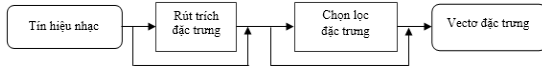


và chỉ chọn các thông tin cần thiết cho việc phân loại nhạc. Ngoài ra, việc chọn lọc đặc trưng được dùng để tạo ra một tập con đặc trưng từ dữ liệu đầu vào nhằm làm tăng hiệu quả về mặt thời gian trong việc nhận dạng vì nó là tiến trình tự động hoá được dùng để giảm số chiều dữ liệu sao cho dữ liệu đầu vào được chuyển đổi sang dạng đơn giản và nhỏ hơn trước khi đưa vào mô hình phân loại.



**Hình 4: Sơ đồ rút trích đặc trưng từ một tín hiệu nhạc**

Nhiều nghiên cứu đã đề xuất các đặc trưng của tín hiệu audio để nhận dạng, phân loại trong các hệ thống nhận dạng, phân loại khác nhau. Mỗi nghiên cứu đều đưa ra một số các đặc trưng của tín hiệu audio và phương thức sử dụng để phân loại. Các đặc trưng của tín hiệu audio thường được chia làm hai nhóm chính: các đặc trưng trong miền thời gian – tần số và các đặc trưng cảm thụ âm thanh của con người (nhịp điệu, cao độ) (Wongso and Santika, 2014). Trong bài báo này, chúng tôi xây dựng hệ thống phân loại nhạc dựa trên hai tập đặc trưng như sau:

- Các đặc trưng về âm sắc (Timbral Texture Features).
- Các đặc trưng về nhịp điệu (Rhythmic Content Features).

### 3.1 Đặc trưng về âm sắc

Tập đặc trưng về âm sắc được sử dụng để biểu diễn các đặc trưng của âm nhạc liên quan đến tiết tấu, âm sắc và nhạc cụ. Vector đặc trưng về âm sắc được sử dụng trong hệ thống phân loại của chúng tôi bao gồm 19 chiều với các đặc trưng: (Trung bình và độ lệch chuẩn của Spectral Centroid, Rolloff, Flux, ZeroCrossing, LowEnergy, và Trung bình và độ lệch chuẩn của 5 hệ số MFCC đầu tiên). Trung bình và độ lệch chuẩn của các đặc trưng này được xác định dựa trên STFT với các cửa sổ phân tích chia tín hiệu đầu vào có độ dài 1s thành các đoạn nhỏ khoảng 20ms. Sau đây là các đặc trưng được xác định trên mỗi cửa sổ phân tích:

#### a. Đặc trưng 1: Spectral Centroid

Spectral Centroid là một độ đo liên quan hình dáng của phổ tần số. Nó xác định điểm cân bằng của phổ tần số. Giá trị Centroid cao tương ứng với phổ có độ sáng chói hơn và chứa nhiều tần số cao. Spectral Centroid được xác định bởi công thức (6):

$$C_t = \frac{\sum_{n=1}^N M_t[n] * n}{\sum_{n=1}^N M_t[n]} \quad (6)$$

Trong đó:  $M_t[n]$  là biên độ của tần số thứ  $n$  trong phổ tần số tương ứng với cửa sổ  $t$ .

#### b. Đặc trưng 2: Rolloff

Rolloff cũng là một độ đo liên quan hình dáng của phổ tần số. Điểm Rolloff của phổ tần số (Rt) được định nghĩa như tần số biên mà ở đó 85% phân bố năng lượng được tập trung trong phổ là dưới điểm này. Công thức (7) xác định Rt - điểm Rolloff của phổ tần số.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^N M_t[n] \quad (7)$$

#### c. Đặc trưng 3: Flux

Flux được xem là độ biến thiên phổ, cho biết sự thay đổi về biên độ tần số của phân phối quang phổ giữa hai cửa sổ phân tích liên tiếp. Nó được xác định là bình phương hiệu giữa các biên độ chuẩn của tần số trong phổ và được xác định bởi công thức (8).

$$F_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n])^2$$

$$N_t[n] = \frac{M_t[n]}{\sqrt{\sum_{i=1}^N (M_t[i])^2}} \quad (8)$$

Với  $N_t[n]$  và  $N_{t-1}[n]$  là biên độ chuẩn của tần số thứ  $n$  trong phổ tần số ở cửa sổ  $t$  và  $t-1$  tương ứng.

#### d. Đặc trưng 4: Zero-crossings

Zero Crossings cho biết mức độ ồn (noisiness) của âm thanh trong tín hiệu. Nó xuất hiện khi các mẫu kề nhau trong tín hiệu khác dấu. Nó được xác định bởi số lần tín hiệu audio vượt qua trục zero trên một đơn vị thời gian và được tính bởi công thức:

$$Z_t = \frac{1}{2} \sum_{n=1}^N |sign(x[n]) - sign(x[n-1])| \quad \text{và} \quad (9)$$

$$sign(x[n]) = \begin{cases} 1 & x[n] > 0 \\ 0 & x[n] \leq 0 \end{cases}$$

$x[n]$  là tín hiệu trong miền thời gian đối với cửa sổ  $t$ .