Audio Features: MFCC & Chroma

MFCC

Mel-Frequency Cepstral Coefficients

Models human auditory perception

Based on: Mel-scale (perceptually motivated frequency scale)

- Speech recognition
- Speaker identification
- Audio classification

Chroma Features

Pitch Content & Harmonic Structure

Represent pitch content and harmonic structure

Focus on: 12 pitch classes (chromatic scale)

- Music analysis
- Chord recognition
- Similarity detection

Both features capture complementary aspects of audio signals

✓

Compact representations suitable for machine learning models

Visual Process Flow

MFCC Extraction Process

1 Audio Signal

Input original audio signal (time domain)

2 STFT

Frequency analysis with Short-Time Fourier Transform

3 Mel Filter Bank

Apply Mel-scale filters that reflect human auditory characteristics

4 DCT

Extract coefficients with Discrete Cosine Transform

MFCC Coefficient Example (13 coefficients)

Chroma Extraction Process

1 Audio Signal

Input original music signal

2 STFT

Analyze frequency spectrum

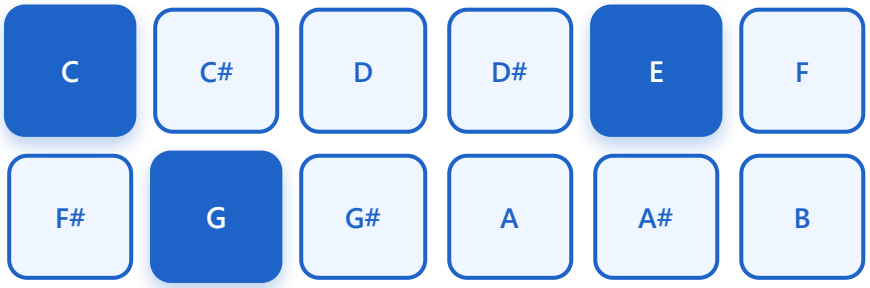
3 Pitch Mapping

Map frequencies to 12 pitch classes (C, C#, D, ...)

4 Octave Folding

Fold all octaves into one (octave invariance)

Chroma Vector Example (12 pitch classes)



Each coefficient represents different frequency characteristics of the audio

Highlighted notes = C Major chord (C, E, G)

The intensity of each pitch class is represented as a vector value