TAI Project 3 - Music Identification System

Using Normalized Compression Distance

Alexandre Ribeiro 108122, Maria Sardinha 108756, Miguel Pinto 107449

Introduction

This project implements a Music Identification System that uses Normalized Compression Distance (NCD) combined with audio feature extraction to identify songs from short audio samples. The system is capable of identifying some music tracks even when they are corrupted with various types of noise (brown, pink, and white) and can work with different audio feature extraction methods.

Goals & Methods

- **Audio Feature Extraction**: Extract frequency-domain features from WAV files using two different methods:
  - **Spectral Method**: Binned frequency spectrum using FFT (Fast Fourier Transform).
  - **Maximum Frequency Method**: Top dominant frequencies per frame.
- **Music Identification**: Identify songs from short samples (10/20 seconds) by comparing extracted features using NCD to obtain the best match.
- **Noise Robustness**: Handle audio samples corrupted with different types of noise (brown, pink, white).
- **Mulltiple Compressors**: Support for various compression algorithms (gzip, bzip2, lzma, zstd) for NCD calculation.
- **Batch Processing**: Automated testing and analysis pipelines for large datasets.
- **Performance Analysis**: Accuracy analysis with visualization tools.

The system was tested in **two datasets**:

- A **smaller** one (**36 songs**): With maximum accuracy 26.5%.
- A **bigger** one (**100 songs**): With maximum accuracy 23%, depending on the music genre (with Hip-Hop/Rap showing the best performance and Classical/Orchestral being the most challenging).

Project Structure

System Architecture:

- **Core Applications**: Feature extraction and music identification executables.
- **Core Library**: Audio processing, feature extraction, and NCD computation modules.
- **Utilities**: Compressor wrapper and JSON configuration support.

Feature Extraction Methods:

- **Spectral Method**:
  - FFT-based frequency analysis.
  - Logarithmically scaled frequency bins.
  - 1024-sample frames with 512-sample hop.
- **Maximum Frequency Method**:
  - Peak detection in frequency domain.
  - Top N dominant frequencies per frame.
  - Compact representation.

NCD Implementation:

- Formula: $\mathrm{NCD}(x,y) = \dfrac{C(x,y) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}$
- Where:
  - *C(x)* = compressed size of file *x*.
  - *C(xy)* = compressed size of concatenated files.
  - Supports: gzip, bzip2, lzma, zstd compressors.

Processing Pipeline:

1. WAV file reading (8/16/24/32-bit PCM).
2. Frame segmentation with overlapping windows.
3. Feature extraction (spectral or maxfreq).
4. Binary/text serialization.
5. NCD-based similarity computation.
6. Ranking and identification.
7. Metrics and visualizations generation.

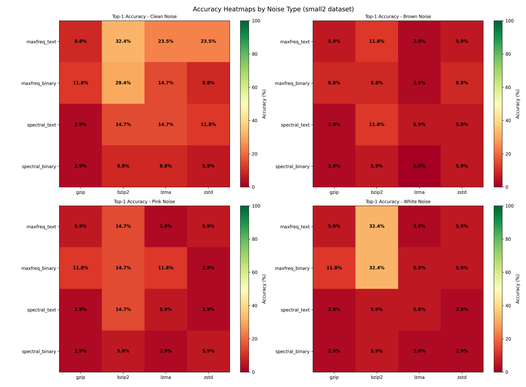
Other characteristics:

- Error handling.
- Temporary file management.
- Ensures valid NCD values.
- Multi-threading support:
  - Parallel feature extraction.
  - Parallel compressor evaluation.
  - Thread-safe I/O.

Results

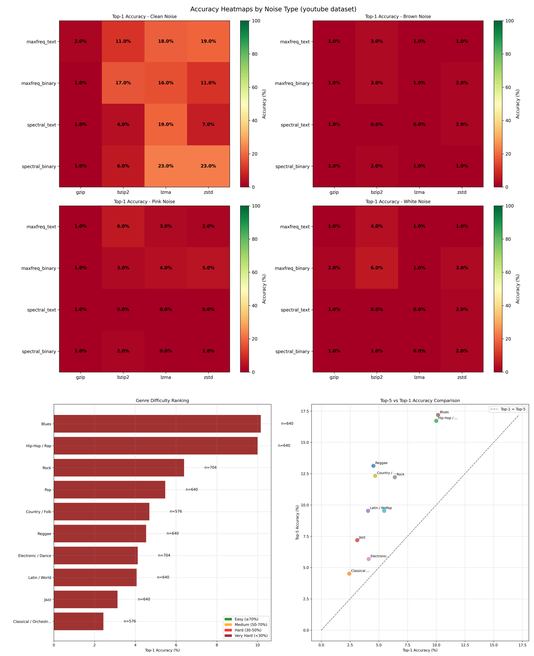
Performance Overview - **36 Songs** Dataset (20s samples):

- Best accuracy with bzip2 & MaxFreq (text).



Performance Overview - **100 Songs** Dataset (10s samples):

- Best accuracy with lzma/zstd & Spectral (binary).
- Accuracy drops with noise and for some genres.



Conclusion

We successfully managed to create a trainning-free music identification system based on Normalized Compression Distance (NCD) and frequency-domain features.

Main Findings

- Up to 26.5% accuracy (36 songs) and 23% (100 songs), with **best** results for **Hip-Hop/Rap** and **worst** for **Classical/Orchestral.**
- **Bzip2** emerges as the **best-performing** compressor, balancing accuracy and speed.
- The system shows **robustness to noise**, though performance **degrades most with white noise**.
- **Binary** features provide **similar accuracy** to text formats, with **improved storage efficiency**.

Limitations

- **Accuracy**: Average accuracy was 6.3% across the full dataset
- **Computational Complexity**: NCD comparison scales quadratically with database size
- **Feature Engineering**: Current methods could be extended with richer features such as MFCCs and chroma

---

Future Work

- **Essemble methods**: Multiple segments per song with NCD averaging.
- **Dataset expansion**: More diverse and curated song collections.
- **Sample length analysis**: Optimal duration for identification.
- **Advanced feature analysis:** Integration of other methods such as chromagrams and tempograms.
- **Hybrid approaches**: Combining NCD with machine learning.

References

Ming Li, Xin Chen, Xin Li, Bin Ma and P. M. B. Vitanyi, "The similarity metric," in IEEE Transactions on Information Theory, vol. 50, no. 12, pp. 3250-3264, Dec. 2004, doi: 10.1109/TIT.2004.838101.

Song websites: https://500audio.com/ & https://youtube.com

universidade de aveiro
theoria poiesis praxis

Departamento de Electrónica, Telecomunicações e Informática
Universidade de Aveiro

ua.pt | ua.pt/en/deti/department