MUSICAL GENRE CLASSIFICATION

Combining Features Extracted From Audio, Symbolic, and Cultural Sources

Burcak Agridag
burcak.agridag@rwth-aachen.de

June 29, 2016

RWTH AACHEN UNIVERSITY

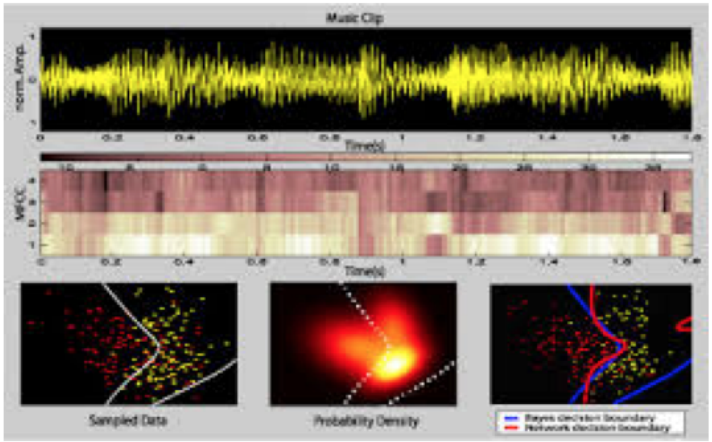
Contents

Introduction

Musical Genres

- A musical genre is characterized by the common characteristics shared by its members as a result of a complex interaction between the public, marketing, historical, and cultural factors.
- These characteristics typically are related to the instrumentation, rhythmic structure, and harmonic content of the music.

Genre Classification



Different Data Sources

Source Types

- Audio (e.g. MP3)
- Symbolic (e.g. MIDI)
- Cultural (e.g. web data, user tags, surveys)
- 7 possible combinations

Why Those Sources

- Audio
  - Audio processing is the fundamental way in which the music is consumed.
    - Spectral, time domain information extracted directly from audio(not intuitively musical).
- Symbolic
  - High-level musical features can be extracted easily compared to mere audio data.
  - Polyphonic audio to symbolic transcription algorithms continue to improve.
    - Instruments, Chord Progressions, Pitches, Rythmic Patterns, etc..
- Cultural
  - Labeling musical pieces based on past experience and knowledge primarily shaped by the culture.
    - Composer, Style, Genre(!), etc.

Accuracy Improvement

Classification Accuracy rate(%)	3-Resources	1-Resource
5 genre taxonomies	96.8%	85.5%
10 genre taxonomies	78.8%	65.1%

History

Previous Studies

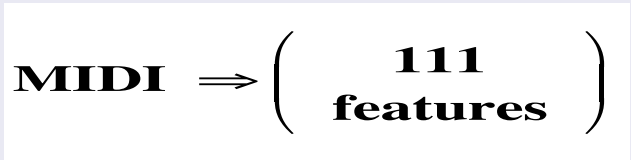
- Focus on either audio, symbolic, or cultural sources on information.
- Significant research on audio-cultural data, Whitman and Smaragdis [1].
- Less work on audio-symbolic data, Lidy et al. [2].
- Significantly increased performance when cultural features are used [3].

Methodology

Methodology Overview

- jAudio: Audio feature extractor.

- jSymbolic: Symbolic feature extractor.



- jWebMiner: Cultural feature extractor.

Audio Features

- 26 core features (e.g. Zero Crossing Rate, MFCCs, Spectral Flux, etc.)
- 5 meta-features to combine with core features (Mean, Derivative of the mean, Standard Deviation, etc.)
- 3 aggregators: Mean, Standard Deviation, and MFCCs (output on per-song basis instead of per window basis).

Audio Features

- e.g. Spectral Flux, spectral correlation btw. adjacent windows as an indication of the degree of change of the spectrum:

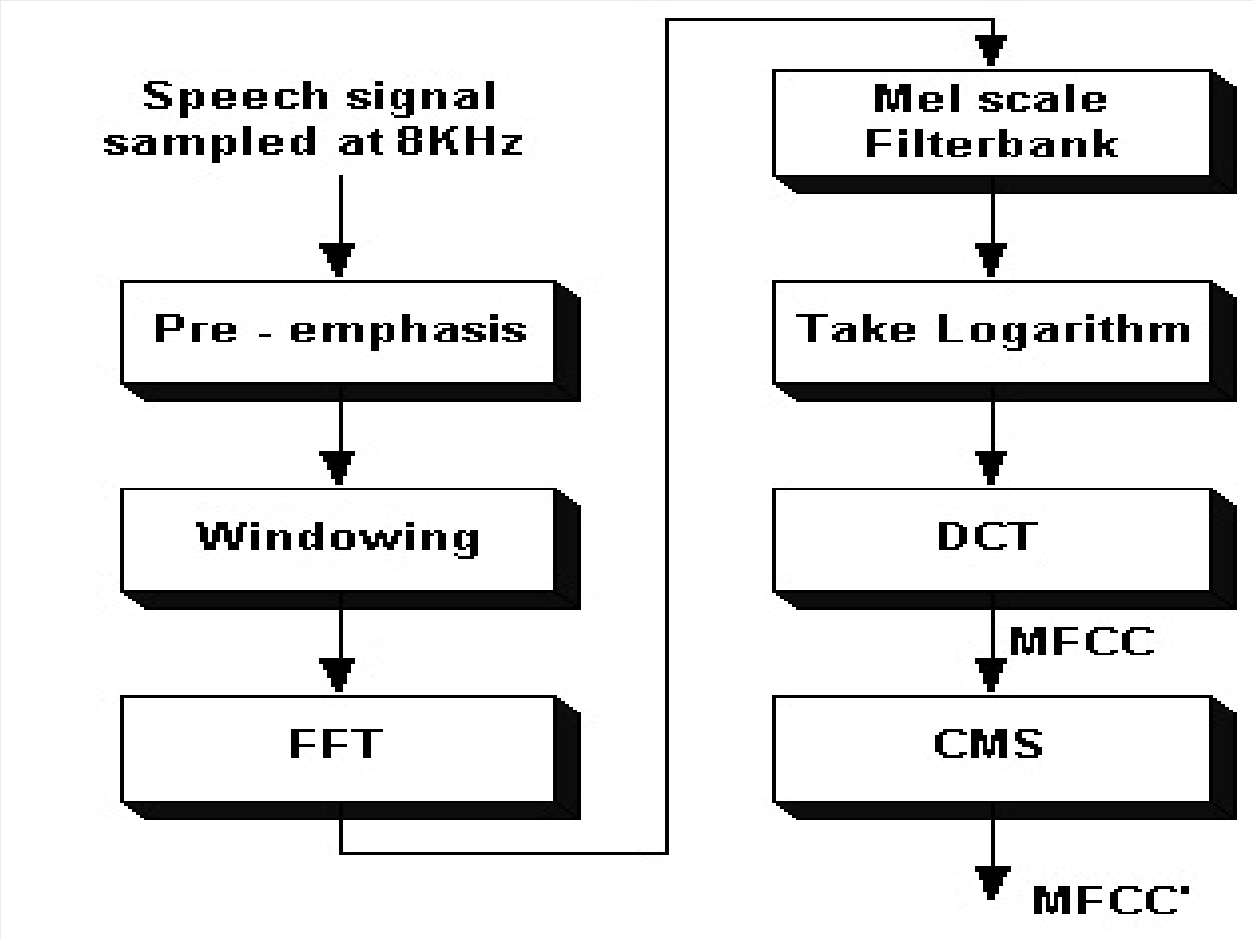
$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2$$

- e.g. RMS, calculate the amplitude of a window:

$$RMS = \sqrt{\frac{\sum_n^N x_n^2}{N}}$$

Audio Features

- e.g. MFCCs.



Symbolic Features

- Low-level features, spectral or time-domain information:
  - Spectral flux, RMS, zero-crossing rate, ...
- High-level features, musical abstractions:
  - Instruments, Melodic Contour, ...
- Cultural Features, sociocultural information:
  - Playlist co-occurrence, purchase correlations, ...

Symbolic Features

- Instrumentation:
  - What types of instruments are present?
  - Which are given particular importance?, ..
- Rhythm:
  - Time intervals between the attacks of different notes
  - Durations of each note,..
- Pitch statistics:
  - Occurrence rates of different notes, ..

Cultural Features

- Given Primary, Secondary search strings, and required filter words search the Web for all possible combinations:
  - e.g. "Beethoven" + "Classical" + "musician"
  - e.g. "Beethoven" + "Metal" + "musician"
  - Count the # of hits for each occurrence of those 3 together and hold statistics.

Experiments

Experiment Setup- SAC Dataset

- The SAC (Symbolic, Audio and Cultural) dataset consists of 250 MIDI files and 250 matching MP3s, accompanying metadata (e.g., title, artist, etc.) in iTunes XML file format.

Blues: Modern Blues *and* Traditional Blues
Classical: Baroque *and* Romantic
Jazz: Bop *and* Swing
Rap: Hardcore Rap *and* Pop Rap
Rock: Alternative Rock *and* Metal

Figure: 10 genres in SAC dataset.

Experiment Setup

Feature Type	5-Genre Code	10-Genre Code
Symbolic	S-5	S-10
Audio	A-5	A-10
Cultural	C-5	C-10
Symbolic + Audio	SA-5	SA-10
Audio + Cultural	AC-5	AC-10
Symbolic + Cultural	SC-5	SC-10
Symbolic + Audio + Cultural	SAC-5	SAC-10

Experiment Setup

- For each of the 10-genre experiments, a normalized weighted classification accuracy rate was calculate to provide insight on error types:
  - Misclassification within a genre pair (e.g., Alternative Rock instead of Metal):

    0.5 x ERROR

  - Misclassification outside a genre pair (e.g., Swing instead of Metal):

    1.5 x ERROR

Experiment Results

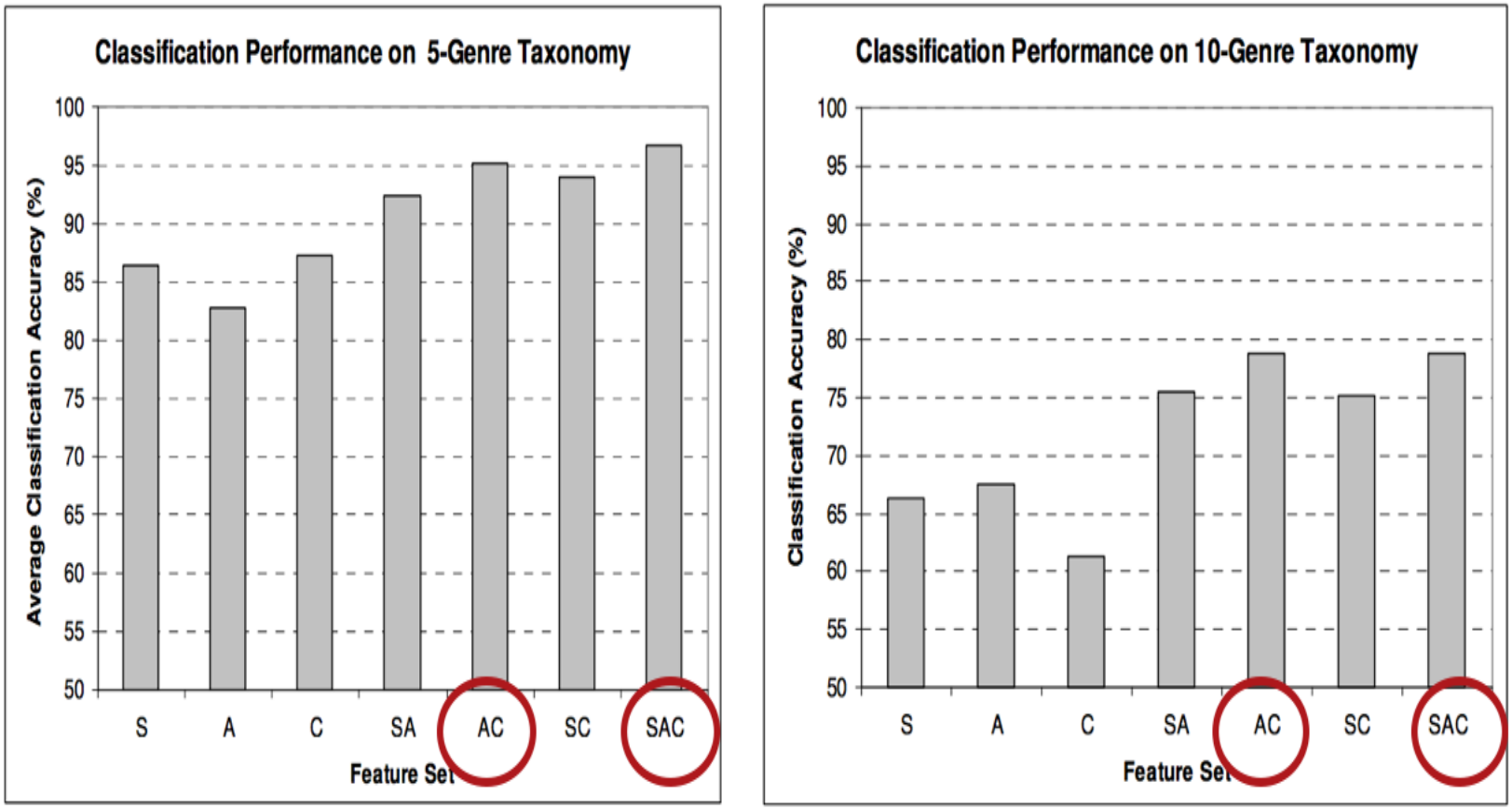
Figure: Classification accuracy (%)

	S	A	C	SA	AC	SC	SAC
5-UW	86.4	82.8	87.2	92.4	95.2	94	96.8
10-UW	66.4	67.6	61.2	75.6	78.8	75.2	78.8
10-W	66.4	67.4	66.6	78.6	84.6	81.2	84.2

	1 Feature Type	2 Feature Types	3 Feature Types
5-UW	85.5	93.9	96.8
10-UW	65.1	76.5	78.8
10-W	66.8	81.5	84.2

Experiment Results

Figure: Classification accuracy (%)



Conclusion

Conclusion

- For automatic genre classification task, it is beneficial to combine features extracted from audio,symbolic and cultural data sources.
- Combining feature types decreased the seriousness of those mis-classifications, particularly when cultural features are included.

Demo & Questions

Demo & Questions

?

References

[1] McEnnis, D., C. McKay, and I. Fujinaga. 2006. jAudio: Additions and improvements. Proceedings of the International Conference on Music Information Retrieval. 385–6.
[2] McKay, C., and I. Fujinaga. 2006. jSymbolic: A feature extractor for MIDI files. Proceedings of the International Computer Music Conference. 302–5.
[3] McKay, C., and I. Fujinaga. 2007. jWebMiner: A web-based feature extractor. Proceedings of the International Conference on Music Information Retrieval. 113–4.

References

[1] http://music.ece.drexel.edu/files/genreJawn.png.
[2]https://www.informatik.uni-
augsburg.de/de/lehrstuehle/hcm/projects/internal/mmgc/img/mmgc_logo.j
[3] http://albumware.ru.
[4] ccrma.stanford.edu.

Appendix

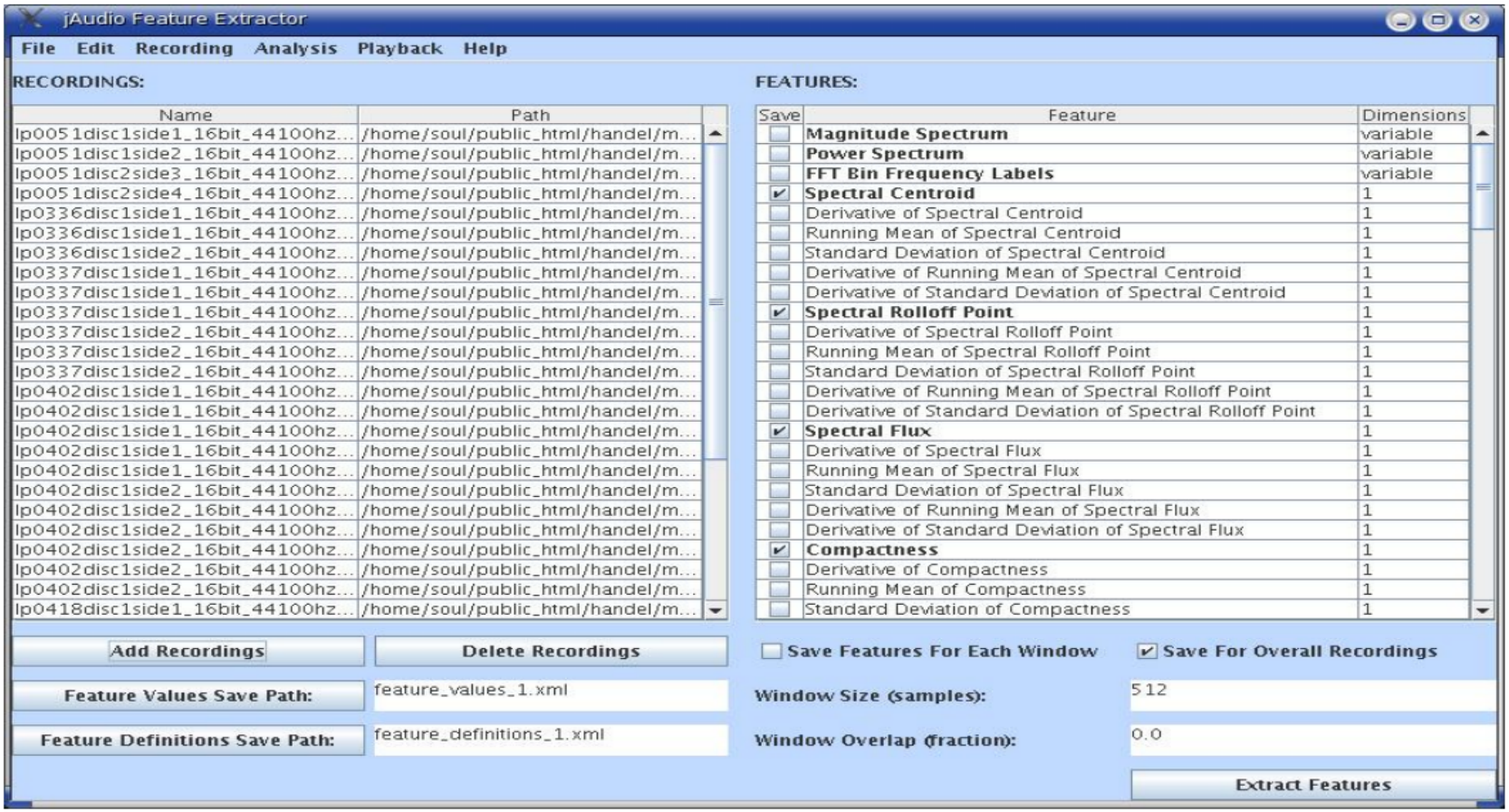


Figure: jAudio GUI

Appendix

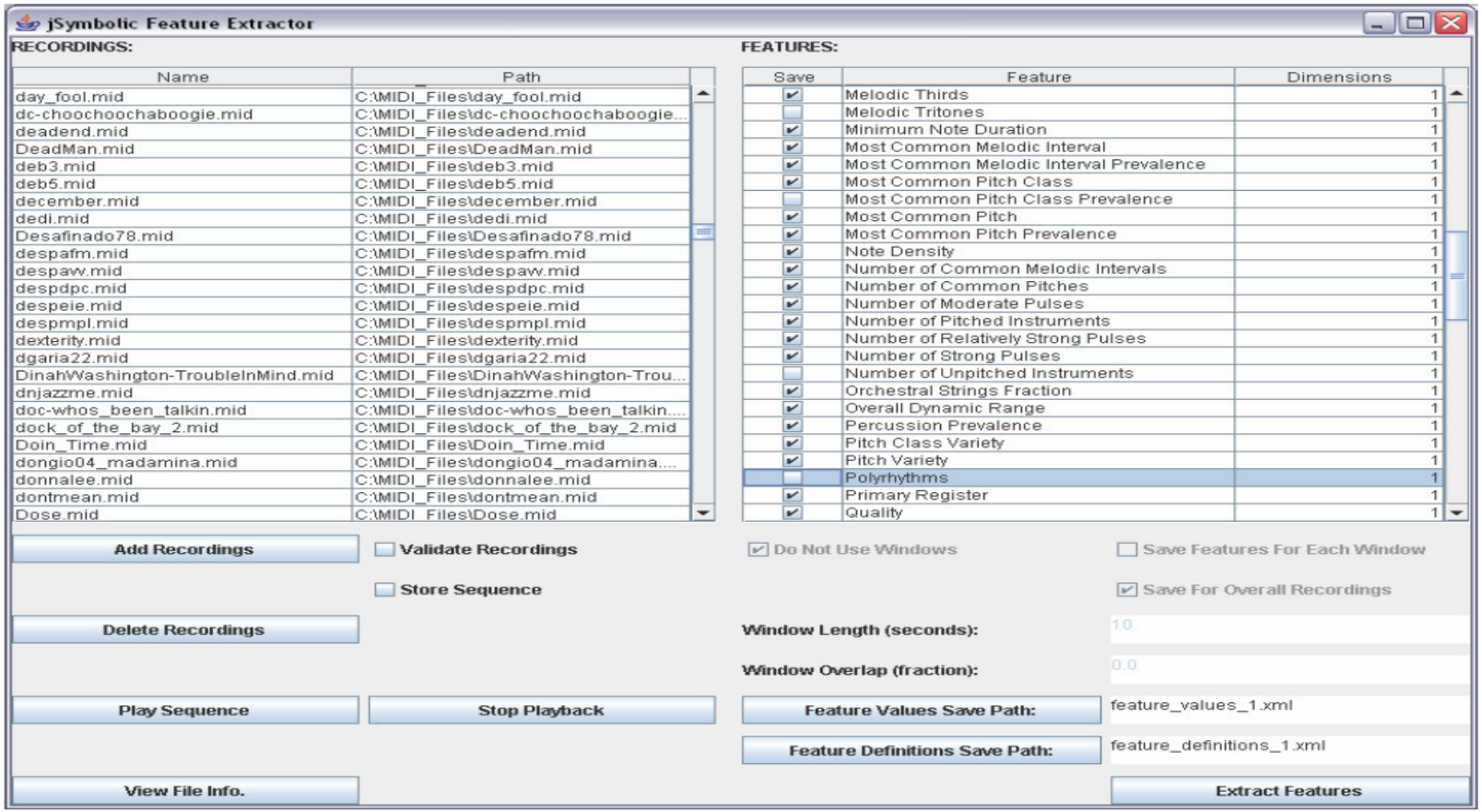


Figure: jSymbolic GUI

Appendix

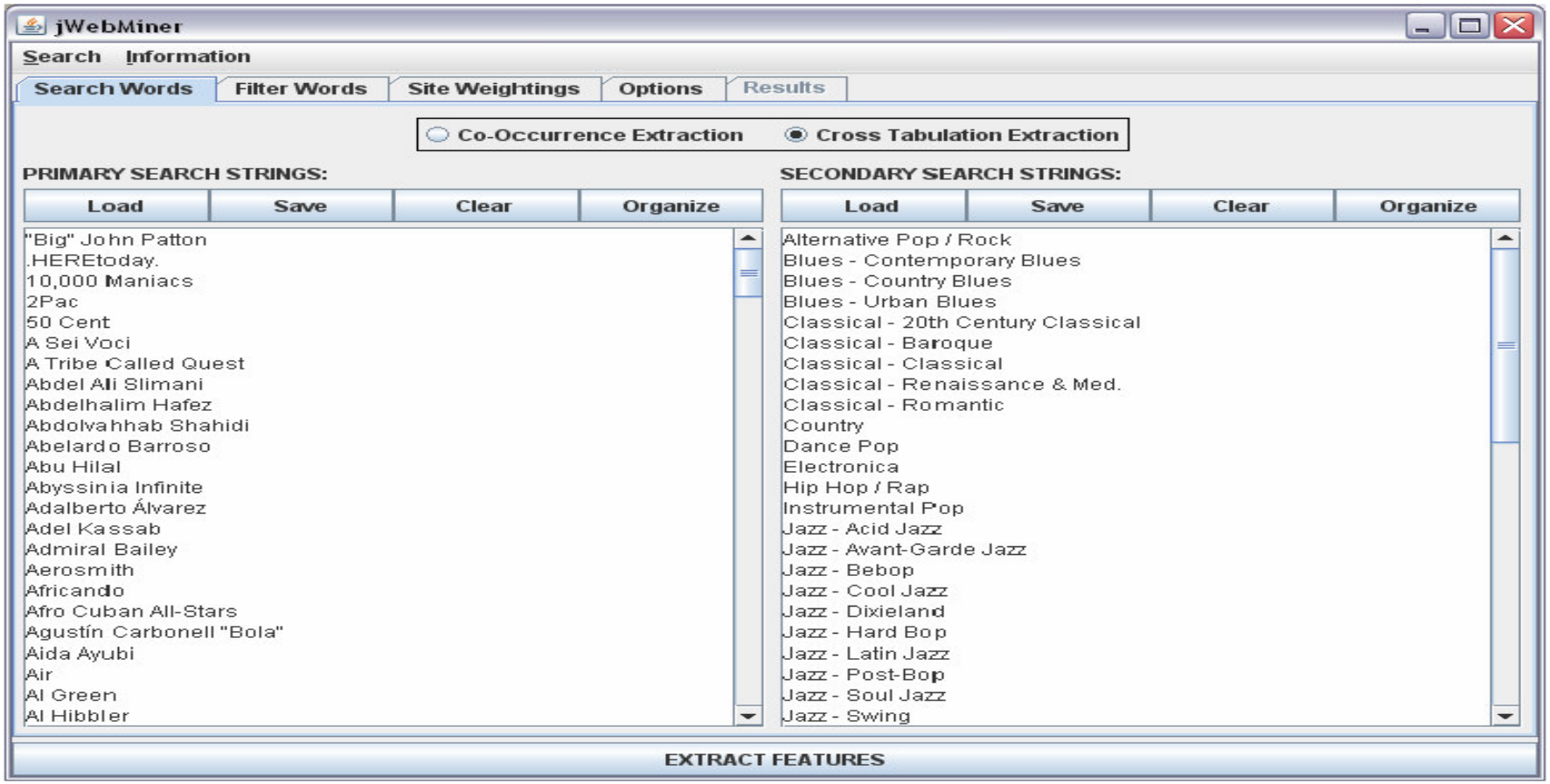


Figure: jWebMiner GUI

Appendix

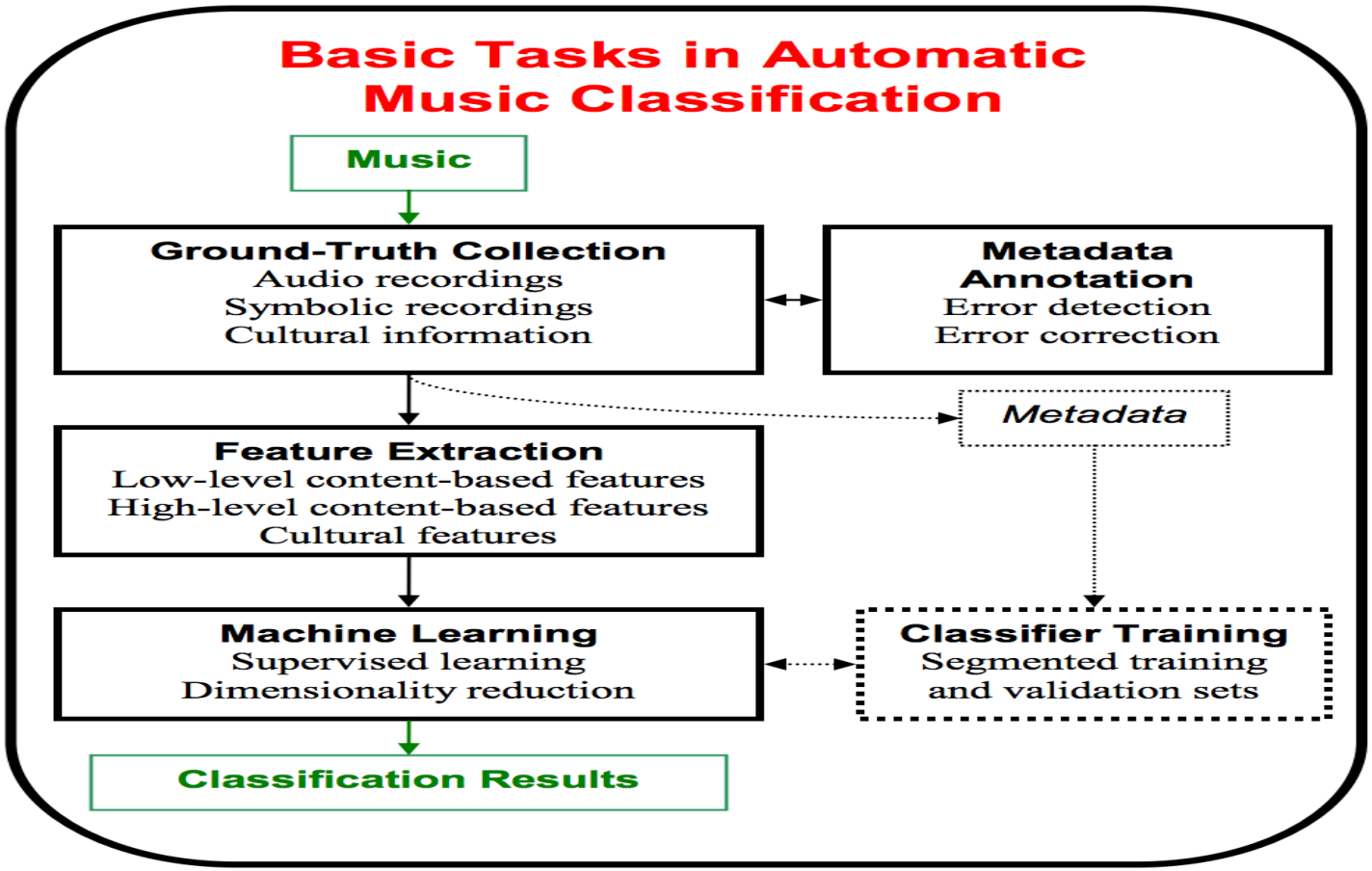


Figure: Classification

Appendix

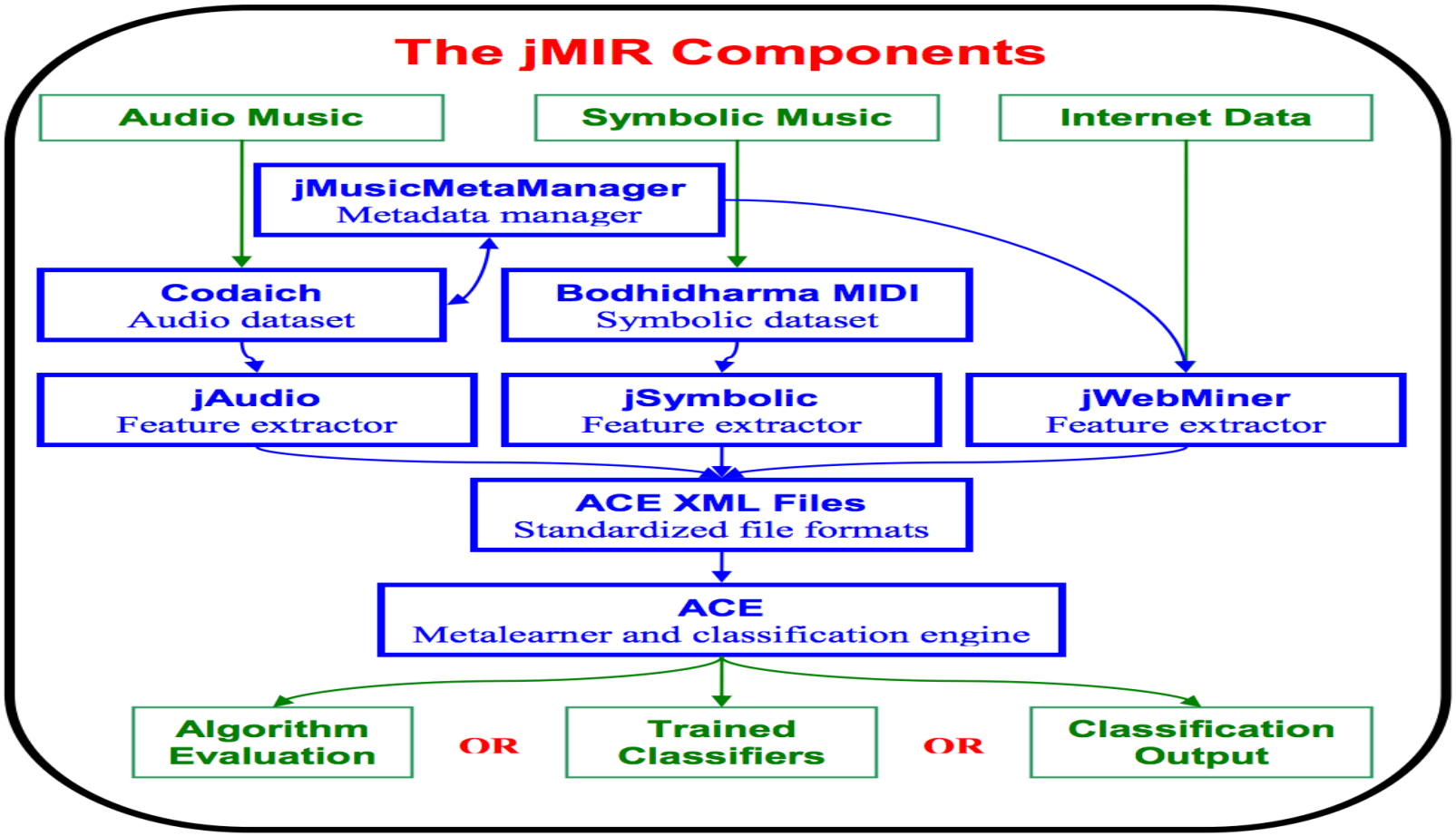


Figure: jMIR