Combining Features Extracted From Audio, Symbolic and Cultural Sources

Cory McKay
McGill University
Montréal, Canada

Presentation overview

- Research context
  - Introduction to Music Information Retrieval and Automatic Music Classification
  - Experimental goals
- Experimental methodology
- Software tools used
  - jMIR
- Results and conclusions

Goals of MIR

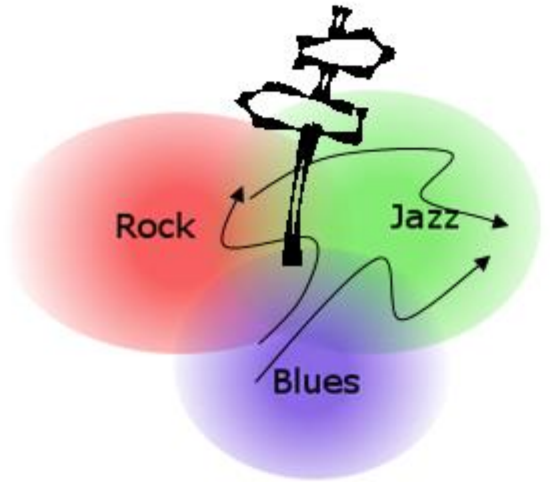
- Extract meaningful information from or about music
- Facilitate music analysis, organization, storage and access

Automatic music classification

- Automatic music classification is the particular focus of my research
- Machine learning and pattern recognition algorithms learn to classify music in various ways based on extracted features
  - Features are various kinds of information distilled from music or from sources of information on music
- Automatic music classification can involve classifying music in almost any kind of way
  - Similar techniques are commonly used regardless of the classification domain

Types of music classification

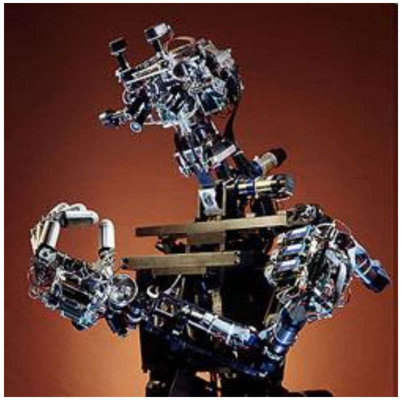
- Examples:
  - Genre or style classification
  - Mood classification
  - Performer or composer identification
  - Music recommendation
  - Playlist generation
    - e.g. iTunes Genius, Last.FM, etc.
  - Hit prediction
  - etc.



- Automatic music classification sub-systems can play an important part in many other MIR research areas
  - Automatic transcription, optical music recognition, etc.

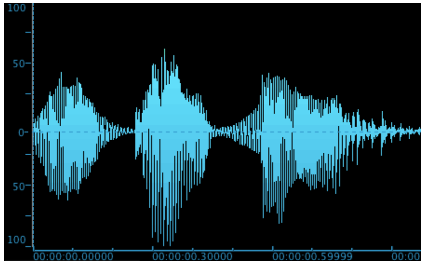
Benefits of automatic classification

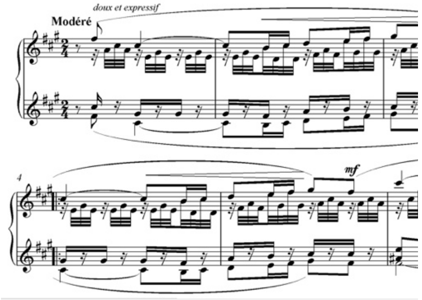
- Computers can perform classifications faster and more consistently than humans
- Computers can analyze music in novel and non-intuitive ways that might not occur to humans
- Computers can avoid human theoretical preconceptions that might contaminate experimental results



Main sources of features

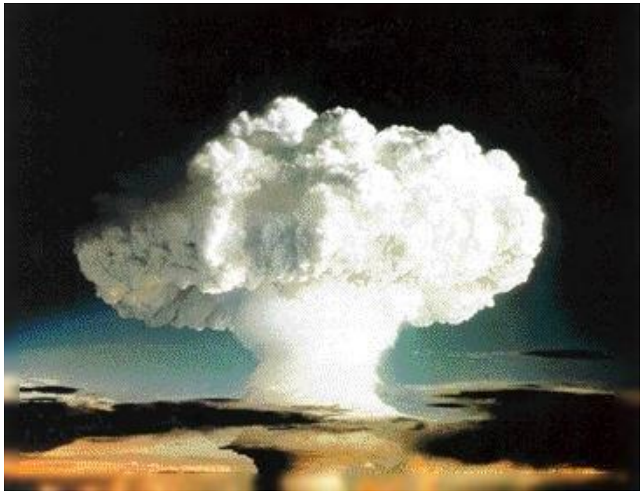
- Audio recordings
  - e.g. MP3 or .wav files
- Symbolic recordings
  - e.g. MIDI or Humdrum files
- Cultural data
  - e.g. text from the web or from metadata tags





End of the MIR Wild West

- Diminishing returns
  - Performance gains in most areas of MIR have been behaving asymptotically in recent years
- Research is increasingly focusing on fine-tuning specialized mini-tasks
  - e.g. differentiating between oboes and bassoons rather than general instrument identification
  - Has already happened in speech recognition
- Unless someone has an unforeseen breakthrough?
  - MIR cold fusion?
  - Perhaps combining feature types?



Research questions addressed

- Can combining features extracted from audio, symbolic and/or cultural sources significantly improve automatic music classification performance?
  - Classification accuracy rates
  - Severity of misclassifications that do occur
    - e.g., John Lennon → Beatles vs. John Lennon → Rihanna
- Can such an approach allow us to break past the seeming performance ceiling recently encountered in tasks like genre classification?

Previous research

- Combining audio and cultural sources (sampling)
  - Whitman and Smaragdis (2002)
  - Baumann, Klüter and Norlien (2002)
  - Dhanaraj and Logan (2005)
  - Aucouturier and Pachet (2007)
  - Eck, Bertin-Mahieux and Lamere (2007)
  - Pampalk and Goto (2007)
  - Reed and Lee (2007)
  - Dopler, Schedl, Pohle and Knees (2008)
- Combining audio and symbolic sources
  - Lidy, Rauber, Pertusa and Iñesta (2007)
    - Found that combining the two sources improved results
- Combining cultural and symbolic sources or all three
  - None?

Our experimental methodology

- Extract features from separate audio, symbolic and cultural sources of data
  - Corresponding to the same musical pieces
- Compare genre classification performance of each of the 7 possible subsets of these 3 feature groups
  - Audio, Symbolic + Audio, Cultural, Symbolic + Cultural + etc.
  - Using 10-fold cross-validation and multiple machine learning algorithms

Musical dataset used: SAC

- The SAC Dataset was assembled for this experiment
  - Symbolic Audio Cultural
  - 250 recordings belonging to 10 genres
  - Audio and MIDI versions of each recording
    - Acquired separately
  - Accompanying metadata could be used to extract cultural features from the web

Genres in SAC

- SAC's 10 genres can be collapsed into 5 genres in order to separately evaluate performance on both moderate and small genre taxonomies
  - Designed to facilitate evaluation of misclassification severity

- **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

Software tools used: jMIR

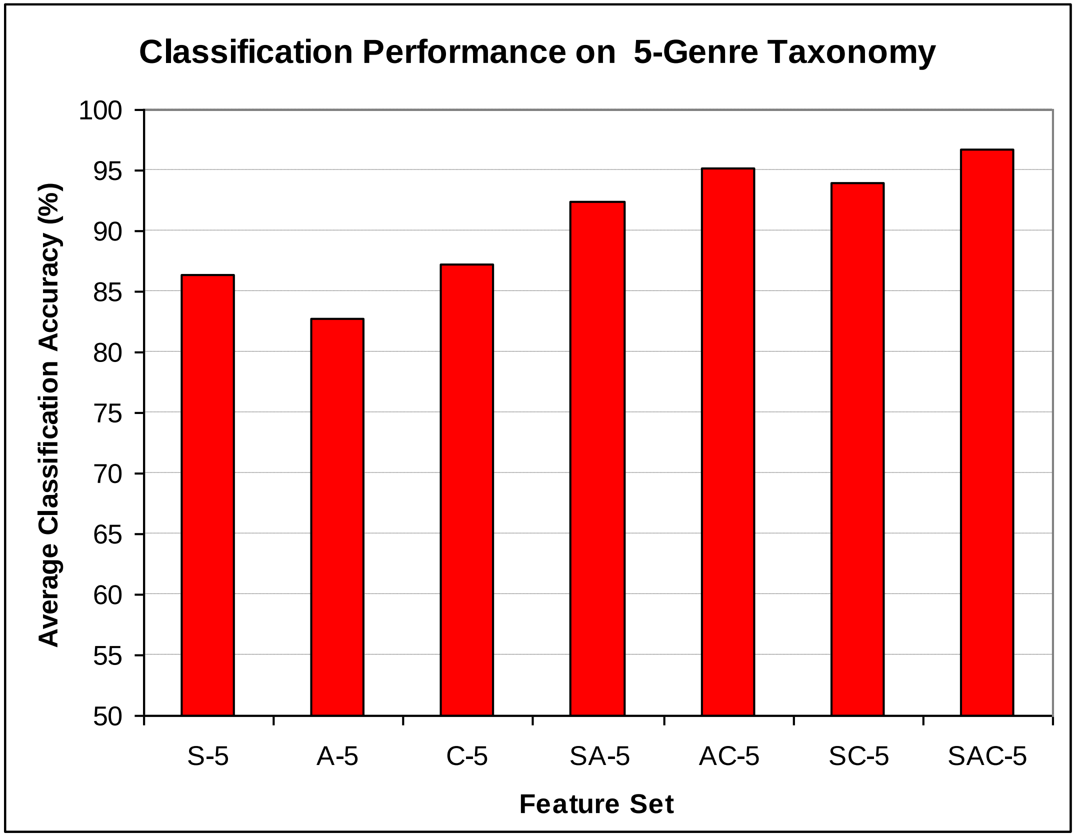
- jMIR is a free and open-source Java software suite designed for general music classification research:
  - jAudio: Audio feature extraction
    - 26 core features + metafeatures and aggregators
  - jSymbolic: Feature extraction from MIDI files
    - 111 mostly original features
  - jWebMiner: Cultural feature extraction
    - Uses search engine co-occurrence page counts
  - ACE: Meta-learning classification system
    - 7 machine learning and 3 dimensionality reduction algorithms
    - Updated version to be released soon

More on jMIR

- jMIR also includes other components
  - ACE XML
  - jMusicMetamanager
  - Codaich
  - Bodhidharma MIDI
  - jMIRUtilities
- More information:
  - jMIR's components have each been described individually in previous publications
  - jmir.sourceforge.net
  - cory.mckay@mail.mcgill.ca

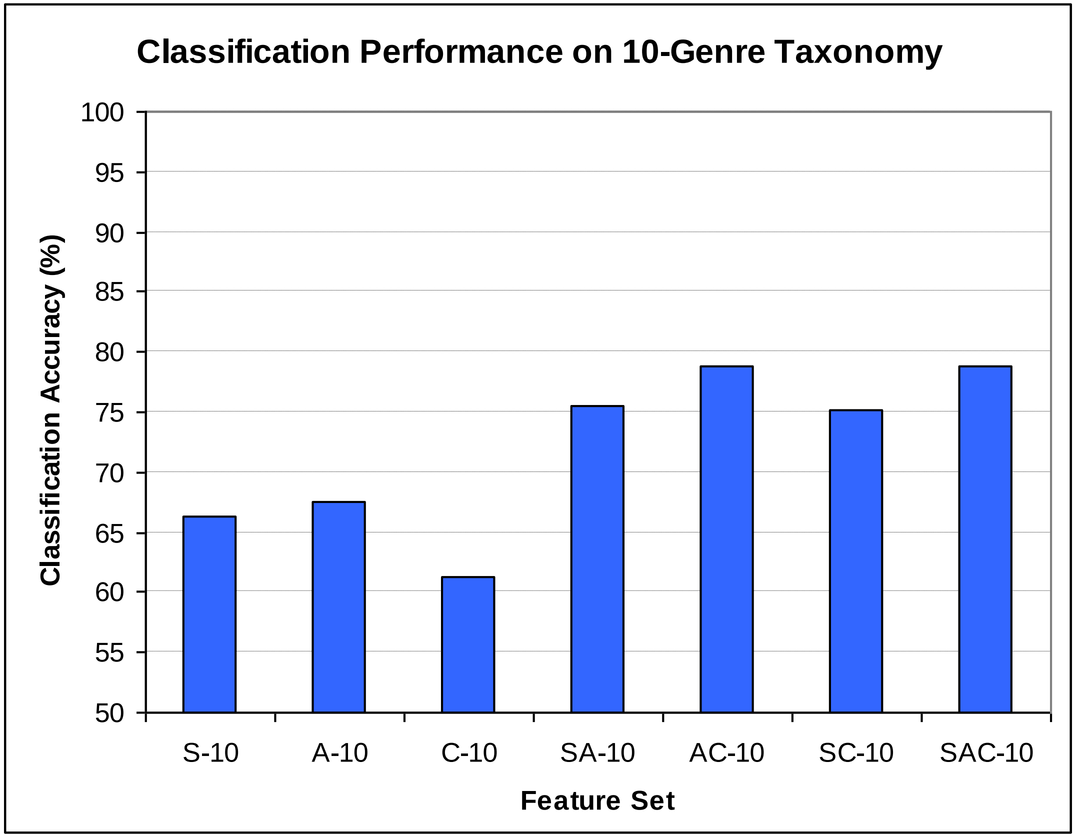
Results: 5-genre taxonomy

- 3 feature types vs. 1
  - 11.3% better
  - Statistically significant (even for small SAC dataset)
  - A 78% decrease in the error rate
- 3 feature types vs. 2
  - 2.3% better
  - Not statistically significant



Results: 10-genre taxonomy

- Trends similar to 5-genre results
- 3 feature types vs. 1
  - 13.7% better
  - Statistically significant
  - A 39.3% decrease in the error rate
- 3 feature types vs. 2
  - 2.7% better
  - Not statistically significant

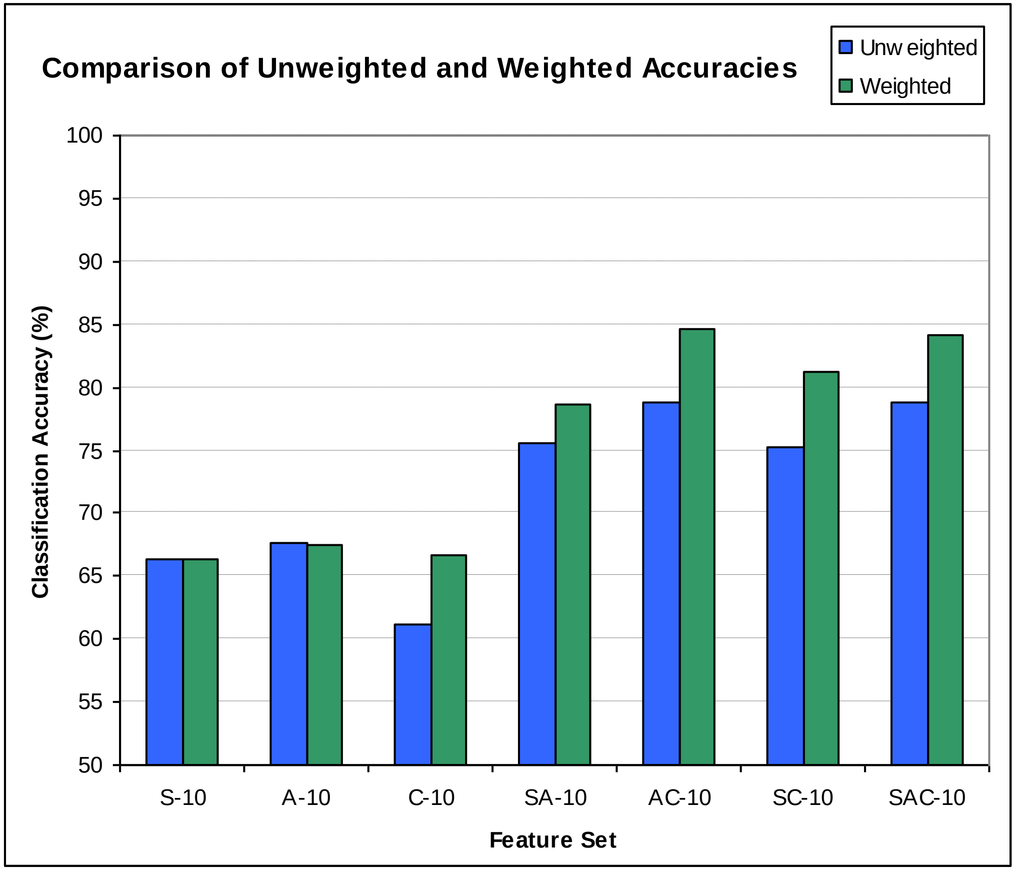


Misclassification seriousness

- Misclassification to a similar genre can be less serious than misclassification to a dissimilar genre
- To investigate this, we calculated normalized weighted classification accuracies for the 10-genre experiments
  - Misclassification within a SAC genre pair: 0.5 error
  - Misclassification outside a SAC genre pair: 1.5 error
- Recall SAC genre pairs:
  - **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

Results: weighted vs. unweighted

- Audio and symbolic
  - No significant difference
  - Although weighted 3% greater than corresponding unweighted when both combined
- Feature groups including cultural features had fewer serious misclassifications than those without cultural features
  - Weighted greater than corresponding unweighted by average of 5.7%
  - Statistically significant

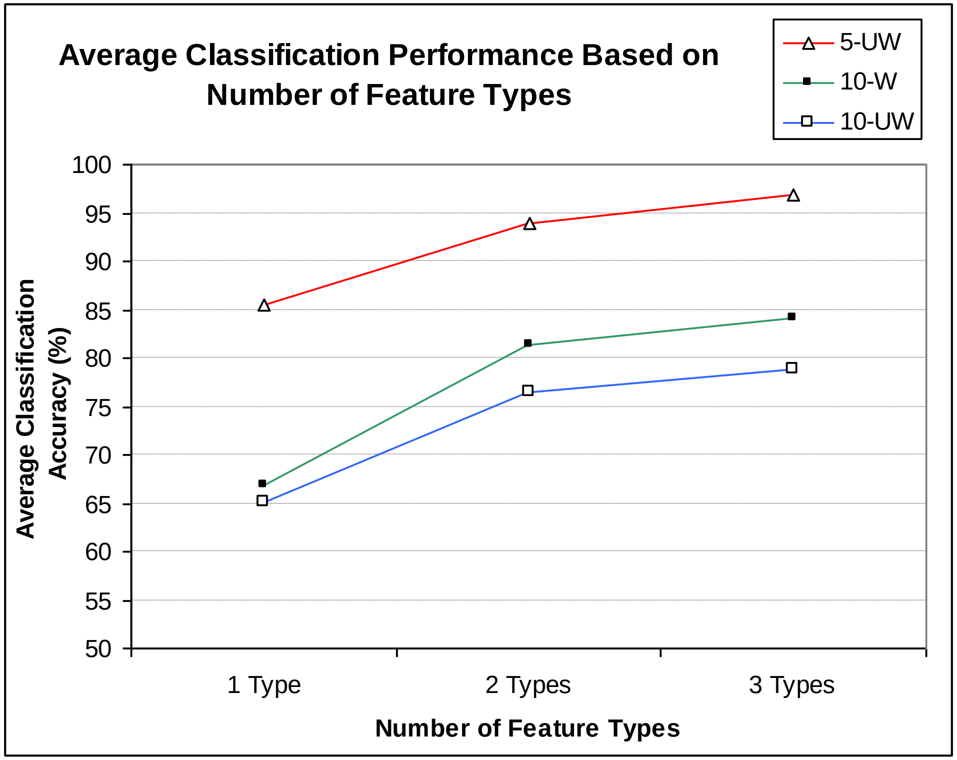


Feasibility of genre classification

- Results still too low for practical application
  - Best on 5-genre taxonomy: 96.8%
  - Best on 10-genre taxonomy: 78.8%
- Results much better than best comparable audio-only results, however:
  - This experiment with jMIR (10 classes): 67.6%
  - MIREX 2008 Audio Genre (10 classes): 66.4%
    - 22,000 tracks
  - MIREX 2007 Audio Genre (10 classes): 68.3%
    - 7000 tracks
- Combining feature types did significantly improve performance results past the seeming ceiling on audio-only classification

Conclusions

- Combining any two or more feature groups improves performance compared to any single feature group
- Using cultural features causes those misclassifications that do occur to be less serious
- The performance ceiling on genre classification performance may not be as low as some have worried



Future research

- Repeat experiments on a much larger dataset
  - Using MIDI files transcribed from audio
  - Incorporating larger class ontologies
- Perform similar experiments with respect to other domains of music classification and similarity measurement
  - Artist, mood, recommendation, etc.
- Experiment with combining feature types and learning models in more sophisticated ways
  - e.g., blackboard classifier ensembles, ontologically structured classification techniques, etc.

Acknowledgements

- Funding, past and future:
  - Social Sciences and Humanities Research Council of Canada
  - The Andrew W. Mellon Foundation
- My doctoral adviser:
  - Ichiro Fujinaga
- Ongoing development of jMIR as part of the NEMA project:
  - Jessica Thompson

- Contact information:
  - cory.mckay@mail.mcgill.ca
  - jmir.sourceforge.net