Automatic Music Classification and Similarity Analysis

Cory McKay
Music Technology Area
McGill University
Montreal, Canada
cory.mckay@mail.mcgill.ca

Abstract

This paper presents a study of automatic music classification and similarity analysis. Important theoretical issues are presented and several experiments are performed. The jAudio feature extraction software and the ACE classification framework are used to perform these experiments, which include percussion timbre recognition, genre classification, speech/music classification and taxonomical evaluation.

1 Introduction

Musical classification and similarity analysis are research topics that have been receiving increasing attention in the music information retrieval community, as well as from machine learning researchers in general. This development parallels a growing theoretical interest in music categorization and overarching theoretical frameworks among musicologists, particularly those researching popular music.

It is unfortunate that there has of yet been relatively little communication between these two groups, as there is important potential for collaboration. Researchers with computer science backgrounds can perform revealing large-scale empirical studies using data mining and pattern recognition techniques, while musicologists can contribute their deep understanding of music and the sociological issues surrounding it to interpret these results. One of the goals of this paper is to highlight certain research that has been done in both the machine learning and musicological domains so that researchers from both communities will have a starting point for future collaboration.

1.1 The intersection of classification and similarity analysis

Although there is certainly a distinction between music classification and "pure" similarity, these fields are highly related, and there are many areas in which they overlap. The essential difference is that, in the case of classification, one has a pre-existing taxonomy upon which recordings are to be classified. The study of similarity, in contrast, attempts to derive some measurement of similarity between pieces of music without reference to an external structure. In both cases, however, one is interested in how one a given recording is related to other recordings. Both classification and similarity also involve the same essential approach of finding appropriate features and supplying these to some pattern recognition algorithm.

One can draw rough parallels between classification and supervised learning and between similarity analysis and unsupervised learning. Although there certainly are exceptions to this division, it does hold in general, as unsupervised learning may be unlikely to cluster recordings into the particular classes that one would like for taxonomies of any sophistication, and supervised learning requires pre-set labellings that may be inappropriate for pure similarity analysis.

1.2 Practical applications

There are numerous practical applications for both music classification and similarity analysis including, to name just a few:

- **Genre classification:** Useful for automatically organizing large databases.
- **Composer identification:** Good for identifying music whose authorship is disputed. This is particularly relevant to early music manuscripts.
- **Performer or song identification:** Useful for automatically indexing unlabelled personal collections or for circumvent copyright protection efforts by detecting false entries on peer-to-peer networks. Can also be used to identify soloists in old recordings with incomplete liner notes.
- **Mood or listening scenario classification:** Can be used for automatic play-list generation.
- **Recommendation systems:** Can help listeners find unknown music that they might be interested in.
- **Hit prediction:** Can attempt to predict the commercial success of recordings in order to efficiently allocate financial resources.

1.3 Theoretical interest

Many of the applications discussed in Section 1.2 are of theoretical interest as well. For example, musical genre is currently poorly understood. Musicologists do not have a general understanding of how genres are created, how they are agreed upon and disseminated, how they are defined, how they are perceived and identified, how they change or how they are interrelate. The writings of Franco Fabbri (1982) and of David Brackett (1995) are of particular interest in understanding musicological issues relating to genre. Experimental research involving both supervised learning and clustering could provide valuable data to musicologists.

Machine learning based research can also help musicologists understand how the musics of different cultures, geographical regions and historical periods have influenced each other. The output of feature selection algorithms can also provide valuable insights on what factors are most important to consider when thinking of different types of music. The clusters output by unsupervised learning algorithms can also be of theoretical interest, as they can present musicologists with new approaches to grouping music.

2 General theoretical issues to consider

The study of pure similarity is both a highly interesting and highly problematic area. Although it is on the surface a simple matter to ask whether two pieces of music are similar, more serious contemplation reveals the many questions that must be considered. How do we quantify similarity? Should we also consider dissimilarity? What features do we base judgments of similarity on? Should we stress cultural or content-based aspects? How do we evaluate similarity analysis systems once they have been implemented? Is there some ground truth, and if so, how do we find it? How does one deal with recordings or musicians that bridge multiple styles?

Although most humans can make intuitive judgments of musical similarity, ranking similarity is a much more difficult task, as is agreeing upon some basis on which similarity judgements can be made. Even listeners from the same culture and with similar backgrounds may use very different considerations.

Even if one is to restrict oneself to the realm of pre-set categories, matters are still far from simple. Different individuals may use entirely different genre categories, for example, or classify individual recordings entirely differently.

Any serious study of music classification should consider psychological research on how humans perform classifications, as the ultimate goal is to produce software that agrees with humans. Eleanor Rosch, for example, has argued that people tend to think of categories as having some typical, or prototypical, members, and other less typical members (Rosch 1975). A robin, for example, can be considered to be a better example of a bird than an ostrich, or a chair a better example of furniture than a magazine rack. This is obviously highly relevant to the study of musical similarity.

There are many other such insights and ideas. George Lakoff (1987) has published a seminal overview relating to human labelling and classification in general. A number of important insights on constructing taxonomies for the particular purpose of automatic music classification have been presented by Pachet and Cazaly (2000) and by Aucouturier and Pachet (2003).

3 Issues relating to features

Any classification algorithm is only as good as the features that it is provided with. It can reasonably be argued that the choice of features is perhaps the most important aspect of most applied pattern recognition. This is particularly problematic in the case of music, as most humans do not think or communicate about music in ways that are easily communicable to a computer. Furthermore, knowledgeable consideration of features can require knowledge of diverse fields, including signal processing, musicology, music theory and text-mining.

Before proceeding to discuss features in more detail, it is important to emphasize that content-based features can be approached in terms of both audio and symbolic data. Due to current limitations in automated transcription, it is currently difficult to extract reliable high-level musical information from audio recordings. Although audio can be synthesized from symbolic recordings, it will certainly not contain much of the useful information that one would find in an original audio recording.

It hoped that, as automatic transcription technology improves, one will be able to extract both high-level and low-level features from audio. As it is now, unfortunately, there is a general division between audio and symbolic research, and relatively little research at all using features based on cultural meta-data. This is unfortunate, as it is probable that research combining the three types of features could be very fruitful.

3.1 Features derived from symbolic recordings

The high-level features that can be extracted from symbolic recordings can be extremely effective, since that they can make use of useful musical abstractions. Some of the best automatic genre classification success rates to date have been achieved using MIDI recordings (McKay and Fujinaga 2004).

Unfortunately, most theoretical musical research has focused on specific types of music, and there is a relative paucity of research on features that could be used for music in general. Furthermore, much of the existing research has focused on sophisticated theoretical models that are subjective and difficult to compute automatically.

There have been, fortunately, a few exceptions. The Cantometrics project quantitatively compared the musics of a variety of cultures (Lomax 1968). Philip Tagg (1982) and David Cope (1991) have published a number of ideas that have general applicability. Arden and Huron (2001) have studied correlations between particular musical features and different geographical regions. McKay and Fujinaga (2005) have found that features based on instrumentation can be particularly powerful with respect to automatic music classification.

3.2 Features derived from audio recordings

Features derived from audio features tend to consist of signal processing based quantities that have little intuitive musical meaning to humans. More high-level information can be derived from them in some cases, however. It is important to note that the degree of accuracy needed for automatic transcription is not necessarily required for features intended for classification. Individual incorrect notes, for example, can be averaged out through the construction of intermediate data structures such as beat histograms or pitch histograms.

Many of the features that have been used for music classification were originally developed for speech processing. Scheirer and Slaney (1997) and Cary, Parris and Lloyd-Thomas (1999) have published useful sets of features with respect to speech/music discrimination. Tzanetakis and Cook (2002) have used many of these features and expanded upon them with the particular needs of music in mind. Researchers in musical instrument identification have also made some important contributions (Eronen 2001; Essed, Richard & David 2004). Further relevant research has been published by Park (2000), McKinney and Breebaart (2003), Pope, Holm and Kouznetsov (2004) and West and Cox (2004).

3.3 Cultural features

Although most research to date has focused on content-based classification, cultural features have the potential to extremely useful as well. Data mining techniques can be used to extract information from the web and other sources to form correlations between particular songs, performers, genre categories, etc. Textual analysis of song lyrics could also prove to be fruitful. Whitman and Smaragdis (2002) and Aucouturier and Pachet (2003) have proposed some important initial steps towards utilizing such cultural information.

Despite the relative paucity of research on cultural features, they are commonly used commercially, albeit in an unsophisticated way. Most on-line music retailers collect statistics on the purchases of individual customers and correlate them with the purchases of others in order to recommend music.

Although this can be effective in many cases, it has three important weaknesses. Firstly, individuals often buy presents for others, thereby introducing potentially significant noise into the system. Secondly, this approach tends to average out the preferences of people with atypical tastes. Thirdly, this approach tends to overlook relatively poorly known music, as there are fewer purchasers of it. Even if popularity weighting is introduced to compensate for this, there is still a much greater error associated with such recordings due to small sample sets. This is problematic, as the most important function of a music recommendation system is to recommend music that users are unlikely to have heard of previously. There is clearly much potential to cultural features, but also much to be done beyond this naïve approach.

4 Previous research on audio classification and similarity

The experiments discussed later in this paper concentrate on audio classification and analysis, so the remainder of this paper will deal solely with research on audio work. McKay (2004) includes a review of an overview of research on classifying of symbolic recordings.

Scheirer and Slaney (1997) have published some of the most influential early modern research on sound classification. Tzanetakis and Cook (2002) are responsible for bringing serious pattern recognition techniques to the forefront of the music information retrieval community through their research on genre classification. These results were later improved by Li and Tzanetakis (2003). There have been too many recent publications on audio classification to cite comprehensively here, but McKay (2004) may be consulted for a detailed summary of automatic genre classification research up to 2004. Much of this research has essentially consisted of straight forwardly applying a variety of basic machine learning techniques to the types of features proposed in the papers discussed in Section 3.2.

There has been somewhat less research with regards to musical similarity, probably because of the difficulty in evaluating such systems. There has been significant commercial interest in this area, however (Cuidado, Goombah, Comparisonics, Polyphonic HMI, etc.), and there is certainly scholarly interest as well, as discussed in Section 1.3.

Foote (1997) performed some of the earliest research in this area, although he used his results to perform classifications rather than studding pure similarity. He performed classifications by generating templates and then measuring distances of test instances from them.

Ellis et al. (2002) and Berenzweig et al. (2003) have published excellent work on attempting to come to terms with the issue of ground truth and musical similarity. They have proposed assessing the opinions of “average listeners” and using the results to train classifiers. They have based their research on a variety of sources, including peer-to-peer statistics, an on-line survey and expert opinion. Their system produces a similarity matrix that compares the similarity of all possible pairs of artists, an approach that has since been adopted by many others.

Aucouturier and Pachet (2002) have made the important contribution of using meta-data to constrain similarity results. For example, one could ask a software system to filter similarity results so that only recordings by a certain artist, in a certain language, belonging to a certain genre, etc. are returned. Aucouturier and Pachet propose further browsing tools for refining searches in a later paper (2004).

Rauber et al. (2002) have proposed using psycho-acoustic models to create a hierarchical organization of music archives based on perceived similarity. This imposition of structure on similarity results is another important way of making them more manageable.

Pampalk (2004) has constructed a MatLab toolbox for measuring audio similarity. This includes an update on the visual *Islands of Music* interface that he has constructed in order to help humans make sense of similarity measurements.

5 Implementation

The experiments presented in Section 6 were performed using the jAudio feature extractor and the ACE classification framework (McKay et al. 2005). This software is implemented entirely in Java, and makes use of the Tritonus alternative Java Sound implementation and the Weka data-mining class library (Witten and Frank 2000). Additional functionality was added to this software during the course of this project.

jAudio and ACE have both been developed to make automatic music classification more accessible and effective. They have been designed to bring powerful general digital signal processing and machine learning algorithms to the music information retrieval community while emphasizing the particular needs of music researchers. Special effort has been exerted to make the software portable and, most importantly, easily extensible. The software has significant functionality outside of the scope of this project, and only the directly relevant aspects of it are presented here.

5.1 Feature extraction with jAudio

The jAudio software extracts features from audio recordings and saves them in ACE XML files, which may also be converted to Weka ARFF format. A variety of pre-processing techniques may be performed, and features may be saved for individual windows or for recordings as a whole. In the particular experiments described here, all recordings were downsampled to 16 kHz, stereo channels were merged into mono and features were extracted for each non-overlapping window of 512 samples. Although a variety of other features are extractable by jAudio, the features used for the experiments described in Section 6 consisted of the overall average and standard deviations over all windows of the following features:

- **Beat Sum:** The sum of all bins in a beat histogram. This is a measure of the importance of regular beats in a signal.
- **Compactness:** A measure of the noisiness of a recording. Found by comparing the components of a window’s magnitude spectrum with the magnitude spectrum of its neighbouring windows.
- **Fraction Of Low Energy Frames:** The fraction of the last 100 windows that had an RMS less than the mean RMS of the last 100 windows. This can indicate how much of a signal section is quiet relative to the rest of the signal section.
- **Root Mean Square (RMS):** A measure of the power of a signal.
- **Root Mean Square Derivative:** The window to window change in RMS.
- **Spectral Centroid:** The centre of mass of the power spectrum.

- **Spectral Flux:** A measure of the amount of spectral change in a signal. Found by calculating the frame to frame change in magnitude spectrum.
- **Spectral Rolloff Point:** The fraction of bins in the power spectrum at which 85% of the power is at lower frequencies. This is a measure the right-skewedness of the power spectrum.
- **Spectral Variability:** The standard deviation of the magnitude spectrum. A measure of how varied the magnitude spectrum of a signal is.
- **Strength Of Strongest Beat:** How strong the strongest beat in the beat histogram is compared to other potential beats.
- **Strongest Beat:** The strongest beat in a signal, in beats per minute, found by finding the highest bin in the beat histogram.
- **Strongest Frequency:** An estimate of the strongest frequency component of a signal, in Hz. This was found three different ways: via the spectral centroid, via taking the power spectrum bin with the highest power and via zero-crossings.
- **Strongest Frequency Variability:** The standard deviation of the frequency of the power spectrum bin with the highest power over the last 100 windows.
- **Zero Crossings:** The number of times the waveform changes sign in a window. A possible indicator of fundamental frequency and of noisiness.
- **Zero Crossings Derivative:** The absolute value of the window to window change in zero crossings.

5.2 Pattern recognition with ACE

The ACE software automatically experiments with a variety of machine learning techniques in order to find ones well suited to a particular problem. All evaluations were done using 10-fold cross-validation.

The machine learning techniques used by ACE include k-NN (unweighted, weighted by distance and weighted by similarity), a naïve Bayesian classifier, feedforward neural networks, support vector machines and a C4.5 decision tree. Experiments are also performed in combining classifiers using bagging and AdaBoost. ACE also experiments with various feature dimensionality reduction techniques. These currently consist of principal component analysis and genetic algorithms.

6 Experiments, results and conclusions

The first test of the ACE system was to run it on several of the UCI benchmark datasets (Blake and Merz 1998). This was done in order to evaluate the effectiveness of the machine learning algorithms themselves. The results are shown on Table 1, and are compared with other recently published results (Kotsiantis and Pintelas 2004) on the same datasets. These results are very encouraging, as ACE performed better than the reference on all data sets but one.

Table 1: Comparative results on UCI datasets

DATA SET	CLASSES	SUCCESS RATE (Kotsiantis 2004)	SUCCESS RATE (ACE)
autos	7	81.8%	86.3%
diabetes	2	76.6%	78.0%
ionosphere	2	90.7%	94.3%
iris	3	95.6%	97.3%
labor	2	93.4%	93.0%
vote	2	96.1%	97.0%

Once the efficacy of ACE had been demonstrated, it was appropriate to turn to data sets specific to music. Three types of tasks were used to test ACE, namely percussive sound identification, genre classification and speech/music discrimination. The first two types of tasks are highly useful but difficult types of music classification that have been receiving significant research attention recently, and the third task is included to see if the system could perform well on a relatively easy task.

Two different experiments were performed involving percussive sound identification. This is a particularly difficult type of classification, as percussive sounds generally consist of unstable transients. The first experiment involved snare drum stroke classification. The data set compiled by Tindale et al. (2004) was used for this experiment, and it involved 58 features and 7 classes (centre, half centre, halfway, halfedge, edge, rimshot and brush). The second experiment involved as of yet unpublished beatboxing data collected by Elliot Sinyor consisting of five classes (kick, open, closed, k-snare and p-snare). Tindale et al.'s features were used for the snare stroke classification and jAudio was used to extract the features from the beatboxing data.

Two genre classification experiments were also performed. The first involved the data set collected by Tzanetakis and Cook (2002), which consists of 10 genres (blues, classical, country, disco, hip hop, jazz, metal, pop, reggae and rock). The second experiment involved an as of yet unpublished data set collected by Douglas Eck and his students. This data set consisted of 6 genres (classical, country, electronic, jazz (bebop), rap and rock). Finally, a speech/music discrimination was performed on the Scheirer and Slaney (1997) data set. jAudio was used to collect features for all of three of these data sets.

Table 2 shows the results of all four of these experiments, including previously published results by others, where appropriate.

Table 2: Results on music-related experiments

EXPERIMENT	PREVIOUS RESEARCH	SUCCESS RATE (PREVIOUS)	SUCCESS RATE (ACE)
Snare drum stroke	Tindale et al. 2004	95%	96%
Beatboxing	--	--	96%
Tzanetakis genre	Li & Tzanetakis 2003	71%	62%
Eck genre	--	--	73%
Speech/music	Scheirer & Slaney 1997	94%	98%

The speech/music discrimination results were very good, and both of the percussion recognition experiments resulted in high success rates. Unfortunately, the genre classification results were somewhat disappointing. However, upon consideration, these results are perhaps not surprising. Existing genre research involved a fine-tuned set of specialized pre-processing, feature extraction and classification methodologies with the particular needs of genre classification in mind. ACE, in contrast, is a general system.

Even though ACE does automatically experiment with a variety of feature selection techniques and classifiers, the possible palette from which the it has to choose is still limited. The relatively poor performance of ACE on genre classification was probably due to the fact that no high-level features were used, other than a few basic statistics derived from beat histograms. Genre classification, unlike the other experiments performed here, is probably particularly well segmented by such high-level features, and Li and Tzanetakis made use of a variety of rhythmic and pitched based features that probably had a significant effect on their success rate. This presents a strong argument for incorporating such features into jAudio in the future.

A final experiment was performed where k-means clustering was applied to the Tzanetakis data set. This was done in order to illustrate just one example of how similarity measurements can be used musicologically. The approach used here was entirely different from that of the papers described in Section 4, where attempts were made to judge the similarity of different artists or songs. Although this type of research is certainly important both practically and theoretically, it was decided to take the original approach here of studying genre categories themselves.

K-means clustering is a simple but effective way of clustering instances into similar groups of varying sizes around strong centres of mass in feature space. K was set to 10 here in order to match the number of model genre classes in the data set. The clustering results are shown in Table 3.

An examination of Table 3 shows some interesting patterns. The instances belonging to classical, country, disco, metal and rock were all each assigned in their entirety to a single cluster, and almost all reggae classes were assigned to the same cluster. This likely indicates a strong grouping of these classes, and reaffirms the appropriateness of the labels that were used. A number of blues and jazz recordings were grouped into the same cluster as country and, similarly, a number of blues and jazz recordings were grouped into the same cluster as rock. This is reasonable, given the similarity that many humans would probably agree exists between these types of music (with the probable exception of jazz and country). This could indicate that it might be more reasonable to assign multiple classes to some instances rather than enforcing only a single class membership.

More interestingly, many jazz recordings were placed in the same cluster as the classical recordings. Although these two genres are generally very different musically, they are the only two "art" genres in the experiment. Perhaps the clustering is recognizing a certain sophistication in the timbre or the dynamics of these genres as opposed to the others.

Both hip hop and pop were each split among two different clusters. This may be an indication that it would be useful to split them both into sub-genres. Blues, in contrast, was spread throughout the clusters, perhaps an unsurprising result given its overlap with many other genres.

Although all of these observations are essentially speculation, this is a good illustration of how even a rudimentary clustering algorithm applied to a rudimentary taxonomy can reveal very interesting insights into the appropriateness of the taxonomy as well as ideas on how it might be modified. There is room for much more research in this area, including varying the number of clusters or experimenting with different clustering algorithms and taxonomies.

Table 3: Unsupervised k-means clustering of the Tzanetakis data set.
Rows correspond to model categories and columns correspond to clusters.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Blues	16	39	1	27	3	5	3	6		16
Classical										110
Country		112								
Disco					102					
Hip Hop			49				57			
Jazz		36		9	1	2	2	4		51
Metal						101				
Pop	64								55	
Reggae			1					112	3	
Rock				111						

7 Conclusions

This paper discussed a number pf important theoretical issues relating to automatic music classification and similarity analysis. Several experiments were performed in order to demonstrate how these theoretical ideas can be applied.

The jAudio and ACE system were found to able to deal effectively with a variety of classification tasks with no manual tweaking or user configuration at all. There is still certainly room for general improvement, however, particularly with regards to the addition of more features and a greater variety of feature selection techniques. Features based on relatively high-level information extracted from audio are particularly promising, and could help ACE to perform well on specialized tasks. In addition, brief experimentation with clustering demonstrated the potential of unsupervised learning when applied to musicological research.

It is hoped that the results and background presented in this paper will help to frame and encourage future collaborative research in musicology and pattern recognition.

References

Aarden, B., and D. Huron. 2001. Mapping European folksong: Geographical localization of musical features. *Computing in Musicology* 12: 169–83.

Aucouturier, J. J., and F. Pachet. 2004. Tools and architecture for the evaluation of similarity measures: Case study of timbre similarity. *Proceedings of the International Conference on Music Information Retrieval.* 198–203.

Aucouturier, J. J., and F. Pachet. 2003. Representing musical genre: A state of the art. *Journal of New Music Research* 32 (1): 1–12.

Aucouturier, J. J., and F. Pachet. 2002. Music similarity measures: What's the use? *Proceedings of the International Conference on Music Information Retrieval*.

Berenzweig, A., B. Logan, D. Ellis, and B. Whitman. 2003. A large-scale evaluation of acoustic and subjective music similarity measures. *Proceedings of the International Symposium on Music Information Retrieval.* 99-105.

Blake, C., and C. Merz. 1998. *UCI repository of machine learning databases.* Retrieved April 13, 2005, from www.ics.uci.edu/~mlearn/MLRepository.html. University of California, Irvine, Dept. of Information and Computer Sciences.

Brackett, D. 1995. *Interpreting popular music*. New York: Cambridge University Press.

Carey, M. J., E. S. Parris, and H. Lloyd-Thomas. 1999. A comparison of features for speech, music discrimination. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. 149–152.

Cope, D. 1991b. *Computers and musical style*. Madison, WI: A-R Editions.

Ellis D., B. Whitman, A. Berenzweig, and S. Lawrence. 2002. The quest for ground truth in musical artist similarity. *Proceedings of the International Conference on Music Information Retrieval.*

Eronen, A. 2001. Comparison of features for musical instrument recognition. *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 753–6.

Essed, S., G. Richard, and B. David. 2004. Musical instrument recognition based on class pairwise feature selection. *Proceedings of the International Conference on Music Information Retrieval.* 560–8.

Fabbri, F. 1982. What kind of music? *Popular Music* 2: 131–43.

Foote, J. T. 1997. A Similarity Measure for Automatic Audio Classification. Proceedings of the *AAAI 1997 Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora*.

Knees, P., E. Pampalk, and G. Widmger. 2004. Artist classification with web-based data. *Proceedings of the International Conference on Music Information Retrieval.* 517–24.

Kotsiantis, S., and P. Pintelas. 2004. Selective voting. *Proceedings of the International Conference on Intelligent Systems Design and Applications*. 397–402.

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Li, T., and G. Tzanetakis. 2003. Factors in automatic musical genre classification of audio signals. *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 143–6.

Lomax, A. 1968. *Folk song style and culture*. Washington: American Association for the Advancement of Science.

Mckay, C., R. Fiebrink, D. McEnnis, B. Li and I. Fujinaga. 2005. ACE: A framework for optimizing music classification. Submitted for publication to the *International Conference on Music Information Retrieval.*

McKay, C. and I. Fujinaga. 2005. Automatic music classification and the importance of instrument identification. *Proceedings of the Conference on Interdisciplinary Musicology*.

McKay, C. 2004. Automatic genre classification of MIDI recordings. *M.A. Thesis*. McGill University, Canada.

McKay, C. and I. Fujinaga. 2004. Automatic genre classification using large high-level musical feature sets. *Proceedings of the International Conference on Music Information Retrieval*. 525–30.

McKinney, M. F., and J. Breebaart. 2003. Features for audio and music classification. *Proceedings of the International Symposium on Music Information Retrieval.* 151–18.

Pachet, F., and D. Cazaly. 2000. A taxonomy of musical genres. *Proceedings of the Content-Based Multimedia Information Access Conference.*

Pampalk, E.. 2004. A MatLab toolbox to compute music similarity from audio. *Proceedings of the International Conference on Music Information Retrieval.* 254–7.

Park, T. H. 2000. Salient feature extraction of musical instrument signals. *Master's thesis.* Dartmouth College, USA.

Pope, S. T., F. Holm, and A. Kouznetsov. 2004. Feature extraction and database design for mu-sic software. *Proceedings of the International Computer Music Conference.* 596–603.

Rauber, A., E. Parnpalk and D. Merkl. 2002. Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by sound similarity. *Proceedings of the International Symposium on Music Information Retrieval.*

Rosch, E. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104: 192–233.

Scheirer, E., and M. Slaney. 1997. Construction and evaluation of a robust multi-feature speech/music discriminator. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*.

Tagg, P. 1982. Analysing popular music: Theory, method and practice. *Popular Music* 2: 37–67.

Tindale, A., A. Kapur, G. Tzanetakis, and I. Fujinaga. 2004. Retrieval of percussion gestures using timbre classification techniques. *Proceedings of the International Conference on Music Information Retrieval.* 541–4.

Tzanetakis, G., and P. Cook. 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing* 10 (5): 293–302.

West, C., and S. Cox. 2004. Features and classifiers for the automatic classification of musical audio signals. *Proceedings of the International Conference on Music Information Retrieval.* 531–7.

Whitman, B., and P. Smaragdis. 2002. Combining musical and cultural features for intelligent style detection. *Proceedings of the International Symposium on Music Information Retrieval.* 47–52.

Witten, I., and E. Frank. 2000. *Data mining: Practical machine learning tools and techniques with Java implementations*. San Francisco: Morgan Kaufmann.