

# Audio Clustering Documentation

## Objective

To perform unsupervised clustering of audio files using MFCCs as audio embeddings.

## Process Overview

Feature Extraction:

- Audio features were extracted using the librosa library.
- MFCCs were computed for each audio file, providing a representation of the audio's spectral properties.

Embeddings Conversion:

- MFCCs were used directly as embeddings for clustering.

Clustering Algorithm:

- K-Means clustering was applied to the embeddings to partition the audio data into clusters.

Cluster Analysis:

- Clusters were analyzed by reviewing the count and sample files from each cluster to understand the similarities within and diversities across different clusters.

## Implementation Details

- Librosa: Used for audio feature extraction, specifically for computing MFCCs.
- K-Means: Chosen for its simplicity and effectiveness, implemented via `sklearn.cluster.KMeans`.
- Cluster Evaluation: The Elbow Method was used to determine the optimal number of clusters before running K-Means.

## Results and Discussion

- The resulting clusters showed grouping based on the spectral features of the audio files.
- Sample files from each cluster were reviewed to qualitatively assess the clustering results.
- The clusters can help in tasks such as speaker diarization, genre classification, or emotion detection in audio samples.

## Future Work

Further improvements can be explored by:

- Incorporating more diverse feature sets such as Chroma features and Spectral Contrast.
- Experimenting with different clustering algorithms like Hierarchical Clustering or DBSCAN.
- Validating clusters with labeled data, if available, to quantify the performance of the clustering.