhỏ hơn f
- n = tỉ lệ cho trước
- f thỏa mãn :
 - o tổng năng lượng của các f_0 nhỏ hơn $n \cdot E$

Lý do lựa chọn :

- giá trị cao có nghĩa là âm thanh có tần số cao
- giá trị thấp có nghĩa là âm thanh có tần số thấp

Chroma Features (chroma_stft)

Cách tính :

- sử dụng STFT chuyển từng frame sang miền tần số
- chuyển tần số sang lớp cao độ
- tổng hợp năng lượng của các tần số cùng 1 lớp cao độ
- vẽ biểu đồ
 - o trục Ox = thời gian
 - o trục Oy = các tần số thuộc cùng 1 lớp cao độ
 - o c = màu của 1 điểm = cường độ của tần số y tại thời gian x , $0 \leq c \leq 1$

Lý do lựa chọn

- các tần số khác nhau nhưng cùng lớp cao độ thì con người vẫn cảm thấy chúng giống nhau
 - o nguồn : https://en.wikipedia.org/wiki/Chroma_feature
- phù hợp để phân tích đặc điểm âm sắc